

# TOWARDS A GENERAL THEORY FOR THE EVOLUTION OF APPLICATION DOMAINS

H.A. PROPER

AND

TH.P. VAN DER WEIDE

*Department of Information Systems, University of Nijmegen,  
Toernooiveld, NL-6525 ED Nijmegen, The Netherlands,  
E.Proper@acm.org*

PUBLISHED AS:

H.A. Proper and Th.P. van der Weide. Towards a General Theory for the Evolution of Application Models. In M.E. Orłowska and M.P. Papazoglou, editors, *Proceedings of the Fourth Australian Database Conference*, Advances in Database Research, pages 346–362, Brisbane, Australia, February 1993. World Scientific, Singapore. ISBN 981021331X

## ABSTRACT

In this article we focus on evolving information systems. First a delimitation of the concept of evolution is discussed. The main result is a first attempt to a general theory for such evolution. In this theory, the underlying data model is a parameter, making the theory applicable for a wide range of modelling techniques.

## 1. Introduction

As has been argued in [Rod91] and [FOP92b], there is a growing demand for information systems, not only allowing for changes of their information base, but also for modifications in their underlying structure (conceptual schema and specification of dynamic aspects). In case of snapshot databases structure modifications lead to costly data conversions and reprogramming.

The intention of an evolving information system ([FOP92a]) is to be able to handle updates of all components of the so-called *application model*, containing the information structure, the constraints on this structure, the population conforming to this structure and the possible operations. The theory of such systems should, however, be independent of whatever modelling technique is used to describe the application model. In this paper,

we discuss a general theory for the evolution of application models. The central part of this theory will make weak assumptions on the underlying modelling technique, making it therefore applicable for data modelling techniques such as ER ([Che76]), NIAM ([NH89]) and PSM ([HPW92]), and action modelling techniques such as Task Structures ([WHO92]), DFD ([BW89]) and ExSpect ([HSV89]).

Version modelling in engineering databases can be seen as a restricted form of evolving information systems ([Kat90], [JMSV92]). An important requirement for evolving information systems, not covered by version modelling systems, is that changes to the structure can be made on-line. In version modelling, a structural change requires the replacement of the old system by a new system. Other research regarding evolving information systems can be found in [MS90], in which an algebra is presented that allows relational tables to evolve by changing their arity.

The structure of the paper is as follows. In section 2 we describe the approach that has been taken to the concept of evolution. We will not focus on a particular modelling technique. In section 3 we describe what the minimal requirements of a modelling technique. By considering all elements that may be subject to evolution, we then introduce in section 4 the application model universe, and describe what constitutes a application model version. Finally, in section 5 the evolution of an application model is formally introduced, and some properties of wellformedness are presented.

## 2. An Approach to Evolving Information Systems

In this section we discuss our approach to evolving information systems. We start with the identification of that part of an information system that may be subject to evolution. From this identification, the difference between a traditional information system, and its evolving counterpart, will become clear. This is followed by a discussion on how the evolution of an information system is modelled.

### 2.1. A hierarchy of models

According to [ISO87], a conceptual (i.e. complete and minimal) specification of a universe of discourse consists of the following components:

1. an *information structure*, a set of *constraints* and a *population* conforming to these requirements.
2. a set of *action specifications* describing the transitions that can be performed by the system.

The set of action specifications is referred to as the *action model*. The *world model* encompasses the combination of information structure, constraints and the population. A conceptual specification of a universe of discourse, containing both the action and world

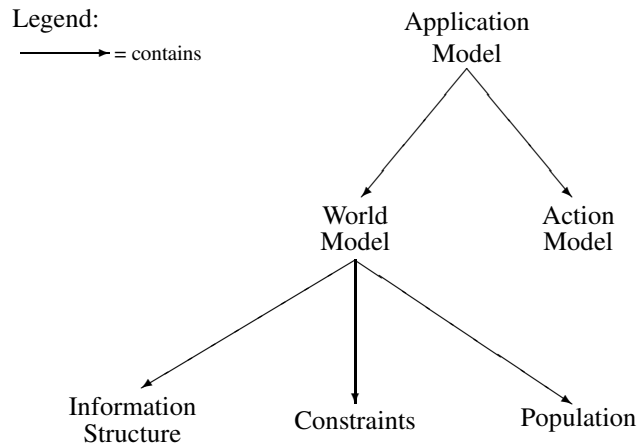


Figure 1: A hierarchy of models

model, is called an *application model* ([FOP92a], [FOP92b]). The resulting hierarchy of models is depicted in figure 1.

## 2.2. An example of evolution

Traditionally, a *world model* consists of a fixed data model according to some data modelling technique, and a population conforming to this data model. The *action model* describes all possible transitions on populations, and is usually modelled by means of Petri-net like specifications (such as ExSpect or Task Structures), or languages such as SQL. The evolvable part of the application model in these cases is restricted to the population.

Some traditional information systems do allow however, to a limited extent, modifications of other components from the application model. For example, adding a new table in an SQL system is easily done. However, changing the arity of a table, or some of its attributes, will result in a time consuming table conversion, which also leads to loss of the old table! In an evolving information system, the entire application model is allowed to evolve on-line, without loss of any information.

As an illustration of an evolving universe of discourse, consider a rental store, for audio records (LP's). In this store a registration is maintained of the songs that are recorded on the available LP's. In order to keep track of the wear and tear of LP's, the number of times an LP has been lent is registered. The information structure and constraints of this universe of discourse are modelled in figure 2 in the style of ER, according to the conventions of [You89]. Note the special notation of attributes ( $\tau$ title) using a mark symbol (#) followed by the attribute (#  $\tau$ title).

