A Framework for the Evolution of Temporal Conceptual Schemas of Information Systems

Juan-Ramón López[1,2] and Antoni Olivé[1]

[1]Universitat Politècnica de Catalunya,
Departament de Llenguatges i Sistemes Informàtics.
Barcelona - Catalonia.
{jrlopez|olive}@lsi.upc.es
[2]Universidade da Coruña,
Departamento de Computación

Abstract. This paper focuses on the problem of information systems evolution. Ideally, changes in the requirements of information systems should be defined and managed at the conceptual schema level, with an automatic propagation down to the logical database schema(s) and application programs. We propose here a framework for the evolution of temporal conceptual schemas of information systems. Our framework uses a reflective architecture with two levels: meta schema and schema, and two loosely coupled information processors, one for each level. We define a temporal minimal meta schema, and we show, using some examples, how to extend this minimal meta schema to support any usual conceptual modeling construct. We also show how the framework can be used to specify schema changes.

1 Introduction

This paper deals with the evolution of temporal conceptual schemas of information systems. We propose a general framework which allows defining schema evolutions and their effects. We illustrate the framework with a limited albeit representative number of widely used conceptual modeling constructs, define some evolution operations at the meta schema level, and analyze their impact on the conceptual schema and the information base.

In order to characterize our contribution, we find it useful to start with the abstract database evolution framework presented in [11], that we rephrase as follows: Assume that an information system satisfies requirements R0. The information system comprises a conceptual schema CS0, one (or more) database(s) with logical schema(s) LS0 and extension(s) E0, and several application programs P0. When requirements change to R1, the information system must evolve. In the general case, the evolution implies changes leading to a new conceptual schema CS1, new logical schema(s) LS1 and application programs P1. Moreover, the extension(s) E0 may not be valid, and need to be converted to E1.

Ideally, the evolution of information systems should follow the strategy of "forward information system maintenance" [11]: changes should be applied to the conceptual schema, and from here they should propagate automatically down to the logical schema(s) and application programs. If needed, the extension(s) should be also converted automatically.

In the context of the above framework, the aspect that has received more research effort is (logical) database schema evolution, including the propagation of changes to the extensional database. Data models that have been studied extensively include the relational and object-oriented models, both non-temporal and temporal. In several cases, research results have been incorporated into commercial or prototype database management systems (e.g., Orion [4], O_2 [27], F2 [1]). We refer to [19, 8, 24, 15] for extensive bibliographies and surveys.

We observe that the ideal strategy described above implies that the conceptual schema is the only description to be defined, and the basis for the specification of the evolution. This observation has led us to our framework, which we see as one step more towards that ideal strategy. Our framework is based on the commonly accepted approach that, from a conceptual point of view, an information system consists of a conceptual schema, an information base and an information processor. The conceptual schema defines all general domain knowledge relevant to the system, which includes entity and relationship types, all kind of integrity constraints and derivation rules, and the effects of external events. The information processor receives messages from the environment, reporting the occurrence of external events. In response to such events, the information processor changes the information base and/or issues messages making known the contents of the information base, in the way specified in the conceptual schema [12].

In this view, evolution can be accommodated elegantly by using the classical reflective architecture [18, 22], as shown in Figure 1. A distinction is made between an information system and a meta information system. The first consists of a conceptual schema, an information base (IB) and an information processor (IP). The second consists of a conceptual meta schema, a meta information base (MIB) and a meta information processor (MIP). In essence, the MIB of the meta information system describes the conceptual schema of the information system [12].

In our framework, possible changes in the conceptual schema are specified as meta external events in the conceptual meta schema. Occurrences of these events are reported to the MIP, which changes the MIB and indirectly the conceptual schema. In some cases, changes in the IB may be necessary too. These latter changes are not performed directly by the MIP. Instead of this, the MIP generates occurrences of external events, which are notified to the IP to perform the desired changes. In this way, the two processors are loosely coupled.

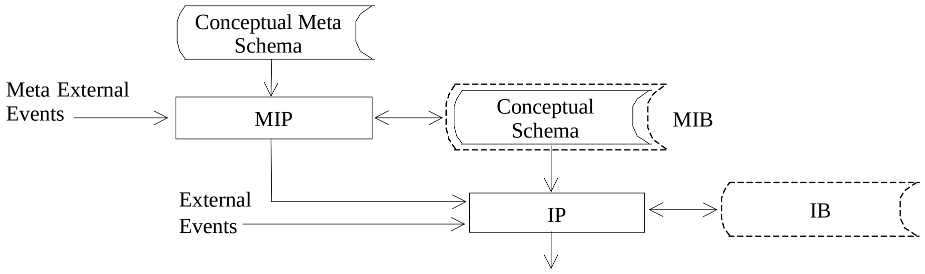

Fig. 1. Framework for the evolution of information systems

There is a lot of previous research and development work relevant to this framework. We can only give in the following a few comments relating a representative part of this work with our paper.

[14] describes SEA, Schema Evolution Assistant, a tool which allows the evolution of schemas in Chimera. Chimera is a rich data model that includes integrity constraints, derivation rules and triggers. SEA uses a reflective architecture [16], as we do. Initially, Chimera was non-temporal, but a temporal extension was published later [5]. SEA was developed for the non-temporal version.

[13] describes ConceptBase, a system based on the O-Telos language, a variant of Telos. O-Telos deals uniformly with objects at different levels of abstraction. O-Telos has a kernel, from which all other constructs can be bootstrapped. We have followed the same idea with our minimal meta schema. However, our framework allows that the information and the meta information processors be different. On the other hand, we deal with the temporal dimension and focus more on schema evolution.

Finally, we mention the work done in the context of the TIGUKAT temporal ODBMS [22, 23, 10]. The TIGUKAT schema objects include types, classes, behaviours, functions and collections. The model allows the representation of the temporal histories of real-world objects. We also adopt the temporal view, but conceptual schemas of information systems need to include other elements like integrity constraints, derivation rules, external events, etc. Evolution is performed by operations (add, drop and modify) on schema objects. These are similar to our meta external events, but we need many more to allow the evolution of other schema objects.

