

**Reuse of a repository of conceptual schemas in a large scale project**

Carlo Batini, University of Milano Bicocca - batini@disco.unimib.it

Tel 0039 +2 6448 7826 Fax 0039 +2 6448 7817

Riccardo Grosso, CSI Piemonte - Riccardo.Grosso@csi.it

Tel 0039 +11 3169253

Total number of words: 4670 words + references + 11 figure + 1 table

Submission category: Completed Research

## **Reuse of a repository of conceptual schemas in a large scale project**

Carlo Batini, University of Milano Bicocca

Riccardo Grosso, CSI Piemonte

**Abstract** . *Large amounts of data are managed by organizations, available to be viewed and analysed from multiple perspectives, which becomes a fundamental resource to the effectiveness of the organizations. An organization can achieve full benefit from the available information by managing its data resource, through the planning of its exploitation and its maintenance. The concept of data repository fulfils these requirements, due to the fact that it contains the description of all types of data produced, managed, maintained and exchanged in an organization. Data descriptions should be organized in a repository to enable all the users of the information system to understand the meaning of data and the relationships between among them.*

*This paper describes an experience of the use of an existing repository of conceptual schema, representing a wide amount of entities of interest for Central Public administration, in order to produce the corresponding repository of the administrations located in a region. Several heuristics are described and experiments are reported.*

Keywords: Enterprise modelling, architecture and governance, Database Modeling, Repositories of conceptual schemas

### **1. Structure of Italian Public Administration and previous experiences of conceptual schema Repositories**

The goal of this paper is to describe an experience of use of a repository of conceptual schemas, related to Central Italian public administration, in order to build a first version of the corresponding repository of conceptual schemas of the local public administration located in one of the 21 regions of Italy. Due to the limited amount of available resources, several approximate techniques have been applied, that allow fast prototyping of the local repository, to be refined by domain expert, resulting in a resource consumption one order of magnitude lower than that with a traditional process.

The Italian Government's policy, in the past few years, similarly to many other governments in the world, has been to improve the quality of services to the citizen, by gradually improving services provided by information systems and databases of its agencies. However, in the past the lack of co-operation among the departments led to the establishment of heterogeneous and isolated systems. As a result, two main problems have arisen:

- duplicated and inconsistent information;
- difficult data access

Moreover, the Government efficiency depends on the sharing of information between the departments, due to the fact that many of them are usually involved in the same procedures, while using different, overlapped, heterogeneous databases.

Therefore, in the long term, a crucial aspect for the overall project is to design a cooperation architecture that allows administration to share information in such a way as to be able to provide services to citizens and businesses on the basis of the "one stop shop" paradigm. A crucial aspect of such cooperation architecture is the data architecture: data have to be interchanged in an interoperable format, all the administration assign the same meaning to the same data, achieving database integration in the long term; this will provide the spread of information within the government branches and will result in a more easily accessible working environment, an increased quality of information management, and an improved state-wide decision making.

Data base integration has to be achieved in the complex scenario of Public Administration. The structure of the Public Administration (PA) in Italy consists of central and local agencies that together offer a suite of services designed to help citizens and businesses to fulfil their obligations towards the PA. Central Public Administrations are of two types, Ministries such as Ministry of Interiors and Ministry of Revenues, and other central Agencies such as Social Security Agency, Accident Insurance Agency and Chambers of Commerce. Main types of Local Administrations correspond to Regions (21), Provinces (about 100) and Municipalities (about 8.000).

The approach to cooperation among administrations followed in Italy to address this problem is based on the concept of Cooperative Information Systems (CIS), i.e., systems capable of interacting by exchanging services with each other. The general cooperative architecture for the Nationwide CIS network of the Italian P.A as shown in Figure 1.

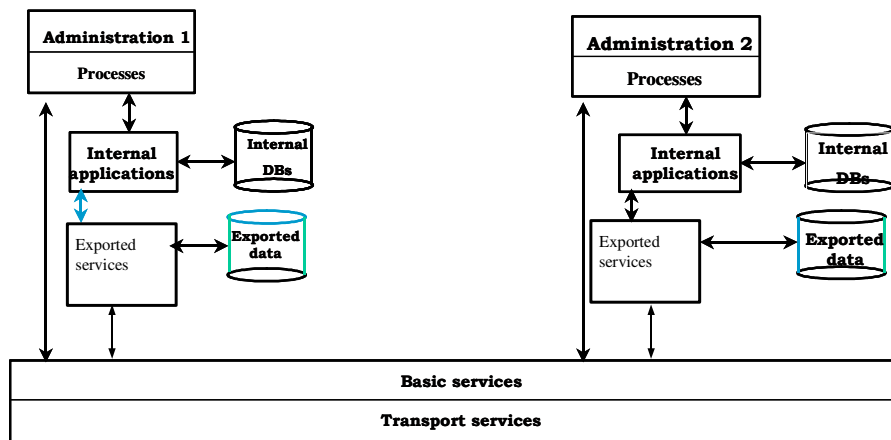


Figure 1: The structure of the cooperative architecture

Besides transport and basic services, a cooperative services layer is shown, including application protocols, repositories, gateways, etc. The main idea is to define a *domain* as the collection of all the computing resources, networks, applications and data that belong to a specific administration. Each domain defines the set of cooperative interfaces that include data and application services available to other systems.