An action specification in this example is the rule *Init-freq*, stating that whenever a new LP is added to the assortment of the store, it's lending frequency must be set to 0:

ACTION *Init-freq* =

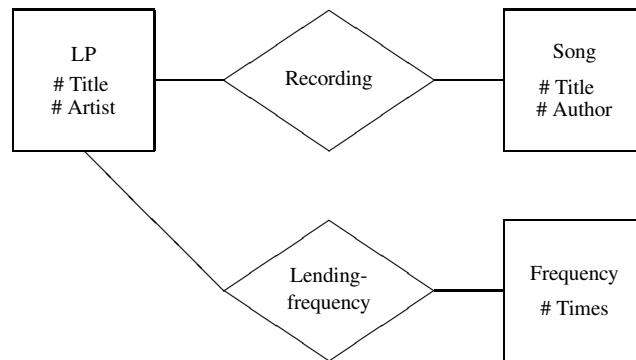


Figure 2: The Data Model of an LP rental store

```

WHEN ADD Lp: x DO
  ADD Lp: x has Lending-frequency of Frequency: 0
  
```

This action specification is in the style of Lisa-D ([HPW93]). Note that the keyword ‘has’ connects object types to relation types, and the keyword ‘of’ just the other way around.

After the introduction of the compact disc, and its conquest of a sizeable piece of the market, the rental store has transformed into an ‘LP and CD rental store’. This leads to the introduction of object type Medium as a generic term for LP’s and CD’s. The relation type Medium-type effectuates the subtyping of Medium into LP and CD. In the new situation, the registration of songs on LP’s is extended to cover CD’s as well. The frequency of lending, however, is not kept for CD’s, as CD’s are hardly subject to any wear and tear. As a consequence, the application model has evolved to figure 3.

The action specification `Init-freq` evolves accordingly, now stating that whenever a medium is added to the assortment of the rental store, it’s lending frequency is set to 0 provided the medium is an LP:

```

ACTION Init-freq =
  WHEN ADD Medium: x DO
    IF Lp: x THEN
      ADD Lp: x has Lending-frequency of Frequency: 0
  
```

### 2.3. The approach

The two ER schemata, and the two action specifications, as discussed above, correspond to two distinct snapshots of an evolving universe of discourse. Several approaches can be taken to the modelling of this evolution. A first approach is to model the history of application model elements by adding birth-death relations to all object types in the information structure ([SA85]). This approach, however, is very limited, as it only enables the modelling of evolution of the population of an information system. For example, the

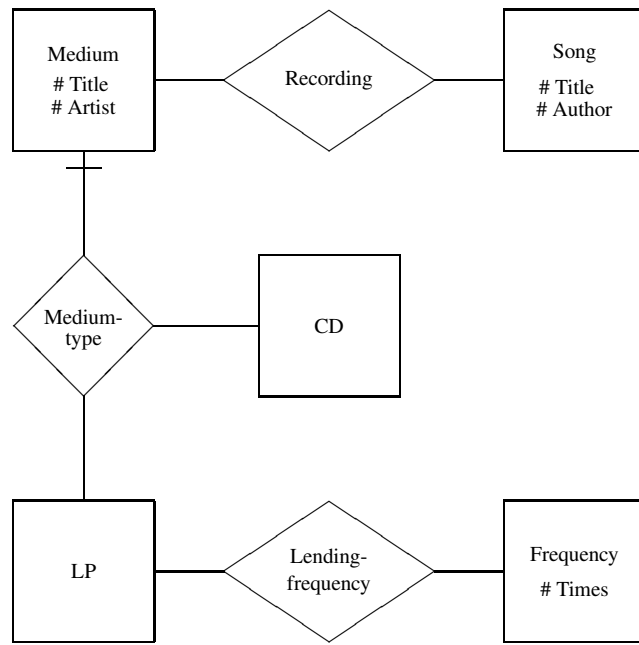


Figure 3: The Data Model of a LP and CD rental store

evolution of the Recording relation type can not be modelled in this approach. Evolution of other application model elements than population, can then be described by a meta model approach.

This paper takes another approach, and treats evolution of an application model as a separate concept. There still are two alternatives to deal with the history of application models. The first one is to maintain a version history of application models in their entirety, leading to a sequence of snapshots of application models. The second one is to keep a version history per element, so keeping track of the evolution of object types, instances, methods, etc.

An advantage of this second alternative is that it enables one to state rules about, and query the evolution of distinct objects. This alternative also allows for the formulation of rules concerning the well formedness of the evolution of application model elements ([PW95]). The first alternative clearly does not offer this opportunity, as it does not provide relations between successive versions of the application model.

Furthermore, the snapshot view from the first alternative can be derived by constituting the application model version of any point of time from the current versions of its components. In the theory of evolving application models we will therefore adapt the second alternative.

An *application model element history* describes the history of an application model element, and is seen as a partial function assigning to points of time the actual occurrence (version) of the element. For example, when CD's are added to the assortment of the rental store, the version of the application model element `Recording` changes from a relation

type that registrates songs on LP's to a relation type that registrates songs on Media. An *application model history* is a set of application model element histories. The current version of an application model then is constituted by the current versions of all application model elements.

### 3. A World Model for Application Models

In this section we take a closer look at application models, and define what constitutes an application model version. An application model version is formulated conforming to some modelling technique. The only assumption on this modelling technique is, that it delimits an *application model universe*, i.e., the space for the evolution of application models.

