Chapter 13
Tunisian Ministry of Agriculture Planning of New Services and Information Systems Integration

As we said in the preface, the eG4M methodology has been applied to eGovernment projects in the Mediterranean area. In this chapter we focus on the data governance part of the methodology that has been applied in the Tunisian Ministry of Agriculture, in parallel with running initiatives on the reorganization of databases managed in the different administrative organizational structures of the ministry. The documentation available has been partial and, consequently, the case study has been simplified w.r.t. the real context and in some cases, modified. The reader has to be conscious that the Tunisian agricultural domain has only inspired the case study that, as a consequence, has to be seen as a paradigmatic example of any initiative of integration of databases, in particular in the agricultural domain. In Sect. 13.1 we provide some insight on the organizational structure of the ministry considered. Section 13.2 describes the activities performed, the roles of the different players, and the organization of work. In Sect. 13.3 we introduce the databases considered in the case study and their conceptual description. Following the methodology discussed in Chap. 2, in Sect. 13.5 the repository of the database conceptual schemas is produced. The analysis of the repository for strategic planning of future possible initiatives discussed in Sect. 13.6 concludes the chapter.

13.1 Organizational Structure of the Ministry of Agriculture

As happens in a great number of countries, we start our description assuming that each department in the ministry has information systems integrated at a vertical level, namely with decentralized agencies providing to the central databases data originated in the Tunisian regions they are in charge of administering. A strategic challenge for the ministry of agriculture and hydraulic resources (MAHR in the following) is the integration at a horizontal level of databases of the central administrative departments; the goal is to have an integrated vision on the whole set of activities and matters the ministry is in charge of, supporting strategic and political decisions and forecasts, e.g., on the productivity of the agricultural sector.

This chapter is authored by Carlo Batini and Gianluigi Viscusi.

G. Viscusi et al., *Information Systems for eGovernment*,
DOI 10.1007/978-3-642-13571-2_13,

Because of their role in MAHR core processes, we assume in the following that four major administrative departments have priority in the horizontal integration objective, namely

- the department in charge of the restructuring of agricultural state-owned domains;
- the department of veterinary services;
- the department of water resources; and
- the department of agricultural production.

Several feasibility studies and preliminary projects have been proposed in the past, always focused on a single department. Indeed, in this case the goal of the application of the eG4M methodology is to plan horizontal integration of the databases of the four considered departments.

13.2 The Activities Performed and Organization of Work

The MAHR has been the focus of the application of the eG4M methodology mainly for the strategic planning phase. State reconstruction has been focused on the following issues:

- The re-engineering of the current databases, namely the production of the corresponding conceptual schemas and the production of the repository of schemas, in order to produce the integrated representation of the whole information system of the MAHR, analyze the drawbacks of the current usage of databases, and conceive the target database architecture.
- The data quality evaluation of the databases, in terms of their currency and completeness, in order to assess the effectiveness of the current administrative processes and services provided by the public administration.
- The representation of the interactions between organizational units of MAHR in terms of services and related processes, together with the types of information involved and the ownership of the databases; the goal is to define the priority intervention areas in the integration initiative.

After a set of courses introducing the methodology and related topics, the above-described three issues have been developed through participatory design workshops involving four teams composed of three civil servants (one senior manager and two middle-level managers) each from the four departments. A three days a week workshop had a monthly frequency. ICT skills and computer literacy among the participants were not homogeneous, where a major divide was mainly related to the age of the attendees and to their functional role in the organization (some departments being represented by the IT staff while others by non-IT human resources).

In the following we focus on the first objective, the evolution of the data architecture. We suggest the reader to anticipate the reading of Appendix B, where different technologies and solutions to achieve integration among databases are described.

The four databases analyzed concern the following topics:

1. Agricultural state property
2. Veterinary services
3. Water resources schema
4. Agricultural production

The topics correspond one to one to the departments mentioned above. Workshops were dedicated mainly to a participated design [110] of the re-engineered conceptual schemas for the databases of each department and of the final integrated conceptual schema; besides this activity, a check of the matrices filled by each department representing organizational units and data flows (see Chap. 5) has been carried out.

Participatory design of the local conceptual schemas and of the integrated conceptual schema (using as formalism the entity – relationship model discussed in Chap. 2) started from the current logical schemas of the department databases, previously designed by external (often private) agencies.

Our main goal has been to build a complete and integrated view of the information content managed by MAHR with the collaboration of administration. To achieve this goal, we applied the data governance methodology described in Chap. 2.

The process followed and the different outputs produced are shown in Fig. 13.1. A different type of line is associated with each schema that will also be used in the following to distinguish the different schemas.