The remainder of this paper is organized as follows. In Section 2, we characterize temporal conceptual schemas and evolution at the temporal IB level. In Section 3 we do the same for the conceptual meta schema and MIB, and define the correspondence between the MIB and the conceptual schema. We also show how usual conceptual modeling constructs can be defined at the conceptual meta schema level. In Section 4 we then define the meta external events that perform schema evolution. Finally, Section 5 gives the conclusions and points out future work.

2 Temporal Conceptual Schema and Information Base

In this section, we briefly characterize the elements of a temporal conceptual schema (or, for short, schema). We wish our work to be of general applicability and language-independent. For this reason, we focus here on a set of minimal core elements of schemas. Any schema can be expressed in terms of these elements. They are entity types, relationship types, derivation rules, integrity constraints and external event types [3]. We represent them (and their instances in the IB) using the language of the first order logic (FOL), with some common extensions.

2.1 Temporal Conceptual Schema

The entities of a domain are instance of *entity types* at particular time points [7]. An entity may be instance of several entity types at the same time (multiple classification). On the other hand, an entity may change the entity types of which it is instance (dynamic classification).

In a temporal conceptual schema, entity types may be represented by binary predicates, whose first term is a symbol that denotes an entity in the domain, and the second a time point. In the IB, we represent by *E(e,t)* the fact that entity *e* is instance of entity type *E* at time *t*.

We assume that time is discrete, linear and totally ordered. Time points are expressed in a common base time unit (or *chronon* [26]), such as second or day. We also assume that a schema includes a special unary predicate *Time*, whose instances are the time points (in the chosen unit) of the information system lifespan. In general, the set of entity types defined in a schema is time-varying. We denote by E_t the set of entity types defined in a schema at time *t*. An entity type may cease to exist in a schema and appear again later. We adopt here the usual *typing rule*, which in our case requires that, for all entity types *E*:

$$\forall e,t\ (E(e,t) \rightarrow E \in E_t)\ . \qquad \mathbf{(1)}$$

That is, we do not allow the IB to represent that *e* is instance of *E* at *t*, if *E* is not defined in the schema at *t*. For convenience, we assume that if an entity type *E* ceases to exist at time *t*, then all its instances at *t-1* (if any) also cease to exist at *t*.

Entity types may be *lexical* or *non-lexical*. For the sake of uniformity we represent lexical entity types also by binary predicates, even if lexical entities are usually instance of their types during the whole lifespan of an information system.

The relationships of a domain are instance of *relationship types* at particular time points. In a temporal schema, a relationship type with *n participants* may be represented by a predicate of degree *n+1*, where the first *n* terms are symbols that denote the participant entities, and the *n+1* term is a time point. In the IB, we represent by $R(e_1,...,e_n,t)$ the fact that entities $e_1,...,e_n$ participate in a relationship instance of relationship type *R* at time *t*. In general, the set of relationship types defined in a schema is time-varying, and we denote by R_t the set of relationship types defined in a schema at time *t*. The above typing rule also applies to relationship types.

Entity and relationship types may be *base* or *derived*. Instances of base types must be stated explicitly, as will be seen later, while those of derived types are defined by *derivation rules*. In a temporal schema, derivation rules must be considered also as time-varying. We denote by DR_t the set of derivation rules defined in a schema at time *t*. Our framework allows derivation rules to be written in the language understood by the IP. In the examples we use the FOL language.

Integrity constraints are conditions that the IB must satisfy. We consider them also as time-varying. We denote by IC_t the set of integrity constraints defined in a schema at time *t*. It is assumed that at any time *t* the IB satisfies all integrity constraints defined in IC_t.

Our framework could be extended by distinguishing, in the schema, several kinds of integrity constraints (and of derivation rules). However, in this paper we prefer not to make this distinction here, but we do it in the conceptual meta schema. The MIP will "generate" constraints in the language required by the IP.

The IB changes due to the occurrence of *external events*. As we did for entity types, we represent external event types in a schema by binary predicates, but now the first term denotes an event occurrence. In the IB we represent by *Ev(ev,t)* the occurrence of event *ev* of type *Ev* at time *t*.

In general, the set of external event types defined in a schema is time-varying, and we denote by EE_t the set of external event types defined in a schema at time *t*. The above typing rule also applies to these types. External event types may be base or derived.

For each external event type, the schema includes the definition of its preconditions and effect rules (transactions). The *preconditions* are conditions that must be satisfied when an external event occurs. The *effect rules* define the changes produced in the IB by the event. Both preconditions and effect rules are formulas expressed in the language understood by the IP.

We find convenient to allow the possibility of defining, for each integrity constraint, a set of *compensating* effect rules. The IP adds these rules to the effect rules of an event type when an occurrence of this event type leads to a violation of the integrity constraint. This, however, does not mean that violations can be prevented in all cases by compensating rules.

2.2 Information Base Evolution

The effect rules of an external event type define the changes in the IB when an instance of that type occurs. We model possible changes in the IB as structural events, and we then say that an external event induces structural events, which, in turn, change the IB.

The *structural event types* are determined by the entity and relationship types defined in a conceptual schema and, thus, they are not defined by the designer. We represent structural event types by predicates, as follows:

- There is an insertion and a deletion event type for each non-lexical entity type *E*. We represent them by the binary predicates *Ins_E* and *Del_E,* respectively, where the first term is a symbol denoting an entity and the second a time point.
- There is an insertion and a deletion event type for each relationship type *R*. We represent them by the n+1-ary predicates *Ins_R* and *Del_R*, respectively, where the first n terms are symbols denoting entities and the n+1 term is a time point.

A structural event type is base (derived) if the corresponding entity or relationship type is base (derived). External events induce directly only base structural events, while derived ones are induced indirectly by the derivation rules. For example, if we have in a schema the derivation rule: $Young(p,t) \leftarrow Person(p,t) \land Age(p,a,t) \land a \leq 18$, then an increase of a person’s age may induce a *Del_Young* event.

3 Temporal Conceptual Meta Schema and Information Base

All conceptual models have a conceptual meta schema (or, for short, meta schema). A meta schema is like an ordinary schema, except that its domain is a schema. The elements of a schema are seen as entities, which are instance of appropriate entity types defined in the meta schema. Similarly, relationships between the elements of a schema are seen as instances of relationship types of the meta schema. The MIB has a representation of the schema.