#### 3.1. The Information Structure Universe

The kernel of the application model universe is formed by the *information structure universe*  $\mathcal{U}_{IS} = \langle \mathcal{L}, \mathcal{N}, \sim, \rightsquigarrow \rangle$ , fixing the evolution space for information structures. Further refinements of the information structure universe depend on the chosen data modelling technique (such as NIAM, ER, PSM and Object Oriented data models), and are beyond the scope of the theory in this paper. For our purposes, an information structure universe is assumed to provide (at least) the above components, which are available in all conventional high level data modelling techniques.

##### 3.1.1. Object Types

The central part of an information structure is formed by its object types (referred to as object classes in object oriented approaches). Two major classes of object types are distinguished. Object types whose instances can be represented directly (denoted) on a medium (strings, natural numbers, etc) form the class of label types  $\mathcal{L}$ . The other object types, for instance entity types or fact (relation) types, form the class  $\mathcal{N}$ . The example of figure 2 contains nine object types: three entity types Record, Song and Frequency, two relation types Recording and Lending-frequency, and four label types Title, Artist, Author and Times.

##### 3.1.2. Type Relatedness

For object types, the (reflexive and symmetrical) relation  $\sim$  expresses *type relatedness* between object types (see [HW93]). Object types  $x$  and  $y$  are termed type related ( $x \sim y$ ) iff populations of object types  $x$  and  $y$  may have values in common. Type relatedness corresponds to mode equivalence in programming languages ([WMP<sup>+</sup>76]). The relation of type relatedness can be recognised in conventional modelling techniques like ER, NIAM, or PSM, as well as object oriented data modelling techniques. Typically, subtyping and generalisation lead to type related object types. For the data model depicted in figure 2, the type relatedness relation is the identity relation:  $x \sim x$  for all object types  $x$ .

### 3.1.3. The Identification Hierarchy

In data modelling, a crucial role is played by the notion of object identification: each object type of an information structure should be identifiable. In a subtype hierarchy however, a subtype inherits its identification from its super type, whereas in a generalisation hierarchy the identification of a generalised object type is inherited from its specifiers. For the data model depicted in figure 3 this means that instances of LP and CD are identified in the same way as instances of Medium. An object type from which the identification is inherited, is termed a *parent* of that object type. The inheritance hierarchy (identification hierarchy) is provided by the relation  $x \rightsquigarrow y$ , meaning that  $x$  is the parent of  $y$ . For figure 3 this leads to:  $\text{Medium} \rightsquigarrow \text{LP}$  and  $\text{Medium} \rightsquigarrow \text{CD}$ .

Object types in an information structure that have no parent are called *roots* as they form the roots of the inheritance hierarchy (a directed acyclic graph). The roots of an object type  $y$  are found by:  $x \text{ RootOf } y \triangleq (x = y \vee x \rightsquigarrow y) \wedge \neg \exists_x [z \rightsquigarrow x]$ .

For every data model from conventional data modelling techniques, a parent ( $\rightsquigarrow$ ) and root (RootOf) relation can be derived. If no specialisations or generalisations are present in that data model, the parent relation will be empty. As a result, the root relation will be the identity relation. For instance the root relation for figure 2 is:  $x \text{ RootOf } x$  for every object type  $x$ .

## 3.2. Properties of Information Structure Universes

Two interesting properties on the root and parent relation are:

**Lemma 3.1** (*common roots*)

$$x \sim y \iff \exists_z [z \text{ RootOf } x \wedge z \sim y]$$

The intuition behind this property is that type relatedness must be based on the type relatedness of some root. Furthermore, type relatedness of roots implies type relatedness of object types:

**Theorem 3.1** (*type relatedness propagation*)

$$\exists_{z_1, z_2} [z_1 \sim z_2 \wedge z_1 \text{ RootOf } x \wedge z_2 \text{ RootOf } y] \iff x \sim y$$

These properties have been proven in [PW95].

## 4. Secondary elements of Application Models

The hierarchy of models (see figure 1) describes how an application model is constructed from other (sub)models. However, this hierarchy disregards relations that must hold between these submodels, for example, the relation between a population and the information structure. These relations are crucial elements of an application model, as they form the fabric of the application model.

## 4.1. Application Model Universe

An application model version is bound to the *application model universe*  $\mathcal{U}_\Sigma$ , which is captured in the tuple:  $\langle \mathcal{U}_{IS}, \gamma, \mu, \mathcal{D}, \Omega, \text{Depends} \rangle$ . The information structure universe  $\mathcal{U}_{IS}$  was introduced in the previous section.

### 4.1.1. Constraints

Most data modelling techniques have a language in which constraints can be expressed, resulting in the set  $\gamma$  of all possible constraint definitions. Constraints are treated as application model elements, that assign to (some) object types a particular constraint definition. A constraint is said to be *owned* by an object type, if the object type has assigned a constraint definition by the constraint. Constraints are inherited via the identification hierarchy. However, as in object oriented data modelling techniques, overriding of constraint definition in identification hierarchies is possible (see for instance [De 91]).