One of the first activities performed in the last decade, with the final goal of designing a suitable data architecture, has been the project of building an inventory of existing information systems operating within the Central Public Administration in Italy. The activity was performed over about 500 data bases, whose logical schemas through reverse engineering activities were translated into Entity Relationship schemas. In order to provide a structure to such a large amount of schemas, a methodology for building repositories of conceptual schemas, described in [Batini C. et al 1993] was used. We describe in brief such methodology in the next section.

In order to achieve cooperation among central and local administrations, it is the moment to design a data architecture that covers both types of administrations, and, consequently, it is necessary to develop a similar repository for local administrations. For this reason, several regional administrations are now designing their own data architecture. The most advanced organizational context among local administrations in a region is when they are coordinated by a

regional agency, that provides services to all or at least to the majority of them. This is the situation of administrations of the Piedmont region, where such central agency exists, and is CSI Piemonte. But also in such a fortunate context, only logical relational schemas are available as input to the process of construction of the local repository. So, a methodology and tools are needed that allow approximate production of conceptual schemas to be arranged in a repository. In this paper we describe such methodology and the experience we achieved so far in applying this to the context of the Piedmont Public Administrations.

The paper is organized as follows. In section 2 we discuss the structure of the Central Administration Repository and we recall the methodology for its construction. In section 3 we describe knowledge available for the design of the local PA repository. In section 4 we provide the methodology for building, starting from the central repository and local logical schemas, a first draft version of the local repository. Section 5 discusses experiences and future research work.

## **2. The Structure of the Central PA Repository**

A repository, in the context of the paper, can be defined as a set of conceptual schemas, each one describing all the information managed by an organisation area within the information system considered. In particular, the data repository referenced in this paper uses the Entity Relationship model to represent conceptual schemas. However a simple set of schemas does not display the relationships among schemas of different areas; the repository has to be organised in a more complex structure, through the use of the structuring primitives.

The primitives are: refinements, views; integration. Refinements allow the description of the same reality at different levels of abstraction. This mechanism is fundamental for a data repository, since it helps the user to perceive a complex reality step by step, going from a more abstract level to a local one. Views are descriptions of fragments of a schema. They allow users to focus their attention just on the part of a complex reality of interest to them. Integration is the mechanism by which it is possible to build a global description of data managed by an organisation area starting from local schemas. By jointly using these structuring primitives we obtain a repository of schemas. Each column of the repository represents an organisation unit while each row stands for a different abstraction level. The left column contains the schemas resulting from the integration of all the other schemas belonging to the same row (views of the integrated schema). In figure 2 we show an example of repository, where the Production, Sales, Department Schemas are represented at different refinement levels respectively in the second, third and fourth column, while the Company schema in the first column is the result of their integration.

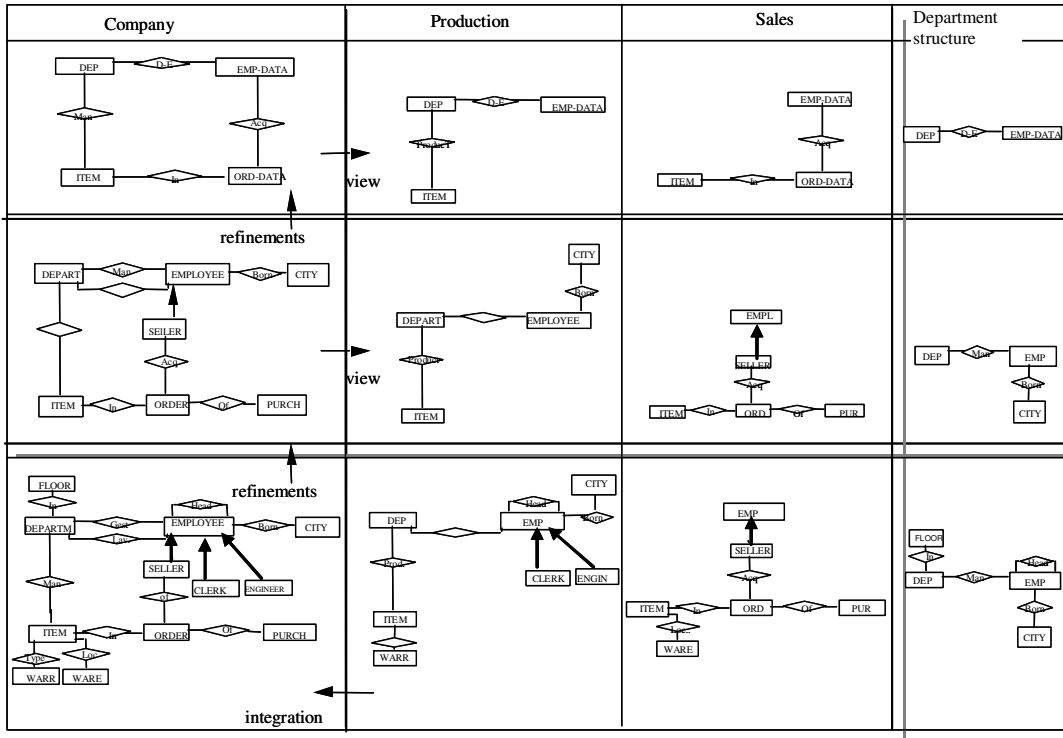


Figure 2 : An example of repository

In practice, when the repository is populated at the bottom level by hundreds of schemas, as in the case that we will examine in the following, it is unfeasible to manage the three structuring primitives, and the view primitive is sacrificed. Furthermore, the integration/abstraction structuring mechanism is iterated, producing a sparsely populated repository such as the one symbolically represented in figure 3, where, for instance, schema S123 results from the integration/abstraction of schemas S1, S2, and S3.