In this section, we first describe the minimal meta schema necessary to represent the core elements defined in the previous section. We then establish the correspondence rules between the MIB and the schema. In the last part of this section we describe how the meta schema can be enriched with other constructs.

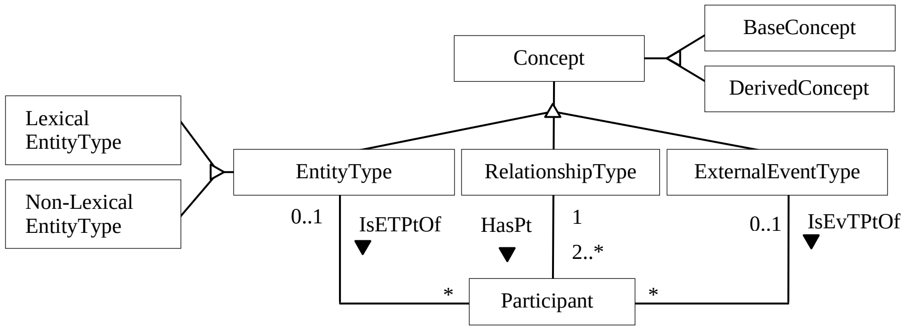

Fig. 2. Minimal conceptual meta schema (Part 1 of 2)

3.1 Minimal Meta Schema

Figures 2 and 3 illustrate, using the UML language [25], the minimal meta schema.[1] We now briefly define it in terms of the core elements explained in the previous section. For clarity, we call sometimes these elements meta elements, in order to distinguish them from schema elements.

– *Entity types.* The entity types are shown as rectangles in Figures 2 and 3. All entity types are non-lexical, except *Formula*, which is lexical. An instance of *Formula* is a formula in the language understood by the IP. For generality, we use here the FOL language. All entity types are base, except *Concept*, *EntityType* and *Rule*, which are derived. As we did in Section 2.1, in the MIB we represent by *E(e,t)* the fact that *e* is instance of meta entity type *E* at time *t*. For example, *EntityType(Employee,T)*.
– *Relationship types.* The meta relationship types are shown as lines linking rectangles in Figures 2 and 3. All of them are binary and base. Note that a schema relationship type *R* may be of degree 2 or higher. *R* would be an instance of *RelationshipType* at *T* (Figure 2) and, if it has degree *n*, there would be *n* instances $HasPt(R,P_i,T)$ in the MIB.
– *Structural events.* Meta structural events are defined as we did in Section 2.2. For clarity, we will prefix them with *M_*. For example, *M_Del_EntityType*, whose intended effect is that an entity type ceases to be instance of *EntityType* in the MIB.
– *Derivation rules.* There are three meta derivation rules, which define *Concept*, *EntityType* and *Rule*. We show here the one corresponding to *Concept*:

[1] Note that, in the figures, we use partitions and cardinality constraints. However, these constructs will be translated into core elements in what follows.

$\text{Concept}(c,t) \leftrightarrow \text{EntityType}(c,t) \vee \text{RelationshipType}(c,t) \vee \text{ExternalEventType}(c,t)$ **MDR1**

– *Integrity constraints.* A schema is a complex structure subject to a large number of integrity constraints. This implies that the meta schema must include many meta integrity constraints, usually called *invariants* in the schema evolution field [4]. We give bellow an example, that states that only existing rules and formulas can be related in *IsExpressedBy*. Other examples can be found in [17].

$$\text{IsExpressedBy}(r,f,t) \rightarrow \text{Rule}(r,t) \wedge \text{Formula}(f,t) \qquad \textbf{MIC1}$$

We define a compensating rule for MIC1, which states that when an rule is deleted so is the relationship, instance of *IsExpressedBy*, in which it participates:

$$\text{M\_Del\_IsExpressedBy}(r,f,t) \leftarrow \text{M\_Del\_Rule}(r,t) \wedge \text{IsExpressedBy}(r,f,t\text{-}1)$$

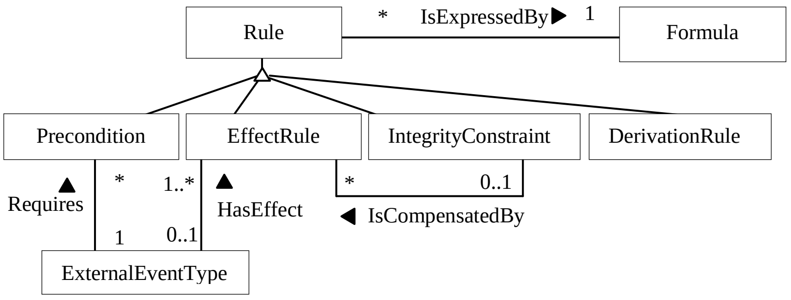

Fig. 3. Minimal conceptual meta schema (Part 2 of 2)

– *External event types.* The minimal meta schema includes meta event types that allow the insertion and deletion of the entity and relationship types shown in Figures 2 and 3. For example, there could be an external event type *NewIC*. Occurrences of *NewIC* would create an instance of *IntegrityConstraint* (Figure 3) and the associated instance of *IsExpressedBy*. We will discuss them in the next section.

3.2 Correspondence Rules

We can now define the correspondence rules in our framework between the MIB and the schema:

- *CR1*. There is a one-to-one correspondence between the instances at time *t* of *EntityType* in the MIB and the elements of E_t in the schema. If an instance of *EntityType* is also instance of *BaseConcept* (*DerivedConcept*) then it is base (resp., derived). If an instance of *EntityType* is also instance of *LexicalEntityType* (*Non-LexicalEntityType*) then it is lexical (resp., non-lexical).
- *CR2*. There is a one-to-one correspondence between the instances at time *t* of *RelationshipType* in the MIB and the elements of R_t in the schema. If an instance of *RelationshipType* is also instance of *BaseConcept* (*DerivedConcept*) then it is base

(resp., derived). If an instance *R* of *RelationshipType* is related (through *HasPt*) to *n* instances of *Participant*, then the degree of *R* in the schema is *n+1*.