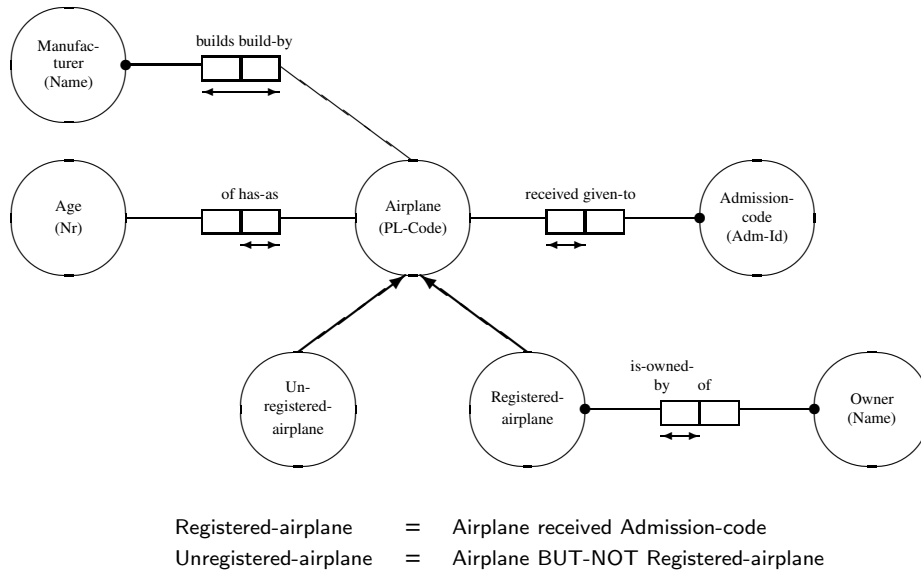


Figure 4: Constraint assignment

As an illustration of the assignment of constraints to object types, consider figure 4. The depicted data model is conforming to NIAM, while the subtype defining rules have been formulated in LISA-D. The modelled universe of discourse is concerned with the administration of airplanes. As airplanes should be replaced in time, the age of an airplane is an important attribute. Furthermore, an airplane may be registered by an aviation association, in which case it has associated an admission code. The owner of registered planes is maintained by the administration.

The graphical constraints contained in this data model, are assigned to object types



in the following way:

$$\begin{aligned} C_1 &: \text{Manufacturer} \\ &\mapsto \text{TOTAL} \{ \text{Manufacturer.builds} \} \\ C_2 &: \text{Airplane} \\ &\mapsto \text{UNIQUE} \{ \text{Airplane.has-as} \} \\ C_3 &: \text{Admission-code} \\ &\mapsto \text{TOTAL} \{ \text{Admission-code.given-to} \} \end{aligned}$$

All these constraints are owned by a single object type. A more interesting case with respect to inheritance results by adding the following constraint:

All airplanes must have associated a manufacturer or an age. Unregistered airplanes must have both.

The object type assignment for this constraint is:

$$\begin{aligned} C_4 &: \text{Airplane} \\ &\mapsto \text{TOTAL} \{ \text{Airplane.build-by}, \text{Airplane.has-as} \} \\ C_4 &: \text{Registered-airplane} \\ &\mapsto \text{TOTAL} \{ \text{Airplane.build-by}, \text{Airplane.has-as} \} \\ C_4 &: \text{Unregistered-airplane} \\ &\mapsto \text{TOTAL} \{ \text{Airplane.build-by} \} \text{ AND } \text{TOTAL} \{ \text{Airplane.has-as} \} \end{aligned}$$

The constraint  $C_4$  is considered to be owned by object types Airplane, Registered-airplane and Unregistered-airplane.

A constraint  $c$ , in an application model version, will therefore be a (usually very sparse) partial function  $c : \mathcal{O} \rightarrow \gamma$ , providing for every object type a *private* definition of the constraint. Each modelling technique will have its own possibilities to formulate inheritance rules, thus governing the mapping  $c$ . The set of constraints in an application model version at time  $t$  is denoted as  $\mathcal{R}_t$ .

#### 4.1.2. Methods

The action model part of an application model version will be provided as a set of action specifications. The set of possible action definitions ( $\mu$ ) depends on the chosen modelling technique for the action model. As for constraints, a modelling technique dependent inheritance hierarchy can be recognized. A method  $m$  is regarded as a partial function  $m : \mathcal{O} \rightarrow \mu$ , assigning action specifications to object types. The set of all methods present in an application model version is:  $\mathcal{M}_t$ .

#### 4.1.3. Domains

The separation between concrete and abstract world is provided by the distinction between the information structure  $\mathcal{I}$ , and the set of underlying (concrete) domains in  $\mathcal{D}$  ([HPW93]). Therefore, label types in an information structure version will have to be related to domains. An application model version contains a mapping  $\text{Dom}_t$  providing the relation between label types and domains. Some examples of such domain assignments, in the context of the rental

store, are:  $\text{Times} \mapsto \text{Natno}$ ,  $\text{Title} \mapsto \text{String}$ , where  $\text{Natno}$  and  $\text{String}$  are presumed to be (names of) concrete domains.

#### 4.1.4. Instances

The population of an information structure is not, as usual, a function that maps object types to sets of instances. Rather, an instance is considered to be an independent thing, which can evolve by itself. Therefore, (non empty) sets of object types are associated to instances, specifying the object types having the instance as an instantiation. This association is contained in the relation  $\text{HasTypes}_t \subseteq \Omega \times \wp(\mathcal{O})$ , where  $\Omega$  is the set of all possible instantiations of object types. An example of such an association is:  $\langle l_1, \{\text{Medium}, \text{Lp}\} \rangle$ , meaning  $l_1$  is an (abstract) instance of entity types  $\text{Medium}$  and  $\text{Lp}$ . The population of an object type, traditionally provided as a function  $p : \mathcal{O} \rightarrow \wp(\Omega)$ , can be derived from the association between instances and object types:  $\text{Pop}_t(x) = \{v \in \Omega \mid v \text{ HasTypes}_t Y \wedge x \in Y\}$ .

#### 4.1.5. Evolution dependency

Every method and constraint specification refers to (uses) a set of object types and denotable instances. This relation is provided in the application model universe by means of the dependency relation ( $\text{Depends}$ ). The relation  $\text{Depends}$  is modelling technique dependent, but is not subject to evolution.