We assume that the four databases analyzed are documented in terms of a hybrid logical – conceptual schema plus a natural language description of requirements.

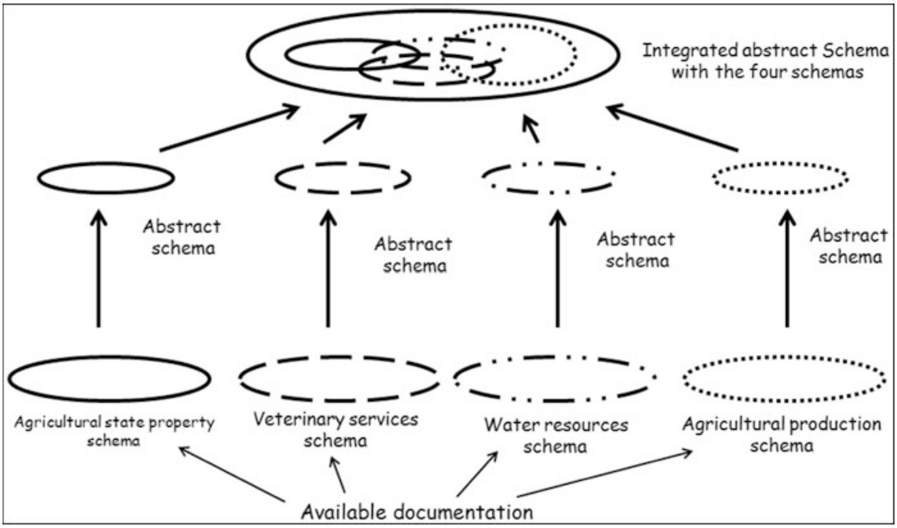

Fig. 13.1 The abstraction integration tree of the four schemas

Starting from this documentation, a reverse engineering activity is performed that produces the four conceptual schemas.

At this point, since the total number of entities is around 50, we first produce an abstract version for each schema and we integrate subsequently the four abstract schemas into the final abstract integrated schema.

13.3 Conceptual Schemas of the Databases

The four conceptual schemas resulting from the reverse engineering step of the methodology are shown in Figs. 13.2, 13.3, 13.4, and 13.5.

For reasons of simplicity, in the representation we have included the entities, the relationships, and IS-A hierarchies defined among them, while attributes, identifiers, and names of relationships are not reported. Note that the terms used for names of entities belong in part to a domain-independent vocabulary, such as in the case of `Project`, `Region`, `Grant`, and in part to domain-specific vocabularies, such as the public administration vocabulary for `Governorship` and `Delegation`, and the agriculture vocabulary, such as `Apiculture`, `Intensive farming`, and `Vaccination`.

The schema integration activities on conceptual schemas are greatly facilitated by the availability of this vocabulary, especially in the issues related to semantic relationships between terms such as synonymies or homonymies. We recall that a *synonymy* among two terms occurs when the two terms have the same meaning.