- *CR3*. There is a one-to-one correspondence between the set of *Formula* in which *IsExpressedBy* the instances of *DerivationRule* at time *t* in the MIB and the elements of DR_t in the schema.
- *CR4*. There is a one-to-one correspondence between the set of *Formula* in which *IsExpressedBy* the instances of *IntegrityConstraint* at time *t* in the MIB and the elements of IC_t in the schema. If *Ic* is an instance of *IntegrityConstraint*, and *IsCompensatedBy(Ic,Eff,t)* and *IsExpressedBy(Eff,F,t)* then formula *F* is a compensating effect rule of *Ic* in the schema at *t*.
- *CR5*. There is a one-to-one correspondence between the instances at time *t* of *ExternalEventType* in the MIB and the elements of EE_t in the schema. If an instance of *ExternalEventType* is also instance of *BaseConcept* (*DerivedConcept*) then it is base (resp., derived). If *Ev* is an instance of *ExternalEventType,* and *Requires(Ev,Pre,t)* and *IsExpressedBy(Pre,F,t)*, then formula *F* is a precondition of *Ev* in the schema at *t*. Similarly, if *Ev* is an instance of *ExternalEventType*, and *HasEffect(Ev,Eff,t)* and *IsExpressedBy(Eff,F,t)*, then formula *F* is an effect rule of *Ev* in the schema at *t*.

Correspondence rules similar to the above are present in a way or another in all frameworks with a reflective architecture. We emphasize the distinction between the MIP and the IP, make explicit the temporal correspondence, and leave undefined how the correspondence rules are implemented.

3.3 Conceptual Modeling Constructs

The above minimal meta schema is hardly satisfactory because it does not take into account the important number of special constructs that have been developed in the conceptual modeling field. These constructs are useful because they ease the definition of the schema and, at the same time, allow the development of efficient implementations. In our framework, the modeling constructs are particular types in the meta schema. However, there is not a direct correspondence between the "instantiation" of these constructs and the schema, but an indirect one. The meta schema includes a set of derivation rules that "translate" the constructs into elements of the minimal meta schema. The above correspondence rules need not to be changed.

In what follows we explain how to extend the minimal meta schema to incorporate the constructs in our framework. Additional details can be found in [17]. The general idea is to partition *IntegrityConstraint* and *DerivationRule* (Figure 3) into two subtypes: a base and a derived one (Figure 4). The base subtypes will correspond to integrity constraints or derivation rules defined explicitly by the designer, while the derived ones are defined by derivation rules of the meta schema.

3.3.1 Cardinality Constraints. Figure 4 shows the conceptualization in the meta schema of the common cardinality constraints. In the MIB, these cardinalities are expressed as instances of relationship types *HasMinCard* and *HasMaxCard*. We will focus only on the minimum cardinality; the maximum one is similar. Each minimum cardinality constraint is an instance of *ICMinCard*, which is a specialization of

DerivedIntegrityContraint. A participant may not have minimum cardinality (i.e., zero) and then there is not a corresponding instance of *ICMinCard*. Entity type *ICMinCard* is derived. The corresponding derivation rule is:

$$\text{ICMinCard(s,t)} \leftarrow \text{HasMinCard(p,int,t)} \wedge \text{s = sym(ICMinCard,p)} \qquad \textbf{MDR2}$$

where *sym(ICMinCard,p)* is a function that denotes a time-independent symbol (*s*), distinct from any other used in the MIB. The MIP could, for example, create this symbol when symbol *p* is created.

Now, we need to define the *Formula* in which the constraint *IsExpressedBy*. We refine the base relationship type between *Rule* and *Formula* to a derived one between *ICMinCard* and *Formula*. The derivation rule is:

$$\text{IsExpressedBy(s,f,t)} \leftarrow \text{ICMinCard(s,t)} \wedge \text{ICMinFormula(s,f,t)} \qquad \textbf{MDR3}$$

In MDR3, *ICMinFormula(s,f,t)* is a predicate that gives in *f* the string representation of the corresponding formula. For a binary relationship type *R*, the general form of *f* for one of its participants (E_1) is "$E_1\ (e_1,t) \rightarrow min \leq |\{e_2 \mid R(e_1,e_2,t)\}|$", where E_1 would be given by *IsETPtOf* (or *IsEvTPtOf*), *min* by *HasMinCard* and *R* by *HasPt*. We omit here the details of predicate *ICMinFormula* because they are a lengthy sequence of trivial string manipulation operations.

Note that, by MDR2 and MDR3, the MIB will have a relationship *IsExpressedBy(s,f,t)* during the period in which *p* exists as participant and has a non-zero minimum cardinality. By CR4, during this period $f \in IC_t$, that is, *f* will be a schema integrity constraint, to be enforced by the IP.

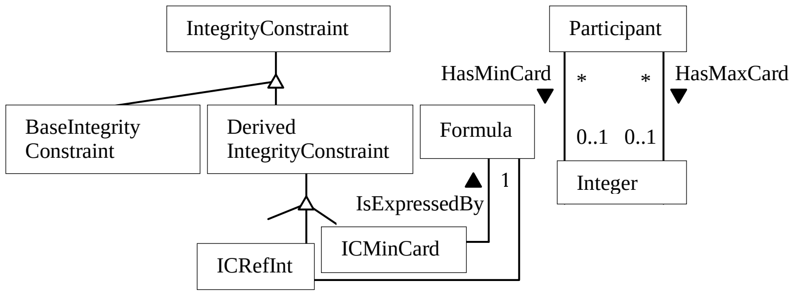

Fig. 4. Meta schema conceptualization of cardinality and referential integrity constraints