The intuition behind this relation is as follows:  $x \text{ Depends } y$  means that if  $y$  is not present in an application model version, then  $x$  can not be used in that version. A consequence is, that in case of evolution of application models, when  $y$  evolves to  $y'$ , then  $x$  must be adapted appropriately.

As an example, consider the second action specification from the rental store example. This action specification depends on object types  $\text{Medium}$ ,  $\text{Lp}$  and  $\text{Frequency}$ . It, furthermore, depends on the domain assignment:  $\text{Frequency} \mapsto \text{Natno}$ . If one of the object types, or the domain assignment, is terminated or changed, the action specification has to be terminated or changed accordingly.

## 4.2. Application Model Versions

An application model version ( $\Sigma_t$ ) over information structure universe  $\mathcal{U}_{\mathcal{IS}}$  is determined by the following components:  $\mathcal{O}_t, \mathcal{R}_t, \mathcal{M}_t, \text{HasTypes}_t, \text{Dom}_t$ . With a version of the application model, we can associate the following version of the information structure:  $\mathcal{I}_t = \langle \mathcal{L}_t, \mathcal{N}_t, \sim_t, \rightsquigarrow_t \rangle$  where:

- $\mathcal{L}_t = \mathcal{O}_t \cap \mathcal{L}$ ,
- $\mathcal{N}_t = \mathcal{O}_t \cap \mathcal{N}$ ,
- $x \sim_t y \triangleq x \sim y \wedge x, y \in \mathcal{O}_t$ , and
- $x \rightsquigarrow_t y \triangleq x \rightsquigarrow y \wedge x, y \in \mathcal{O}_t$ .

As an overview of the components of an application model version, a meta model is provided in figure 5. This (meta) data model is conforming to the PSM data modelling technique, an extension of the NIAM modelling technique. The object types  $\mathcal{R}_t$  and  $\mathcal{M}_t$  in figure 5 are power types, the data modelling pendant of power sets in set theory.

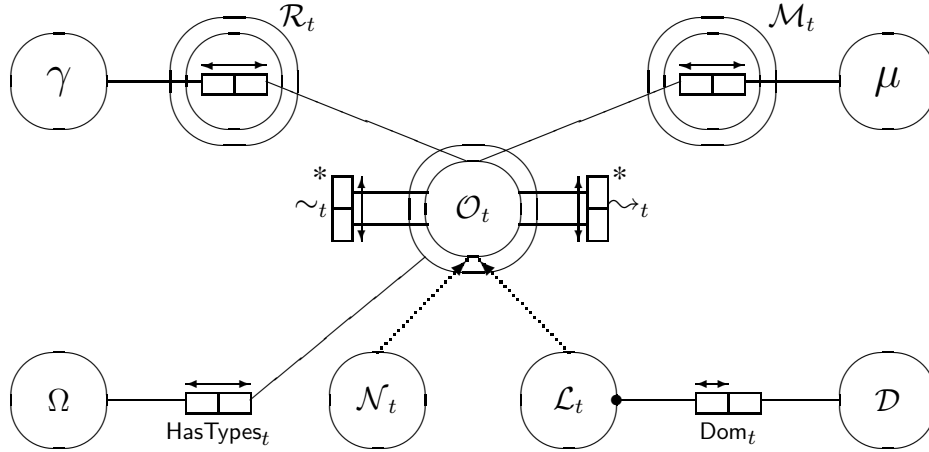


Figure 5: A meta model for information structures

Every application model version must adhere to some rules of wellformedness. Some of these rules are modelling technique dependent. Nonetheless, some general rules about application model versions can be stated. In [PW95] a formalisation of these rules is provided.

An object type  $x$  is called *alive* at a certain point of time  $t$ , if it is part of the application model version at that point of time ( $x \in \mathcal{O}_t$ ). Furthermore, an object type  $x$  is termed *active* at a certain point of time, if there is an instance with  $x$  as one of its associated types ( $v \text{ HasTypes}_t X \wedge x \in X$ ).

A first rule of wellformedness states that every active object type must be alive as well. This rule can be popularised as: ‘I am active, therefore I am alive’. Furthermore, the set of object types associated to an instance must be mutually type related, as their populations share at least this value.

From the very nature of the root relation it immediately follows that instances are included upwards, towards the roots. As a result, every instance of an object type is also an instance of its parents (if any). This rule has a structural pendant as well. Every living object type should be accompanied by one of its parents (if any). If all object types in a data model have a non empty instantiation, i.e. every object type is active, the latter (structural) rule immediately follows from the former (population) rule. However, when some non-root (alive) object type is not active, implying that its parent(s) do not have to be active, it does not follow that at least one of its parents must be alive. Therefore, a rule demanding this explicitly is required.

Constraints and methods are defined as mappings from object types to constraint and method definitions respectively. Thus, object types that own a constraint or a method must be alive. Furthermore, object types that own the same constraint or method, have to be type related. Finally, due to inheritance, if a constraint is defined for a parent object type, it must be defined for its children as well.

For populations some interesting properties have been proven [PW95]. An first example states: every instance of an object type is also an instance of one of its roots.

**Lemma 4.1**  $\text{Pop}_t(x) \subseteq \bigcup_{y \text{ RootOf } x} \text{Pop}_t(y)$

From the nature of type relatedness it follows that the populations of roots which are not type related are mutually exclusive. This rule can be generalised to all object types, leading to:

**Lemma 4.2**  $x \not\sim y \Rightarrow \forall_{t,u} [\text{Pop}_t(x) \cap \text{Pop}_u(y) = \emptyset]$

For the proofs of these properties, refer to [PW95].