SI12345678										
	SI123				SI456				SI78	
		S1	S2	S3		S4	S5	S6		S7 S8

Figure 3: A fragment of repository

The repository structure described previously has been adopted for representing the conceptual content of a wide amount of conceptual schemas related to the most relevant databases of Italian central public administration in an integrated structure.

In order to build the whole repository, the following methodology has been adopted, defined in more detail in [Batini et al 1993].

1. First, conceptual schemas representing the different organisation areas are grouped in terms of homogeneous classes, corresponding to meaningful administrative matters of interest in central public administration; 27 different matters have been defined: examples of matters are social security, finances, cultural heritage, education. As we said, at the bottom level of the repository we have about 500 schemas, corresponding to the logical schemas of data bases of the 21 most relevant central administrations in Italy, with approximately 5.000 entities and a similar number of relationships. We name in the following basic schemas the conceptual schemas defined at the bottom of the repository.
2. Each group of basic schemas is integrated and, at the same time, abstracted, resulting in a unique schema for each matter, that populates the second level of the repository, resulting in 27 second level abstract schemas. In figure 3 the different levels of the repository are represented, starting from the second level; for instance, the Internal security second level schema results from the integration/abstraction process, performed over 6 schemas corresponding to 130 concepts.
3. The integration/abstraction process is iterated, producing higher level schemas, corresponding to more abstract matters, such as financial resources, human resources, social services, economic services, finally producing a unique integrated schema, that is further abstracted resulting, at the topmost level of the repository, in the schema shown in figure 5, that represents at a very high level the most significant concepts represented in the information systems of any Public Administration, i.e. Individuals, Legal persons, Properties, Places, Documents. The resulting pyramid of schemas provides a natural representation of concepts at different abstraction levels, and allows, with suitable approximation, to find the common parts among databases pertaining to different agencies.

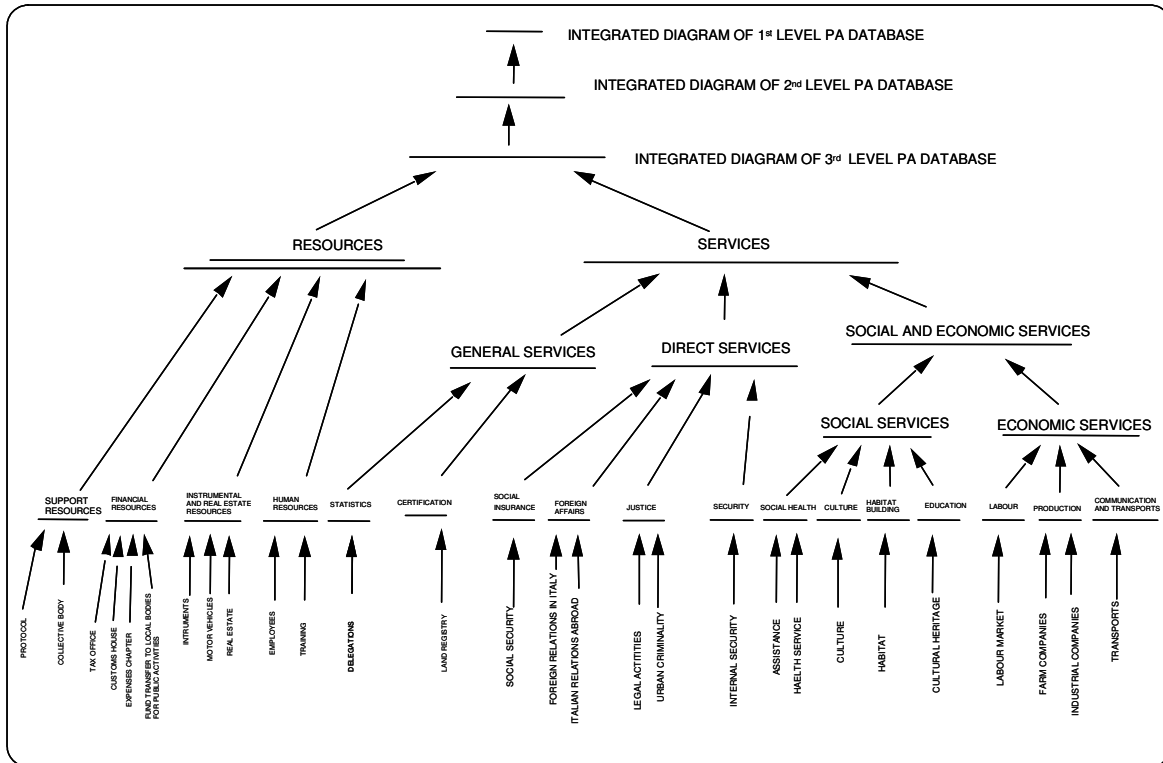


Figure 4: The whole repository of schemas.