3.3.2 Referential Integrity Constraints. A referential integrity constraint must be defined for each participant of a base relationship type. We assume that for derived relationship types the corresponding derivation rule is correct and, therefore, also are the types of the participants in its instances. To represent these constraints, we define in the meta schema (see Figure 4) a derived subtype of *DerivedIntegrityConstraint*, called *ICRefInt*, with the following derivation rule:

$$\text{ICRefInt(s,t)} \leftarrow \text{HasPt(r,p,t)} \wedge \text{BaseConcept(r,t)} \wedge \text{s=sym(ICRefInt,p)} \qquad \textbf{MDR4}$$

where *sym* is again a function that gives us a time-independent symbol, that in this case denotes the referential integrity constraint related to a participant. We refine again *IsExpressedBy* to a derived type, in this case between *ICRefInt* and *Formula*:

$$\text{IsExpressedBy(s,f,t)} \leftarrow \text{ICRefInt(s,t)} \wedge \text{ICRefIntFormula(s,f,t)} \qquad \textbf{MDR5}$$

where *ICRefIntFormula* is a predicate that gives in *f* the string representation of the corresponding formula.[2] For a binary relationship type *R*, the general form of *f* for one of its participants (E_1) is „$R(e_1,e_2,t) \rightarrow E_1(e_1,t)$„, where *R* would be given by *HasPt*, and E_1 by *IsETPtOf* (or *IsEvTPtOf*).

3.3.3 Partitions. Partitions are an important modeling construct in many conceptual models [20]. All generalizations and specializations can be transformed into partitions. We are going to see that partitions are represented in the core elements of a schema by zero or one derivation rule, and zero or more integrity constraints. Figure 5 shows the conceptualization of partitions in the meta schema.

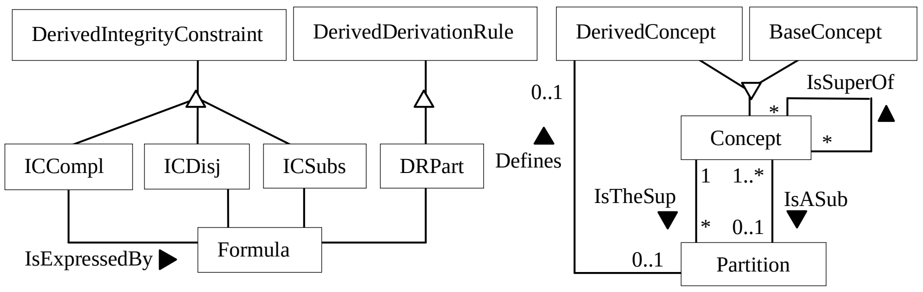

Fig. 5. Meta schema conceptualization of partitions

We allow the partition of any *Concept*. Thus, we deal also with partitions of relationship types [2]. A concept may be the superconcept (*IsTheSup*) of several partitions. A partition has only one superconcept, but it may have one or more subconcepts. We assume that a concept may be the subconcept (*IsASub*) of zero or only one partition. The concepts involved in a partition may be base or derived. Almost all combinations can be acceptable. One of the exceptions is when all concepts are base, because if all subconcepts are base, then the superconcept must be derived. For example, if P_1 is a partition with super *Person* and subs *Man* and *Woman*, and these are base, then *Person* must be derived, since it is the "union" of *Man* and *Woman*. In the meta schema, there is a relationship type (*Defines*) between *Partition* and *DerivedConcept*. An instance *Defines(P,C,T)*, where *C* must be a derived concept and the super of *P*, indicates that, at time *T*, concept *C* is derived as the union of the subconcepts of *P*. In the above example, we could have *Defines(P_1,Person,T)*. Entity type *Person* may be the super of another partition P_2 (for example, into the base entity

[2] This formula depends on the relationship type of the participant being *synchronous* or *asynchronous* (see [21] for more details), but for the sake of simplicity in this paper we are considering only synchronous relationship types.

types *Married* and *Unmarried*) but it is defined only by one of them, because the other definition would be redundant.

Partitions require several meta integrity constraints. See [17] for some of them.

In the schema, we define a *disjointness integrity constraint* for each concept that is a sub of a partition.[3] In the above example, a *Man* cannot be a *Woman*, and the inverse. These integrity constraints are modeled as instances of *ICDisj*. Entity type *ICDisj* is derived, with derivation rule:

$$\text{ICDisj}(s,t) \leftarrow \text{IsASub}(c,p,t) \wedge s = \text{sym}(\text{ICDisj},c) \wedge \text{IsASub}(c',p,t) \wedge c \neq c' \qquad \textbf{MDR6}$$

The *Formula* in which *ICDisj IsExpressedBy* is given by the derivation rule:

$$\text{IsExpressedBy}(s,f,t) \leftarrow \text{ICDisj}(s,t) \wedge \text{ICDisjFormula}(s,f,t) \qquad \textbf{MDR7}$$

where predicate *ICDisjFormula* gives in *f* the string representation of the formula. As an example, the general form of *f* for a base relationship type R_1 in a partition of a binary relationship type *R* into $R_1,..., R_n$ is "$R_1(e1,e2,t) \rightarrow \neg\ (R_2(e1,e2,t) \vee ... \vee R_n(e1,e2,t))$".

In the schema, we define also a *subset integrity constraint* for each subconcept, except when the partition defines a derived superconcept. For example, if *Person* is a derived concept, defined by the partition into *Man* and *Woman* and also partitioned into base concepts *Married* and *Unmarried*, then we must ensure that all instances of *Married* or *Unmarried* are also instances of *Person*. The general form of an integrity constraint for a base entity type E_1 in a partition of a derived entity type *E* into $E_1,...,$ E_n would be "$E_1(e,t) \rightarrow E(e,t)$". These integrity constraints are modeled as instances of the derived entity type *ICSubs* using two meta derivation rules similar to the preceding ones.

Finally, we also include in the schema a *completeness integrity constraint* for each partition, except when the partition defines a derived superconcept. In the above example, we must ensure that all instances of *Person* are also instance of *Married* or *Unmarried*. The general form of *f* for a base entity type *E* partitioned into $E_1,..., E_n$ would be "$E(e,t) \rightarrow (E_1(e,t) \vee ... \vee E_n(e,t))$". These integrity constraints are modeled as instances of the derived type *ICCompl*, using other two meta derivation rules.