## 5. Evolution of Application Models

The basis of the theory for evolving application models is formed by the concept of *evolution continuum*, capturing both dimensions of evolution of application models, being the universe of application model and time. The evolution continuum serves as a state space for evolutionary navigation:  $\mathcal{LS} = \langle \mathcal{T}_s, \mathcal{U}_\Sigma \rangle$

Time, essential to evolution, is incorporated into the evolution continuum through the algebraic structure  $\mathcal{T}_s = \langle \mathcal{T}, F \rangle$ , where  $\mathcal{T}$  is a (discrete) time axis, and  $F$  a set of functions over  $\mathcal{T}$ . In this paper, the time axis is regarded as an abstract data type. Several ways of defining a time axis exist, see e.g. [CR87], [WJL91] or [All84]. The time axis is the axis along which the application model evolves (through the application model universe).

An application model element history is a partial function, that assigns to points of time the actual version of the element, and thus is a partial mapping  $\mathcal{T} \mapsto \mathcal{AM}\mathcal{E}$ , where  $\mathcal{AM}\mathcal{E}$  is the set of all evolvable elements of an application model:  $\mathcal{O} \cup \mathcal{R} \cup \mathcal{M} \cup \text{HasTypes} \cup \text{Dom}$ . An application model history over an information structure universe  $\mathcal{U}_{\mathcal{IS}}$  then is defined by:

### Definition 5.1

$$\mathcal{AMH} \subseteq \mathcal{T} \mapsto \mathcal{AM}\mathcal{E} \quad \square$$

As a first rule of wellformedness, the evolution of application model elements is bound to classes. For example, an object type may not evolve into a method, and a constraint may not evolve into an instance. A consequence of this rule is that an application model history can be partitioned into the history of its object types, its constraints, its methods, its populations, and its concretisations (of label types):

## Definition 5.2

*object type histories:*

$$H_{type} = \{h \in \mathcal{AMH} \mid \exists_t[h(t) \in \mathcal{O}]\},$$

*constraint histories:*

$$H_{constr} = \{c \in \mathcal{AMH} \mid \exists_t[c(t) \in \mathcal{R}]\},$$

*method histories:*

$$H_{meth} = \{m \in \mathcal{AMH} \mid \exists_t[m(t) \in \mathcal{M}]\},$$

*population histories:*

$$H_{pop} = \{g \in \mathcal{AMH} \mid \exists_t[g(t) \in \text{HasTypes}]\},$$

*concretisation histories:*

$$H_{dom} = \{d \in \mathcal{AMH} \mid \exists_t[d(t) \in \text{Dom}]\}. \quad \square$$

An application model version  $\mathcal{I}_t = \langle \mathcal{L}_t, \mathcal{N}_t, \sim_t, \rightsquigarrow_t \rangle$  is easily derived from the application model history:

## Definition 5.3

*object types:*

$$\mathcal{O}_t = \{h(t) \mid h \in H_{type} \wedge h \downarrow t\},$$

*constraints:*

$$\mathcal{R}_t = \{c(t) \mid c \in H_{constr} \wedge c \downarrow t\},$$

*methods:*

$$\mathcal{M}_t = \{m(t) \mid m \in H_{meth} \wedge m \downarrow t\},$$

*population:*

$$\text{HasTypes}_t = \{g(t) \mid g \in H_{pop} \wedge g \downarrow t\},$$

*concretisations:*

$$\text{Dom}_t = \{d(t) \mid d \in H_{dom} \wedge d \downarrow t\}. \quad \square$$

In this definition  $f \downarrow t$  is used as an abbreviation for  $\exists_s [(t, s) \in f]$ , stating that (partial) function  $f$  is defined at time  $t$ . As an example of evolution, the following table represents three object type evolutions ( $h_1, h_2, h_3 \in H_{type}$ ) from the rental store (see section 2):

$H_{type}$	$h_1$	$h_2$	$h_3$
$t_1$	$o_1$	$o_2$	—
$t_2$	$o_1$	$o_2$	—
$t_3$	$o_1$	$o_2$	—
$t_4$	$o_1$	$o_3$	$o_4$
$t_5$	$o_1$	$o_3$	$o_4$

where  $t_1, \dots, t_5 \in \mathcal{T}$ , and  $o_1, \dots, o_4 \in \mathcal{O}$  are object types, such that:

- $o_1$  = ‘Entity type: Record’
- $o_2$  = ‘Fact type: Recording of Song on Record’
- $o_3$  = ‘Fact type: Recording of Song on Medium’
- $o_4$  = ‘Entity type: Medium’

Note that the evolution step (from figure 2 to figure 3) takes place at point of time  $t_4$ . Two example instance evolutions ( $g_1, g_2 \in H_{pop}$ ), obeying the above schema evolution, are:

$H_{pop}$	$g_1$	$g_2$
$t_1$	$\langle i_1, \{o_1\} \rangle$	$\langle i_2, \{o_2\} \rangle$
$t_2$	$\langle i_1, \{o_1\} \rangle$	$\langle i_2, \{o_2\} \rangle$
$t_3$	$\langle i_1, \{o_1\} \rangle$	$\langle i_3, \{o_2\} \rangle$
$t_4$	$\langle i_1, \{o_1, o_4\} \rangle$	$\langle i_4, \{o_3\} \rangle$
$t_5$	$\langle i_1, \{o_1, o_4\} \rangle$	$\langle i_4, \{o_3\} \rangle$

where  $i_1, \dots, i_4 \in \Omega_\infty$  are instances such that:

$$\begin{aligned}
 i_1 &= \text{'Brothers in Arms'} \\
 i_2 &= \langle \text{'Money for nothing'}, \text{'Brothers in Arms'} \rangle \\
 i_3 &= \langle \text{'Brothers in Arms'}, \text{'Brothers in Arms'} \rangle \\
 i_4 &= \langle \text{'Brothers in Arms'}, \text{'Brothers in Arms'} \rangle
 \end{aligned}$$