In order to produce the repository, about 200 person months were needed to produce the 500 basic conceptual schemas, while about 24 person months were needed to produce the 55 abstract schemas of the upper part (approximately 2 weeks per schema, both for basic and for abstract schemas).

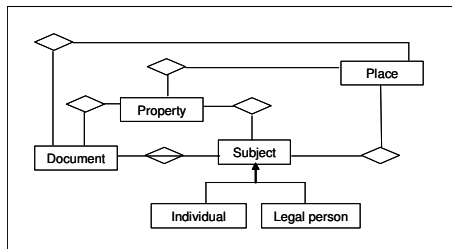


Figure 5: The schema at the top level of the repository

### 3. The Repository of Piedmont Local Administrations: basic knowledge available

In this section we describe in more detail the knowledge available for the design of the Piedmont Local public administration (LPA) repository and the assumptions that have been made in the activity.

A first relevant input available for the process is the Central Public Administration (CPA) Repository of schemas, made of basic and abstract schemas. A second input concerns Piedmont databases. Piedmont local public administration are centrally served by a unique consortium, CSI Piemonte, that created in the last years approximately 450 databases of 12 main local administrations, whose logical schemas are documented in terms of: relational database schemas, tables (approximately 17.000), textual descriptions of tables, referential integrity constraints defined among tables, attributes, definitions of attributes, identifiers. A very thin conceptual documentation has been created, that concern so called “supertypes of attributes” and “supertypes of relations”, corresponding to generalization abstractions of a few attributes and tables (about 10%) defined in the logical schemas. They have not been used so far in the process.

The basic sources of knowledge available for the production of the LPA repository, as results from the above discussion, are very rich, but characterized by two significant heterogeneities: the conceptual documentation concerns central administration, while for local Piedmont administration the prevalent documentation concerns logical schemas.

A second relevant condition of our activity has concerned budget constraints; for the first year of the project we had only one person year available, so less than one tenth of the resources that were available for the construction of the central repository. So, in conceiving the methodology for the LPA repository production, we made a few significant assumptions, and used heuristics and approximate reasoning, in order to reduce human intervention as much as possible.

A first assumption we made has been that, while basic schemas of the CPA repository and the LPA repository may probably differ, due to the different functions among central and local administrations, the similarity should be much higher among the abstract schemas of the CPA repository and basic + abstract schemas of the LPA repository.

In consequence of the above assumption and resource constraints, we decided to use in some steps of the methodology a more manageable knowledge base than the 500 central basic schemas + the 50 abstract schemas. Such schemas can be represented in terms of a much more dense conceptual structure, that corresponds to the generalization hierarchies that have at their top level the five concepts defined in the schema of figure 4, and having at lower levels the concepts in more refined abstract schemas and basic schemas, obtained applying top down the refinements along the integration/abstraction hierarchy. We show in figure 5 a fragment of one of the hierarchies, the one referring to individuals.



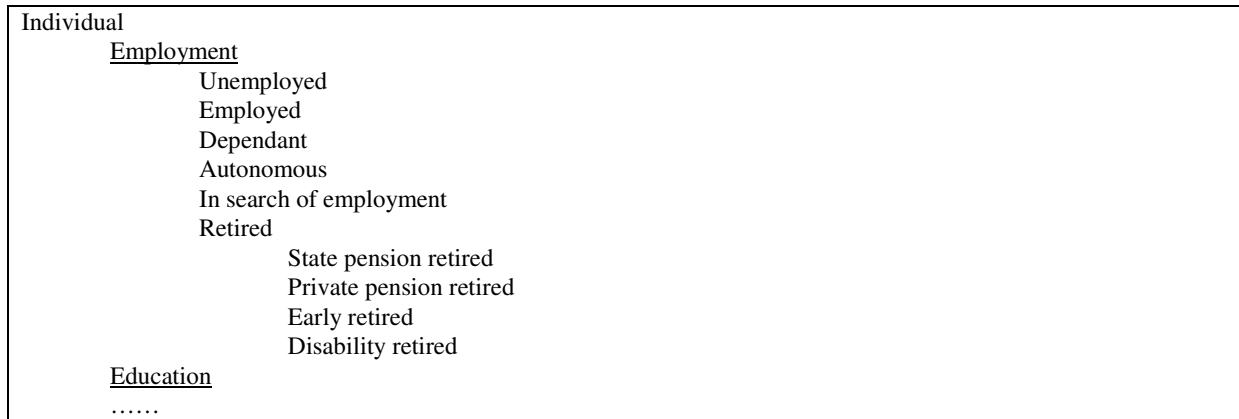


Figure 6: A fragment of the Individual generalization hierarchy

So, a second idea we implemented has been to use, besides the basic schemas and the abstract schemas, the five generalization hierarchies of Individual, Legal Person, Property, Document, Place.

As a consequence of the above assumptions, constraints and choices, the inputs to the methodological process, shown in figure 7, have been:

1. The CPA Repository of 550 basic + abstract schemas
2. The five CPA Generalization hierarchies
3. The logical schemas of the 450 local PA databases.