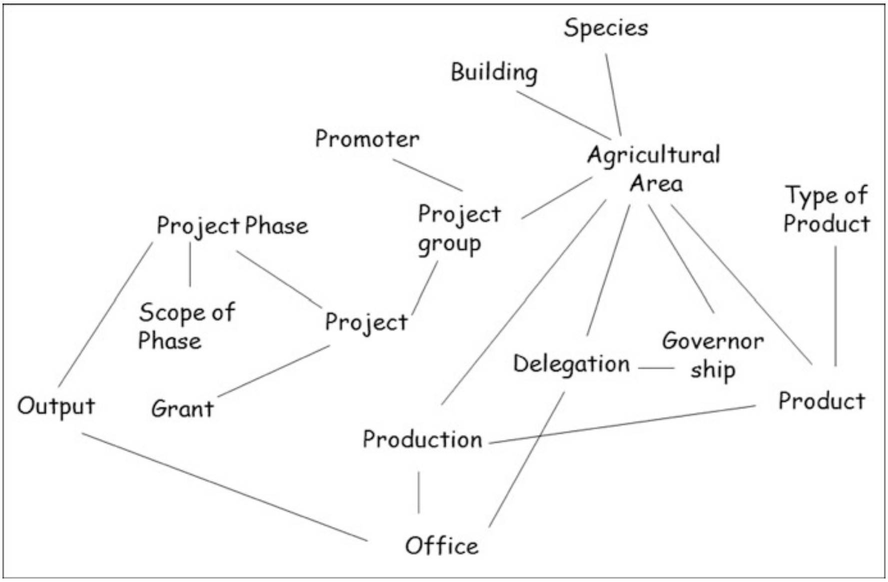

Fig. 13.2 The agricultural state property schema

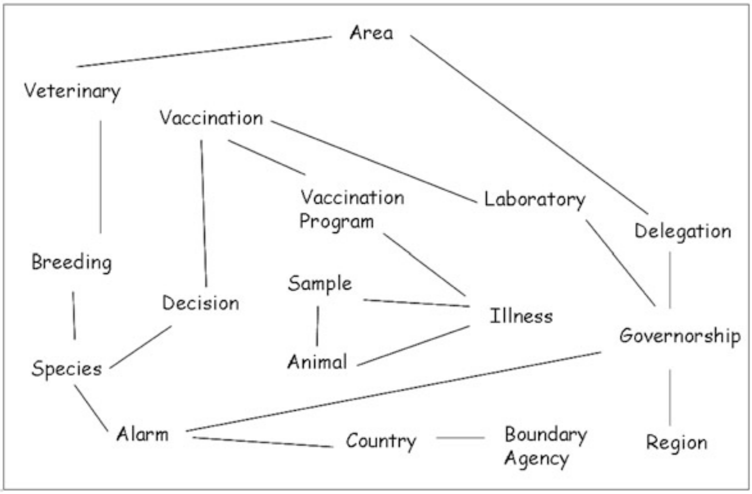

Fig. 13.3 The veterinary services schema

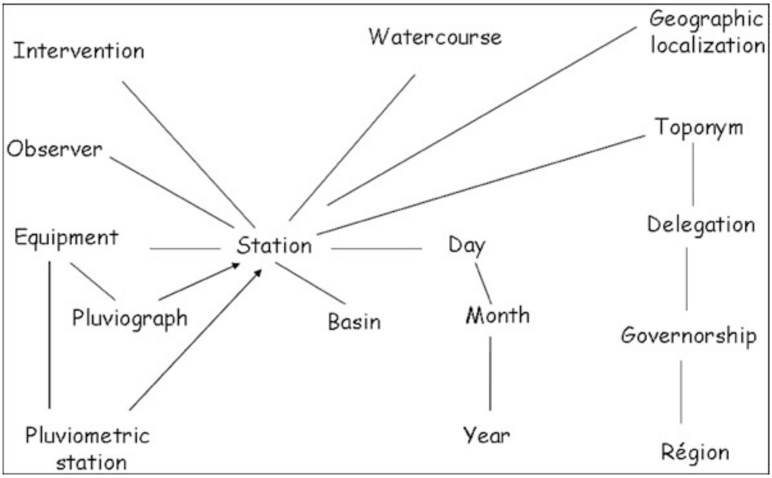

Fig. 13.4 The water resources schema

A *homonymy* among two occurrences of the same term in different schemas occurs when the two occurrences of the term have different meanings (for a complete discussion on semantic relationships among terms, see [78]). When a vocabulary is enriched with semantic relationships between terms it is called *lexicon*. For a discussion on lexicons the reader may refer to [78].

In the case of four schemas, we have been very careful in the reverse engineering activity to choose exact terms, not too general, not too specific, in order to minimize the occurrences of synonymies and homonymies. Looking with attention at similar terms in different schemas, we conclude that a synonymy exists between

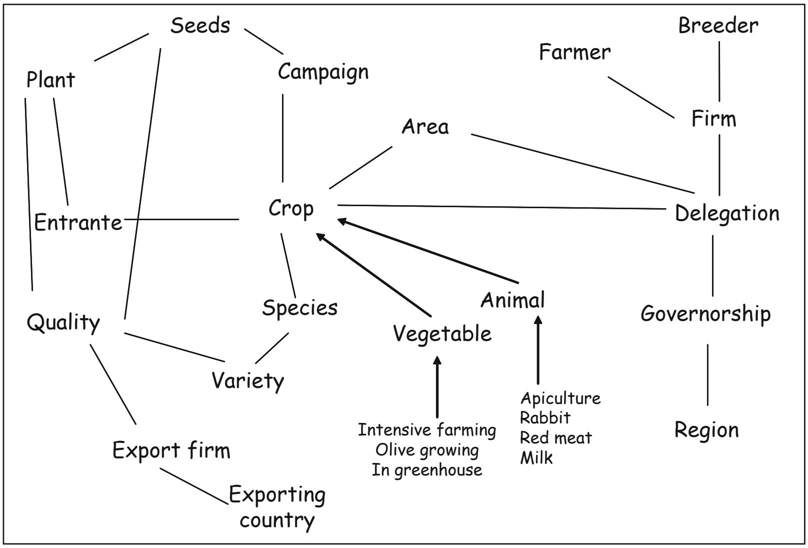

Fig. 13.5 The agricultural production schema

the entity `Agricultural area` in the `Agricultural state property` schema and the two entities `Area` in the `Veterinary services` and in the `Agricultural production` schemas. We may solve the synonymy by changing `Agricultural area` into `Area`.