The interpretation of this table leads to:

$$\begin{aligned}
 g_1(t_4) = \langle i_1, \{o_1, o_4\} \rangle &\text{ means: 'Brothers in Arms' is both a Record and a Lp at } t_4, \\
 g_2(t_3) = \langle i_3, \{o_2\} \rangle &\text{ means: Song 'Brothers in Arms' is Recorded on Record 'Brothers in Arms',} \\
 g_2(t_4) = \langle i_4, \{o_3\} \rangle &\text{ means: Song 'Brothers in Arms' is Recorded on Medium 'Brothers in Arms'}.
 \end{aligned}$$

As an outline of the hitherto defined concepts, a (meta) data model, relating all defined concepts, is provided in figure 6. The data model depicted there is conform the PSM modelling technique, and uses the notion of schema objectifications (object type  $\mathcal{AMH}$ ), and powertyping (object type  $\wp(\mathcal{O})$ ). The population of an objectified schema at hand is to be looked upon as one single abstract object instance of the object type corresponding to the objectified schema. Powertyping is, as stated before, the data modelling pendant of powersets from set theory.

## 6. Conclusions and Further research

In this paper we presented a first attempt to a general theory for the evolution of application models, supporting evolving information systems. As a next step a set of well formedness rules on evolution steps has to be defined, since not all evolution steps of application models will correspond to sensible/valid evolution steps in an actual universe of discourse.

Furthermore, in order to validate the theory, it must be applied to some modelling techniques. In the near future we will therefore apply this theory to PSM, a data modelling technique serving as a common base for data modelling techniques such as NIAM, ER, and IFO, and to Task Structures, a powerfull action modelling technique. A query and manipulation language must be defined supporting the evolution of information systems. Finally, the consequences of evolution for the internal representation of information structures should be studied.

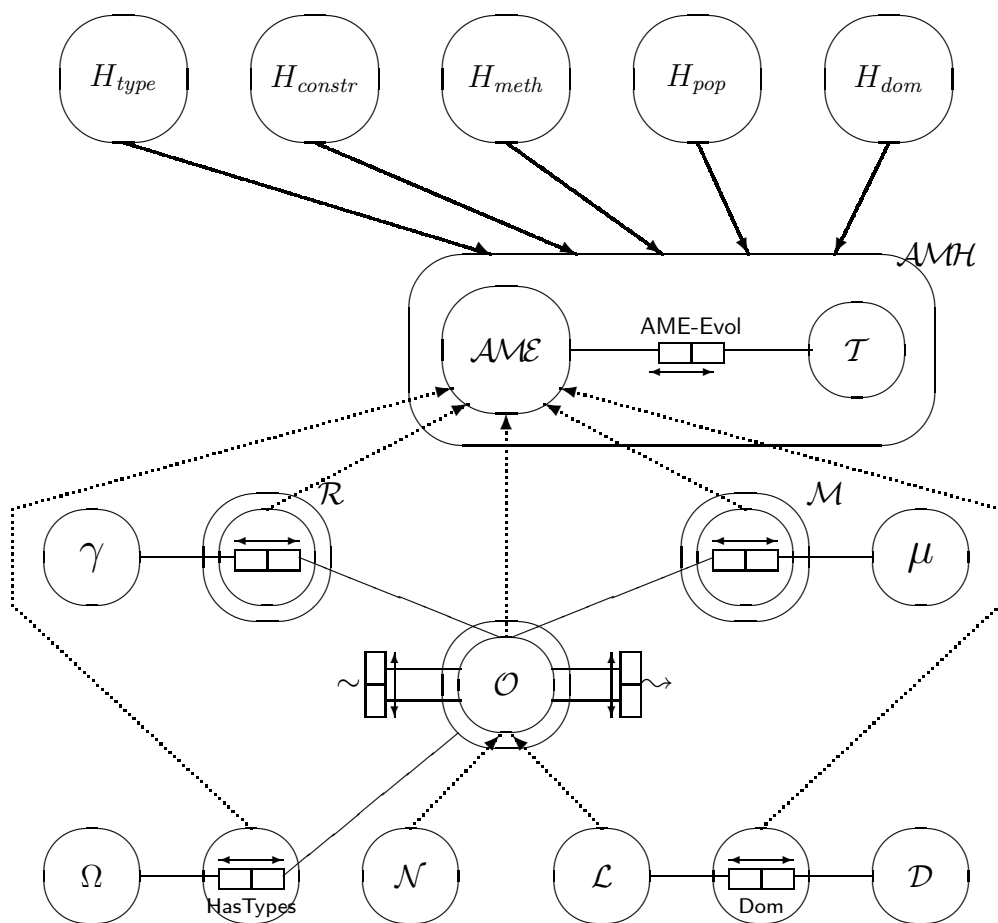


Figure 6: A meta model for the evolution system

## Acknowledgements

The investigations were partly supported by the Foundation for Computer Science in the Netherlands (SION) with financial support from the Netherlands Organization for Scientific Research (NWO).

## References

- [All84] J.F. Allen. Towards a General Theory of Action and Time. *Artificial Intelligence*, 1984(23):123–154, 1984.
- [BW89] P. D. Bruza and Th.P. van der Weide. The semantics of data flow diagrams. In N. Prakash, editor, *Proceedings of the International Conference on Management of Data (CISMOD)*, pages 66–78, Hyderabad, India, 1989. McGraw-Hill Publishing Company.