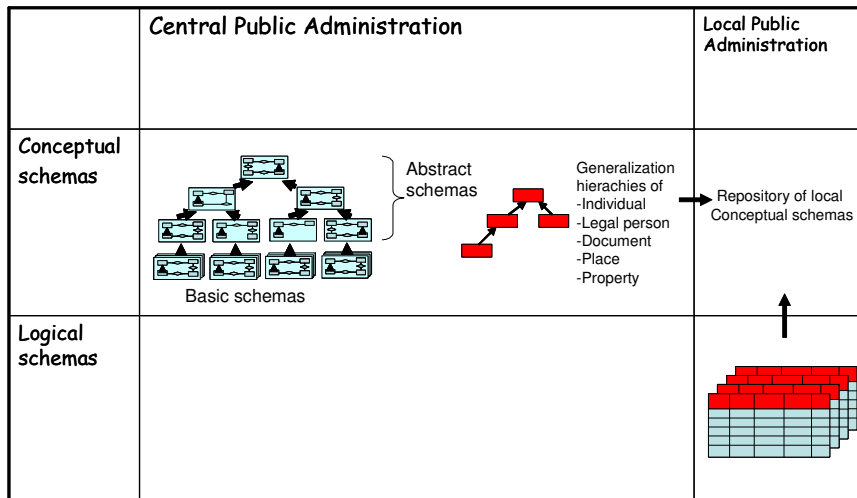


Figure 7: Input knowledge for the production of the Repository of local conceptual schemas

#### 4. The methodology for the construction of the local Repository

First, we provide the rationale of the methodology, and then we detail specific steps. The methodology follows a mixed approach in building the basic schemas and the upper part of the repository, and consequently, can be seen divided in two phases.

For each local logical schema, available conceptual/central and logical/local knowledge is used in phase 1 to build a basic local conceptual schema (steps 1-5 of the methodology); then, in phase 2, the abstract local schemas conceptually related to the new local basic schema are produced, resulting in a progressive construction of the upper part of the local repository (step 6). Local logical schemas are chosen on the basis of their relevance and the quality of the documentation, thus producing a stable skeleton of abstract schemas in the early phases of the iterative activity, that is monotonically enriched with new knowledge, as long as new basic schemas are produced.

We examine now the two phases in detail, while in section 5 we discuss the result of a first set of experiments performed in order to measure the effectiveness of the methodology and tune it. Each step is described with a common documentation frame, describing the inputs to the step, the procedure, and in some cases, when relevant, the outputs of the step. An example is provided, related to a logical schema concerning grant monitoring of industrial business activities.

Phase 1: Construction of a basic schema

1. Extract entities

Inputs: Central PA generalization hierarchies of concepts, one Local PA logical schema

Names of entities in hierarchies are compared with names and description of each table, and set of attributes of the logical schema. The comparison function makes use presently of a simple distance function among the different

strings. The entities and corresponding frequency of matching are sorted, and a threshold is fixed: all the entities with frequency over the threshold are selected, resulting in a first draft schema made only of entities. The output is a draft schema made of disconnected entities.

## 2. Add generalizations

Inputs: the draft schema obtained in the previous step and the four CPA generalization hierarchies.

Visit the generalization hierarchies and add to the draft schema subset relationships present in hierarchies, defined among the entities in the draft schema.

## 3. Extract relationships

Inputs: the draft schema + all the basic schemas in the CPA repository

Entities of the draft schema are pairwise compared with all the basic schemas in the CPA repository. For each pair of entities E1 and E2 several types of relationships are extracted by the basic schemas:

- a. relationships defined exactly on E1 and E2;
- b. relationships corresponding to chains of relationships defined among pairs  $E1-E_i$ ;  $E_i-E_{i+1}$ ; ...;  $E_{i+j}-E2$ ;
- c. relationships defined among entities  $E1^*$  and  $E2^*$  corresponding to ancestors of E1 and E2 in the four generalization hierarchies.

Relationships collected in steps a and c are sorted according to the frequency of names. Here we have several possibilities:

- a. The most frequent name is chosen as the name of the relationship
- b. The name is assigned by the domain expert.

## 4. Check the schema with referential integrity constraints defined among logical tables

Input: the draft schema + constraints defined in tables

For each referential integrity constraint defined among two tables T1 and T2 in the logical schema, it is checked whether T1 and/or T2 have been already selected as entities in the draft schema, and in case added as new entities. Furthermore, it is checked whether a relationship is defined among the entities, and in case added.

## 5. Domain expert check of the draft schema and construction of the final schema

Input: the draft schema

In this step the schema produced by the semi automated process is examined by the knowledge domain expert that may add new concepts, cancel existing concepts, or else modify some concepts

Since step 5 is performed after addition of relationships and entities resulting from integrity constraints, it may happen that too many concepts have been added, and the manual check of the domain expert leads to delete concepts. Sometimes new concepts are added, resulting in an enriched schema whose kernel is the initial schema. More frequently schemas obtained after integrity constraints check and after domain expert check coincide

Output: the final schema

We show in figure 8 the schemas obtained as a result of the execution of steps 1 to 5 of the methodology in our case study. In this case, schemas obtained after integrity constraints check and after domain expert check coincide, and, consequently, are not distinguished in the figure.