Now, we describe the schema derivation rules implied by partitions. If a derived concept is defined by a partition, then the schema must include its derivation rule. We model it as an instance of *DRPart*, which is a derived entity type subtype of *DerivationRule* (Figure 5). Its derivation rule is:

$$\text{DRPart}(s,t) \leftarrow \text{Defines}(p,c,t) \wedge s = \text{sym}(\text{DRPart},c) \qquad \textbf{MDR8}$$

Note again the use of *sym(DRPart,c)*, which denotes a symbol we use to refer to the derivation rule. We refine the base relationship type *IsExpressedBy* between *Rule* and *Formula* to a derived one between *DRPart* and *Formula*. The derivation rule is:

$$\text{IsExpressedBy}(s,f,t) \leftarrow \text{DRPart}(s,t) \wedge \text{DRPartFormula}(s,f,t) \qquad \textbf{MDR9}$$

[3] For flexibility, we allow a partition with only one subtype (Figure 5). In this particular case, it is obvious that disjointness constraints would not apply.

where predicate *DRPartFormula* gives the string representation of *f*. The general form of *f* in a partition of a derived entity type *E* into $E_1,\ldots, E_n$ is "$E(e,t) \leftarrow (E_1(e,t) \vee \ldots \vee E_n(e,t))$".

4 Schema Evolution

In our framework, schema evolution is performed by meta external events. When the designer wants to change a schema, she generates an instance of the appropriate meta external event type. The corresponding effect rules will induce meta structural events, which change the MIB. By the correspondence rules (Section 3.2), such changes imply changes to the schema.

The set of external event types included in the meta schema define the possible schema evolutions, either applied individually or by means of their composition. There is no other way to evolve a schema. Given that usually a schema may change in many different ways, it is necessary to include many external event types in the meta schema. For example, [1] defines 38 possible F2 schema changes. More are needed for richer schemas. We cannot show here the complete set of needed external event types. Rather, we are going to show how external event types and their effects can be defined in our framework.

Some meta external event types change only the schema, without affecting the IB. For example, *M_RemoveIntegrityConstraint* could have only the effect rule:

$$\text{M\_Del\_BaseIntegrityConstraint(ic,t)} \leftarrow \text{M\_RemoveIntegrityConstraint(rm,t)} \wedge \\ \text{RemovesIC(rm,ic,t)}$$

which, by CR4, would remove *ic* from the set of schema integrity constraints IC_t. Note that, in this case, the relationship *IsExpressedBy* between *ic* and its formula at *t* would be deleted by the compensating rule of MIC1 (Section 3.1).

Other meta external event types require changing the schema and the IB. We can distinguish two cases here: when the required changes to the IB are given by the typing rule (Section 2.1), and when they must be defined by explicit effect rules. An example of the former could be *M_RemoveRelType*, with appropriate preconditions and the effect rule:

$$\text{M\_Del\_RelationshipType(r,t)} \leftarrow \text{M\_RemoveRelType(rm,t)} \wedge \text{RemovesRT(rm,r,t)}$$

which, by CR2, would remove *r* from the set of the schema relationship types R_t and, by the typing rule, would delete existing instances of *r* at *t* in the IB. We will show a more complete example of this case in Section 4.1 below.

In other cases, the changes implied by the typing rule are not enough. We need to define an appropriate external event type in the schema, generate an occurrence of it and notify the IP of this occurrence. We generate “on the fly” an instance *EvT* of *ExternalEventType*, without parameters, and then we notify the IP of an occurrence of *EvT*. We will use in the meta effect rules the notation “*EvT* ← to mean that the MIP notifies the IP of an occurrence of *EvT*. We will show a detailed example of this case in Section 4.2 below.

4.1 M_RemoveEntityType

We start with a rather simple external event type, *M_RemoveEntityType*. It has only one parameter, given by relationship type *Removes* between that type and *Non-LexicalEntityType*. A particular instance of *M_RemoveEntityType*, occurring at time *T* and removing *E*, represents the design decision that, from time *T*, *E* ceases to be instance of *EntityType* and, therefore, the IB must not record the instances of entity type *E*. However, *E* may be again instance of *EntityType* later. *E* may be base or derived. For illustration purposes, we only deal here with the simple case when *E* does not participate in any relationship type and is not involved in any partition.

Preconditions. We define first the preconditions. One of them is that E is a non-lexical entity type at time T-1. Formally,

M_RemoveEntityType(rm,t) $\wedge$ Removes(rm,e,t) $\rightarrow$ Non-LexicalEntityType(e,t-1)

The if-part of the rule serves only to declare that an instance of the event type has occurred and to give name to its parameters. We omit this part in the next rules. Then, the other preconditions are:

$\neg\exists$participant IsETPtOf(e,participant,t-1)
$\neg\exists$partition IsTheSup(e,partition,t-1)
$\neg\exists$partition IsASub(e,partition,t-1)

Effect rules. The first effect rule induces the meta structural event that deletes E from Non-LexicalEntityType:

M_Del_Non-LexicalEntityType(e,t)$\leftarrow$M_RemoveEntityType(rm,t)$\wedge$Removes(rm,e,t)

As we did before, we omit the if-part of this rule in the next ones. Now, we must remove *E* from *BaseConcept* or from *DerivedConcept*:

M_Del_BaseConcept(e,t) $\leftarrow$ BaseConcept(e,t-1)
M_Del_DerivedConcept(e,t) $\leftarrow$ DerivedConcept(e,t-1)

Finally, if *E* is derived, we must delete its derivation rule:

M_Del_DerivationRule(s,t) $\leftarrow$ DerivedConcept(e,t-1) $\wedge$ s = sym(DerivationRule,e)

Induced effects on the MIB. The above effect rules are the only ones that need to be defined. We now reason (informally) about induced effects on the MIB:

– By the MDR's described in Section 3.1 (e.g. MDR1), *e* ceases to be instance of *Concept* and *EntityType*. The derived structural events *M_Del_Concept(e,t)* and *M_Del_EntityType(e,t)* are induced. Also, *M_Del_DerivationRule* induces *M_Del_Rule*. By the compensating rule of MIC1, *M_Del_IsExpressedBy(s,f,t)* is induced if *IsExpressedBy(s,f,t-1)*.

Induced effects on the schema and IB. We now reason about the induced effects on the schema and the IB:

– By CR1, *e* is not an entity type in the schema at *t* (that is, $e \notin E_t$)
– By the typing rule, all instances of *e* at *t-1* cease to be instances of *e* at *t*.