- [Che76] P.P. Chen. The entity-relationship model: Towards a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, March 1976.
- [CR87] J. Clifford and A. Rao. A simple, general structure for Temporal Domains. In C. Rolland, F. Bodart, and M. Leonard, editors, *Temporal Aspects in information Systems*, pages 17–28. North-Holland/IFIP, Amsterdam, The Netherlands, 1987.
- [De 91] O.M.F. De Troyer. The OO-Binary Relationship Model: A Truly Object Oriented Conceptual Model. In R. Andersen, J.A. Bubenko, and A. Sølvsberg, editors, *Proceedings of the Third International Conference CAiSE'91 on Advanced Information Systems Engineering*, volume 498 of *Lecture Notes in Computer Science*, pages 561–578, Trondheim, Norway, May 1991. Springer-Verlag.
- [FOP92a] E.D. Falkenberg, J.L.H. Oei, and H.A. Proper. A Conceptual Framework for Evolving Information Systems. In H.G. Sol and R.L. Crosslin, editors, *Dynamic Modelling of Information Systems II*, pages 353–375. North-Holland, Amsterdam, The Netherlands, EU, 1992. ISBN 0444894055
- [FOP92b] E.D. Falkenberg, J.L.H. Oei, and H.A. Proper. Evolving Information Systems: Beyond Temporal Information Systems. In A.M. Tjoa and I. Ramos, editors, *Proceedings of the Data Base and Expert System Applications Conference (DEXA'92)*, pages 282–287, Valencia, Spain, EU, September 1992. Springer Verlag, Berlin, Germany, EU. ISBN 3211824006
- [HPW92] A.H.M. ter Hofstede, H.A. Proper, and Th.P. van der Weide. Data Modelling in Complex Application Domains. In P. Loucopoulos, editor, *Proceedings of the Fourth International Conference CAiSE'92 on Advanced Information Systems Engineering*, volume 593 of *Lecture Notes in Computer Science*, pages 364–377, Manchester, United Kingdom, EU, May 1992. Springer Verlag, Berlin, Germany, EU. ISBN 3540554815
- [HPW93] A.H.M. ter Hofstede, H.A. Proper, and Th.P. van der Weide. Formal definition of a conceptual language for the description and manipulation of information models. *Information Systems*, 18(7):489–523, October 1993.
- [HSV89] K.M. van Hee, L.J. Somers, and M. Voorhoeve. Executable Specifications for Distributed Information Systems. In E.D. Falkenberg and P. Lindgreen, editors, *Information System Concepts: An In-depth Analysis*, pages 139–156. North-Holland/IFIP, Amsterdam, The Netherlands, 1989.
- [HW93] A.H.M. ter Hofstede and Th.P. van der Weide. Expressiveness in conceptual data modelling. *Data & Knowledge Engineering*, 10(1):65–100, February 1993.
- [ISO87] *Information processing systems – Concepts and Terminology for the Conceptual Schema and the Information Base*, 1987. ISO/TR 9007:1987.  
<http://www.iso.org>
- [JMSV92] M. Jarke, J. Mylopoulos, J.W. Schmidt, and Y. Vassiliou. DAIDA: An Environment for Evolving Information Systems. *ACM Transactions on Information Systems*, 20(1):1–50, January 1992.
- [Kat90] R.H. Katz. Toward a Unified Framework for Version Modelling in Engineering Databases. *ACM Computing Surveys*, 22(4):375–408, 1990.



- [MS90] E. McKenzie and R. Snodgrass. Schema evolution and the relational algebra. *Information Systems*, 15(2):207–232, 1990.
- [NH89] G.M. Nijssen and T.A. Halpin. *Conceptual Schema and Relational Database Design: a fact oriented approach*. Prentice-Hall, Sydney, Australia, 1989. ASIN 0131672630
- [PW95] H.A. Proper and Th.P. van der Weide. A General Theory for the Evolution of Application Models. *IEEE Transactions on Knowledge and Data Engineering*, 7(6):984–996, December 1995.
- [Rod91] J.F. Roddick. Dynamically changing schemas within database models. *The Australian Computer Journal*, 23(3):105–109, August 1991.
- [SA85] R. Snodgrass and I. Ahn. A Taxonomy of Time in Databases. In *Proceedings of the ACM SIGMOD International Conference on the Management of Data*, pages 236–246, Austin, Texas, 1985.
- [WHO92] G.M. Wijers, A.H.M. ter Hofstede, and N.E. van Oosterom. Representation of Information Modelling Knowledge. In V.-P. Tahvanainen and K. Lyytinen, editors, *Next Generation CASE Tools*, volume 3 of *Studies in Computer and Communication Systems*, pages 167–223. IOS Press, 1992.
- [WJL91] G. Wiederhold, S. Jajodia, and W. Litwin. Dealing with the Granularity of Time in Temporal Databases. In R. Andersen, J.A. Bubenko, and A. Sølvsberg, editors, *Proceedings of the Third International Conference CAiSE'91 on Advanced Information Systems Engineering*, volume 498 of *Lecture Notes in Computer Science*, pages 124–140, Trondheim, Norway, May 1991. Springer-Verlag.
- [WMP<sup>+</sup>76] A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.T. Meertens, and R.G. Fisker. *Revised Report on the Algorithmic Language ALGOL 68*. Springer-Verlag, Berlin, Germany, 1976.
- [You89] E. Yourdon. *Modern Structured Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1989.