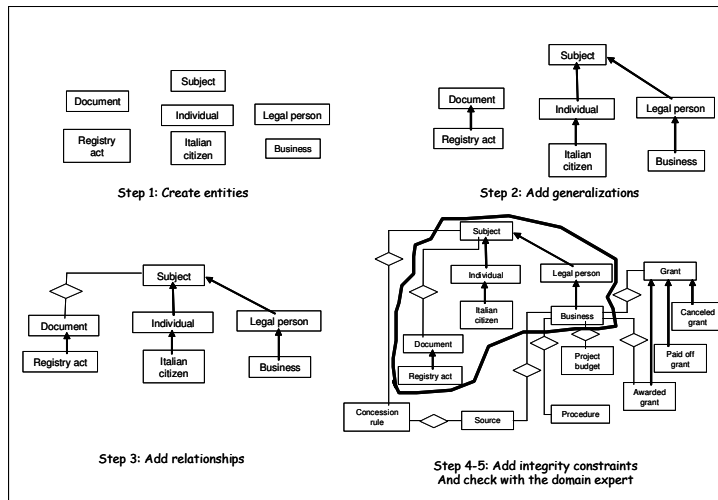


Figure 8: Schemas obtained after steps 1-5

## 6. Repository update

This step may effectively use the results of previous steps 1-5. In fact, the initial schema obtained after steps 1-3 inherits high level abstract knowledge from the CPA Repository and basic knowledge from the LPA logical schemas, while the enriched schema obtained in steps 4-5 encapsulates basic knowledge from the LPA logical schemas. So we may conjecture that the initial schema is a candidate for abstract schema for the upper levels of the LPA repository, while the enriched schema, being a more detailed description representing a logical schema, populates the basic level of the repository (see figure 9).

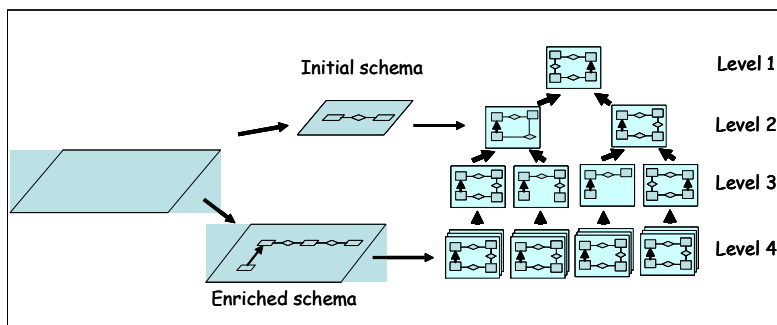


Fig 9: Presumable locations in the repository of initial and enriched schemas

So, we may conceive two possible strategies for the repository update step (see figure 9), that we describe with the usual format.

Input: the actual version of the LPA repository and a new pair of initial + enriched schemas.

First strategy: starting from the initial schema and the enriched schema (dotted lines in figure 10.a) we first (step 1 in fig 10.a) complete the “local” repository of abstract schemas corresponding to the enriched schema (see in figure 11 the result of the step for our example); we then (step 2 figure 10.a) integrate the local repository with the actual one: it may happen that we have to update, due to similarities between concepts, the abstract schemas of the actual repository, or else add new schemas, autonomous with respect to the previous ones.

Second strategy: the new repository is obtained through abstraction/integration activities on the actual LPA repository and the initial and refined schemas.

The final output of the step is the enriched LPA repository.

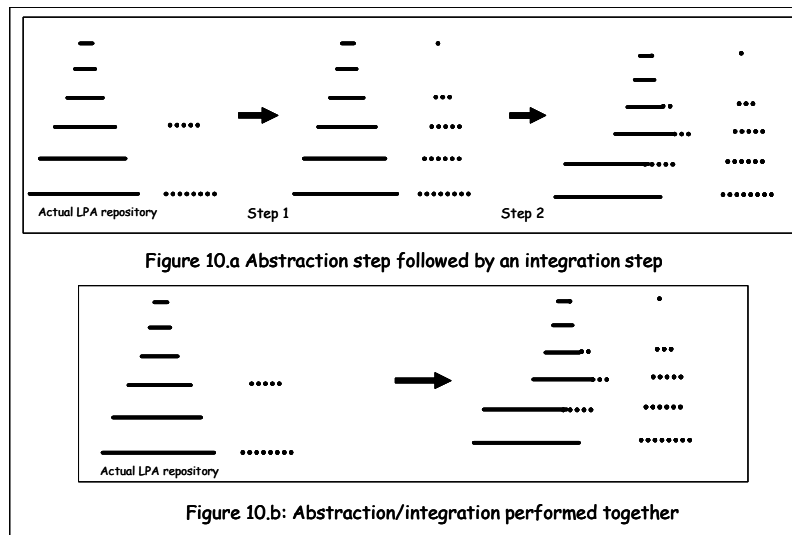


Fig 10: Two possible strategies to update the LPA repository

The first strategy is probably more effective when the actual LPA repository and the new schema represent very different knowledge, while the second strategy has the advantage of natively using the structuring paradigm of the repository, the abstraction/integration operation. We are currently experimenting with the two strategies, and other possible strategies, such as building small homogeneous repositories and then integrating them to obtain a larger repository.