– By CR3, all instances of *DerivationRule* that have been deleted at *t* cease to exist in the schema at *t*.

Remarks. The removal of an entity type, or any other schema change, can affect indirectly to integrity constraints, derivation rules, or external event types defined explicitly in the schema by the designer. Some changes in those elements can be needed. The analysis of those changes results in a very complex task [15, 27], that is outside the scope of this paper.

4.2 M_ChangeParticipantToSET.

We now deal with a complex external event type, *M_ChangeParticipantToSET*. It has two parameters, given by relationship types *Changes*, with *Participant*, and *HasNewType*, with *EntityType*. An occurrence *M_ChangeParticipantToSET(CP,T)* of this event type, with *Changes(CP,P,T)* and *HasNewType(CP,NT,T)* represents the fact that, from time *T*, the participant *P* of a relationship type *R* changes its type to a supertype or a subtype,[4] the entity type *NT*. This means that, from *T*, instances of *R* are only valid if their corresponding participant is instance of *NT*.

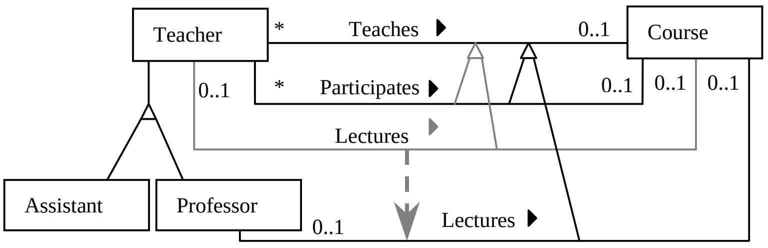

Fig. 6. An example of changing a participant in a relationship type

We consider the general case in which the relationship type has any degree, and is base or derived, subtype in zero or one partition, and super in any number of partitions. Changing a participant of a binary relationship type (attribute) has been studied extensively in the literature. However, we have not found the study of the change in our general settings.

If the relationship type is base, some of its instances may become invalid after the change. We have two possible strategies here: either to reject the change or to delete explicitly the invalid instances. The two possibilities must be offered to the designer. We could define a different external event type for each strategy, or add a new parameter to the event type, indicating the desired strategy [27]. In the following, we assume the existence of two different external event types, and present here the one that performs the strategy of deleting explicitly the invalid instances. In that case, if the relationship type is involved in a partition, deletion of some of its instances must be propagated through the partition hierarchy. For base types, the propagation must be

[4] *ToSET* stands for „***To** a **S**uper or a **S**ub **E**ntity **T**ype*„.

done explicitly. For derived types, the associated derivation rules might have to be corrected in the right way. On the other hand, if the relationship type changed is derived, the corresponding derivation rule needs to be changed too, so the invalid instances will be deleted implicitly.

This is a more complex change than the one exposed in Section 4.1, so for clarity we present here only an example. The complete formalization of the general case can be found in [17]. Assume a university information system (see Figure 6) with two base entity types, *Teacher* and *Course*, and also a base relationship type *Teaches* between *Teacher* and *Course*, partitioned into two subtypes, *Lectures* (base) and *Participates* (derived), with the same participants as the supertype. A teacher participates in a course if she teaches but does not lecture it. Any teacher can lecture courses or participate in them.

Assume now that the domain changes, and *Teacher* is transformed into a derived type, partitioned into *Professor* and *Assistant*. Now, among other changes, only a professor can be the lecturer of a course. That implies changing the participant of the relationship type *Lectures* from *Teacher* to *Professor*. The effects of this change, in our framework, would be the following:

The type of the participant must be explicitly changed in the MIB:

M_Del_IsETPtOf (Teacher, P, T)
M_Ins_IsETPtOf (Professor, P, T)

with P being the symbol representing the participant being changed, and T the time instant when the change is done.

As *Lectures* is base, P has a referential integrity constraint that must change. This change is performed automatically by MDR5, which induces:

M_Del_IsExpressedBy(S,*'Lectures(tch,c,t)* $\rightarrow$ *Teacher(tch,t)'*, T)
M_Ins_IsExpressedBy(S,*'Lectures(tch,c,t)* $\rightarrow$ *Professor(tch,t)'*, T)

with S=*sym(ICRefInt,P)*.

Some of the instances of *Lectures* may violate the above new integrity constraint. In that case, they must cease to exist. As *Lectures* is base, we must define a new external event type in the conceptual schema to explicitly eliminate its invalid instances; and also to eliminate them as instances of *Teaches*, since it is a base supertype of *Lectures*. We use the special function *newSymbol()* to obtain new symbols in the MIB, when needed. The explicitly defined effects in the MIB must be the following:

et=newSymbol()
M_Ins_ExternalEventType(et,T)

er_1=newSymbol()
M_Ins_EffectRule(er_1,T)
M_Ins_HasEffect(et,er_1,T)
M_Ins_IsExpressedBy(er_1,*'Del_Lectures(tch,c,t)* $\leftarrow$ *Lectures(tch,c,t-1)*
$\wedge$ $\neg$*Professor(tch,t-1)'* ,T)

er_2=newSymbol()
M_Ins_EffectRule(er_2,T)
M_Ins_HasEffect(et,er_2,T)

M_Ins_IsExpressedBy(er$_2$,'*Del_Teaches(tch,c,t)* ← *Del_Lectures(tch,c,t)*',T)

Finally, we have to notify the IP of an occurrence of the external event type we have just defined:

et ←

Remarks. As pointed out at the beginning of this section, this event type can require the change of some derivation rules, corresponding to derived relationship types of the partition hierarchy. As we think that this kind of change must be modeled as an independent external event type, in this case would be needed the composition of instances of both event types to achieve the whole desired change.

5 Conclusions

We have presented a framework that allows the definition of the evolution of information systems, at the conceptual level. The framework is based on a reflective architecture, with two loosely coupled conceptual processors.

Our framework allows the definition of all conceptual modeling constructs in the meta schema. This has been illustrated with three representative constructs. On the other hand, our framework allows the definition of any evolution operation, and this has been illustrated also by two particular operations, which have been analyzed in detail.