## 5. Experiments

In the present stage of the project we have experimented the above methodology in three different matters and nine related areas, namely:

### Businesses

- Grants to industrial businesses
- Economic/productive activities
- Farms register

HealthCare

Abortions/miscarriages statistics

Regional territory

Provincial Road register

Dam register

Laws on water management

Hydrogeological network

City toponymic

The total number of tables of the nine databases is approximately 350, corresponding to 2% of the total.

We were interested in measuring two relevant qualities of the process:

1. the correctness of the conceptual schema with respect to the “true” one, i.e. the schema that could be obtained directly by the domain expert through a traditional analysis or else a reverse engineering activity. Correctness is measured with an approximate indirect metrics, corresponding to the percentage of new/deleted concepts in the schema produced by the expert at the end of step 5 with respect to concepts produced in the semi automatic steps 1-4.
2. the completeness of the conceptual schema with respect to the corresponding reengineered logical schema. Completeness is measured by the percentage of tables that are caught in steps 1-5, in comparison with the total number of tables, after excluding tables not carrying relevant information, such as redundant tables, tables of codes, etc.

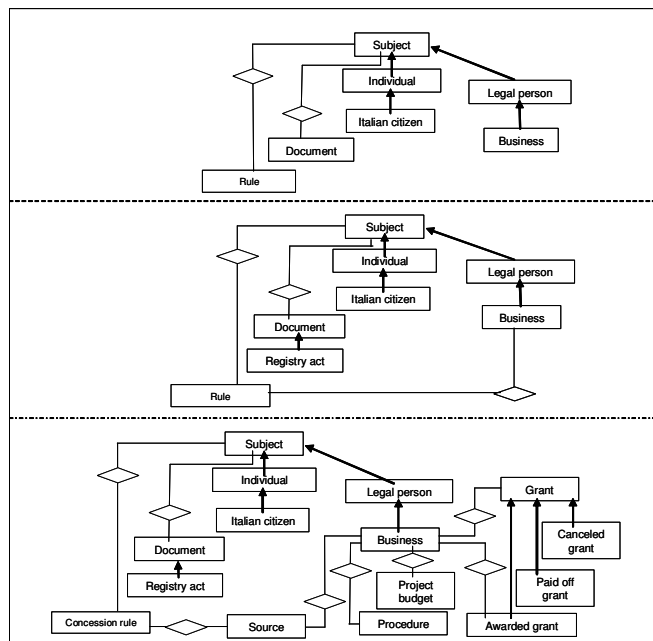


Fig 11: abstract schemas obtained from the basic schema

Table 1 summarizes main results of experiments. Concerning correctness, in general the schemas obtained after step 4: check with integrity constraints, and after step 5: domain expert check are very similar, i.e. domain experts tend to confirm and consider complete entities and relationships added in the previous step; the overall figure for the nine experiments results in more than 80% of concepts common to the two types of schemas. We see also that the add constraints step introduces approximately 30% of new concepts in comparison with the extract entities step. Consequently the joint application of the Central PA knowledge and Local PA knowledge reveals effective. These are, in our opinion, encouraging results, considering the highly heuristic nature of the methodology.

Concerning completeness, results are less reassuring. On the average, only 50% of tables are caught. This value changes significantly in the different areas. Furthermore, as was to be expected, completeness decreases significantly when the referential integrity constraints are not documented or partially documented, resulting in lower quality (completeness) conceptual schema when the input schema is characterized by poor documentation. Apart the quality of the documentation, another cause of reduced completeness is the static nature of generalization hierarchies used in step 1, and the unequal semantic richness in representing related top level concepts. For instance, in the initial Subject hierarchy, 20 concepts represent individuals, while only 3 represent legal persons. An improvement we are presently applying concerns their incremental update with abstract concepts generated in steps 6-7. Such enriched hierarchies are progressively reconciled and brought near to hierarchies characteristic of local administrations, resulting in a corresponding more effective selection mechanism.

Step	# of tables extracted	% of tables extracted
Create entities	172	30
Add constraints	219	41
Domain expert check	275	51

Table 1: Experiments results

A final comment on resources. The amount of resources spent in the experiments has been on the whole 30 person/days, corresponding to 3 person/day per schema. About 30% of time has been spent in steps 1-4, and 60% of time has been spent in the manual check. So, the domain expert has been engaged for 2 days per schema; we have to add to this variable cost a fixed cost of a 3 days course. We may expect a greater efficiency as long as the activity proceeds, and fix in 1 person day the average final due effort, significantly lower than the typical 2-3 person/weeks needed for traditional design of one schema.

## 6. Concluding remarks

The problem addressed in this paper, and the related conceptual tools, are not new in the literature.

Repositories of conceptual schemas are proposed in several application areas (e.g. biosciences [TDWG 2004]) for reuse ([Ruggia et al 1997]). [Castano et al 1997] ). [Castano et al 1998] propose criteria and techniques to support the establishment of a semantic dictionary for database interoperability are described. Similarity-based criteria are used to evaluate concept closeness and, consequently, to generate concept hierarchies. The techniques allow the analysis of

conceptual schemas of databases in a federation and the definition and maintenance of concept hierarchies. Experimentation of the techniques in the public administration domain is discussed.