Our work can be extended in at least five directions. First, we could define, in the meta schema, other modeling constructs (aggregation, materialization, etc.). Second, we could develop the complete (both at schema and instance-level [8]) set of meta external event types, for the minimal meta schema and for the considered constructs. Other complex external event types could be added to that set to facilitate the evolution to the designer [6]. Of course, this should be done on the basis of existing work. Third, a prototype implementation could be developed, preferably using two processors of different type. In view of practical applications, the two processors should be loosely coupled, as suggested by our framework. Fourth, that prototype could include a tool to assist and guide the designer in the definition and analysis of evolution operations [27, 9]. For instance, warning her about external event types possibly affected by a particular schema change. Finally, building on top of existing knowledge and experience, one could attempt the development of an industrial system, with several possible degrees of ambition.

Acknowledgments

The authors are indebted to Dolors Costal, Cristina Gómez, Camilo Ocampo and Toni Urpí, and to the anonymous referees, for their helpful comments. This work has been partially supported by CICYT program projects TIC99-1048-C02-1 and TEL99-0335-C04-02.

References

1. Al-Jadir, L., Léonard, M.: Multiobjects to Ease Schema Evolution in an OODBMS. In: Proc. ER'98, Singapore, LNCS 1507, Springer (1998) 316-333
2. Analyti, A., Spyratos, N., Constantopoulos, P.:Property Covering: A Powerful Construct for Schema Derivations. In: Proc. ER'97, LNCS 1331, Springer-Verlag (1997) 271-284
3. Boman, M., Bubenko jr., J.A., Johannesson, P., Wangler, B.: Conceptual Modelling. Prentice Hall (1997)
4. Banerjee, J., Chou, H-T., Garza, J.F., Kim, W., Woelk, D., Ballou, N.: Data Model Issues for Object-Oriented Applications. In: ACM TOIS Vol.5, No.1, January (1987) 3-26
5. Bertino, E., Ferrari, E., Guerrini, G.: T_Chimera: A Temporal Object-Oriented Data Model. TAPOS 3(2) (1997) 103-125
6. Brèche, P.: Advanced Primitives for Changing Schemas of Object Databases. 8Th. In: Intl. Conf. CAiSE'96, LNCS 1080, Springer (1996) 476-495
7. Bubenko jr., J.A.: The Temporal Dimension in Information Modelling. In: Architecture and Models in Data Base Management Systems, North-Holland (1977)
8. Casais, E.: Managing Class Evolution in Object-Oriented Systems, In: D.C. Tsichritzis (ed.): Object Management (1990) 133-195
9. Erwald, C.A., Orlowska, M.E.: A procedural approach to schema evolution. In: Proc. CAiSE'93, LNCS 685, Springer (1993) 22-38
10. Goralwalla, I., Szafron, D., Özsu, T., Peters, R.: A Temporal Approach to Managing Schema Evolution in Object Database Systems. In: Data&Knowledge Eng. 28(1), October (1998) 73-105
11. Hainaut, J-L., Englebert, V., Henrard, J., Hick, J-M., Roland, D.: Database Evolution: the DB-MAIN Approach. In: 13th. Intl. Conf. on the Entity-Relationship Approach - ER'94, LNCS 881, Springer-Verlag (1994) 112-131
12. ISO/TC97/SC5/WG3: Concepts and Terminology for the Conceptual Schema and Information Base. J.J. van Griethuysen (ed.), March (1982)
13. Jarke, M., Gallersdörfer, R., Jeusfeld, M.A., Staudt, M., Eherer, S.: ConceptBase - a deductive object base for meta data management. In: Journal of Intelligent Information Systems, 4(2) (1995) 167-192
14. Lemke, T., Manthey, R.: The Schema Evolution Assistant: Tool Description. IDEA.DE.22.O.004, University of Bonn (1995)
15. Lemke, T.:Schema Evolution in OODBMS: A Selective Overview of Problems and Solutions. IDEA.WP.22.O.002, University of Bonn (1994)
16. Lemke, T.: DDL = DML?. An Exercise in Reflective Schema Management for Chimera. IDEA.WP.22.O.003, University of Bonn (1995)
17. López, J.R., Olivé, A.: A Framework for the Evolution of Temporal Conceptual Schemas of Information Systems – Extended Version. LSI-00-14-R. Department of Software (LSI). Universitat Politècnica de Catalunya (2000)
18. Manthey, R.: Beyond Data Dictionaries: Towards a Reflective Architecture of Intelligent Database Systems. In: DOOD'93, Springer-Verlag (1993) 328-339
19. Nguyen, G.T., Rieu, D.: Schema evolution in object-oriented database systems. In: Data&Knowledge Eng. 4 (1989). 43-67
20. Olivé, A., Costal, D., Sancho, M.-R.: Entity Evolution in IsA Hierarchies. In: Proc. ER'99, LNCS 1728, Springer-Verlag (1999) 62-80
21. Olivé, A.: "Relationship Reification: A Temporal View". In: Proc. CAiSE'99, LNCS 1626, Springer (1999) 396-410
22. Peters, R.J., Özsu, T.: Reflection in a Uniform Behavioral Object Model. In: Proc. ER'93, Arlington, LNCS 823, Springer-Verlag. (1993) 34-45
23. Peters, R.J., Özsu, T.: An Axiomatic Model of Dynamic Schema Evolution in Objectbase Systems. In: ACM TODS Vol. 22, No. 1, March (1997) 75-114

24. Roddick, J.F.: Schema Evolution in DataBase Systems – An Updated Bibliography. In: ACM SIGMOD Rec., 21(4), May (1994) 35-40
25. Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modeling Language Reference Manual. Addison-Wesley (1999)
26. Tansel,A., Clifford,J., Gadia,S. et al.: Temporal Databases: Theory, Design and Implementation. Benjamin/Cummings (1993)
27. Zicari, R.: A Framework for Schema Updates in Object-Oriented Database System. In: Bancilhon,F.; Delobel,C.; Kanellakis, P. (eds.): Building an Object-Oriented Database System - The Story of O_2. Morgan Kaufmann Pub. (1992) 146-182