[Perez et al 2002] presents a solution and methodology for reverse engineering of legacy databases using formal method-based techniques. Repositories of ontologies are proposed in several papers. An approach for interoperation across pre-existing Ontologies is discussed in [Mena et al 1996]. The alignment and integration of ontologies is investigated in [Wang et al 2002], [Di Leo et al 2002], [Farquhar et al 1995], where information integration is enabled by having a precisely defined common terminology. A set of tools and services is proposed to support the process of achieving consensus on such a common shared ontologies by geographically distributed groups. Users can quickly assemble a new ontology from a library of modules.

In [Pan et al 2003] multi-agent systems rely on shared ontologies to enable unambiguous communication between agents. An ontology defines the terms or vocabularies used within messages encoded, using an agent communication language (ACL). In order for ontologies to be shared and reused, ontology repositories are needed to store and maintain ontologies.

Repositories of ontologies for public sector organizations are proposed in [Slota et al 2003]. The repository was used in a system supporting organizational activity by formalizing, sharing and preserving operational experience and knowledge for future use.

What seems new in our approach as regards the above mentioned papers is the abstraction/integration primitive adopted for structuring the repository and the attention devoted to feasibility aspects and resource constraints, and the consequent heuristic strategy. On the other side, we are conscious that our conceptual model is less powerful than ontology based models. A complete comparison with existing approaches is out of the scope of this paper.

We are now analyzing lessons learned and improving the methodology along the lines discussed in previous sections; we are also investigating new techniques that use more complex similarity measures in matching between generalization hierarchies and logical schemas. Furthermore, since some of the local PA schemas (and corresponding hierarchies) have been independently developed, especially in the regional territory area, we are using such schemas as training examples to tune semiautomatic steps of the methodology and similarity measures adopted.

## References

- [Batini et al 1993] C. Batini, G. Di Battista, G. Santucci - Structuring primitives for a dictionary of entity relationship data schemas -IEEE Transactions on Software Engineering vol.19 no.4 April 1993.
- [Batini et al 1996] C. Batini, S. Castano, V. De Antonellis, M.G. Fugini, B. Pernici - Analysis of an Inventory of Information systems in the Public Administration - Requirements Engineering, 1996, 47-62
- [Batini et al. 1996] C. Batini, S: Castano, B. Pernici - Tutorial on Reuse Methodologies and Tools - Entity Relationships International Conference, Cottbus, Germany, 1996.
- [Castano et al 1997] S. Castano, V. De Antonellis, Semantic dictionary design for database interoperability 13th International Conference on Data Engineering April 07 - 11, 1997, University of Birmingham, Birmingham, U.K.
- [Castano et al 1998] S. Castano, V. De Antonellis, B. Pernici – Conceptual Schema analysis: techniques and applications. ACM Transactions on Data Base Systems, Vol. 23, No. 3, 1998.
- [Di Leo et al 2002] Jonathan DiLeo, Timothy Jacobs, and Scott DeLoach. Integrating Ontologies into Multiagent Systems Engineering. Fourth International Bi-Conference Workshop on Agent-Oriented Information Systems (AOIS-2002). 15-16 July 2002, Bologna (Italy).



- [Farwuhar et al 1995] Adam Farquhar, Richard Fikes, Wanda Pratt, James Rice - Collaborative Ontology Construction for Information Integration - Knowledge Systems Laboratory Department of Computer Science, KSL-95-63, August 1995.
- [Fonseca et al 2003] F. Fonseca, C. Davis, G. Camara – Bridging Ontologies and Conceptual schemas in Geographic Information systems – *Geoinformatica* 7:4, pp. 355 – 378, 2003.
- [Pan et al 2003] J. Pan, S. Cranefield, D. Carter International Conference on Autonomous Agents. Proceedings of the second international joint conference on Autonomous agents and multiagent systems Melbourne, Australia Pages: 632 – 638, 2003.
- [Perez et al. 2003] J. Perez, I. Ramos, J. Cubel, F. Dominguez, A. Boronat, J. Carsi Data reverse engineering of Legacy Databases to object oriented conceptual schemas - *Electronic Notes in Theoretical Computer Science* 74 No. 4, 2002.
- [R. Ruggia et al 1997] R. Ruggia, A.P. Ambrosio – A toolkit for Reuse in Conceptual Modelling – Proceedings Caise Conference, 1997.
- [Slota et al 2003] R. Slota, M. Majewska, M. Dziewierz, K Krawczyk, M. Laclavik, Z. Balogh, L. Hluchy, J. Kitowski, S. Lambert Ontology Assisted Access to Document Repositories for Public Sector Organizations, PPAM 2003, Czestochowa, Poland
- [TDWG 2004] Taxonomic Databases Working Group on Biodiversity Informatics 2004, Taxonomic Databases Working Group Annual Meeting, 11-17 October 2004 - University of Canterbury, Christchurch, New Zealand.
- [Wang et al 2002] J. Wang, L. Gasser, “Mutual Online Ontology Alignment” AAMAS-2002 Workshop on Ontologies for Agent Systems.