13.4 The Abstractions on Schemas

We move now to the second step of the methodology, namely the production of abstract schemas. For each schema we have to choose the groups of concepts (entities, relationships, IS-A hierarchies) to be abstracted into a unique concept in the corresponding abstract schema. The abstractions chosen for the four schemas are shown in Figs. 13.6, 13.7, 13.8, and 13.9. We put in evidence the groups of concepts to be abstracted by means of closed lines and add a box to the border of the line with a name for the abstract concept.

In order to avoid the introduction of new synonyms and homonyms, in the assignment of abstract names it is worthwhile to have a look at the entire set of schemas, e.g., a choice that we made has been to assign the same name, `Agricultural production resource`, to two groups of concepts in the water resources schema and in the agricultural production schema. Looking at the two groups, they are dissimilar. If the repository had more than two levels, it would be worthwhile to assign `Agricultural production water resource` and `Vegetable and animal production resource` as abstract names, respectively, to put in evidence the difference between the two types of resources. Another possible choice could be to distinguish among input resources and output resources.

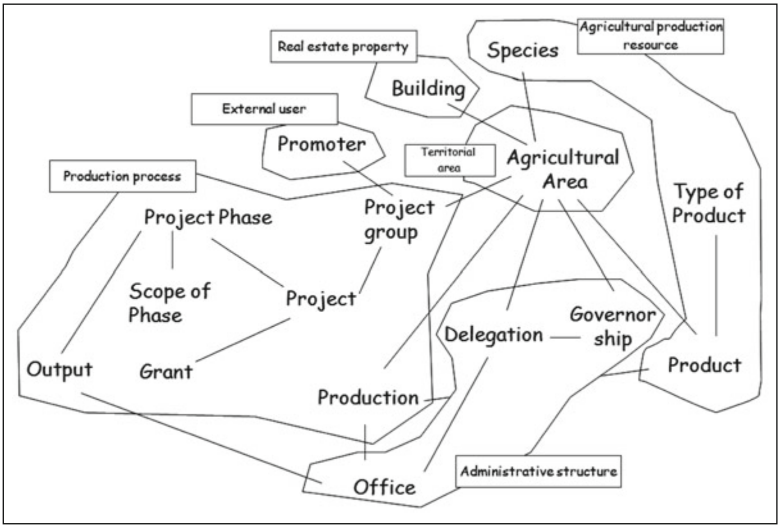

Fig. 13.6 Abstractions for the agricultural state property schema

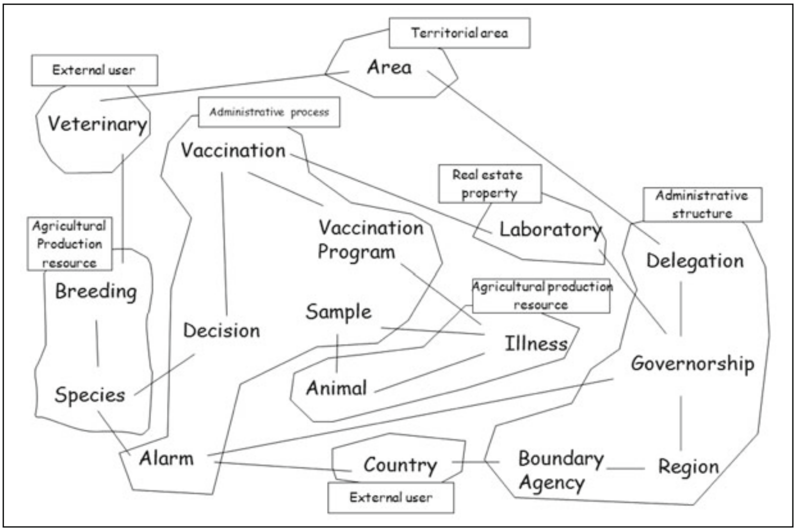

Fig. 13.7 Abstractions for the veterinary services schema

A second example concerns the abstract concepts `Geographic localization` and `Territorial area`. Starting from the same motivation as before, namely, the unique level of abstraction available, we could assign the same name to the two abstract concepts. This choice would lead to an error, since `Geographic localization` is a concept with a very specific meaning that has to be represented in the abstract schema to put in evidence the differences w.r.t. `Territorial area`.

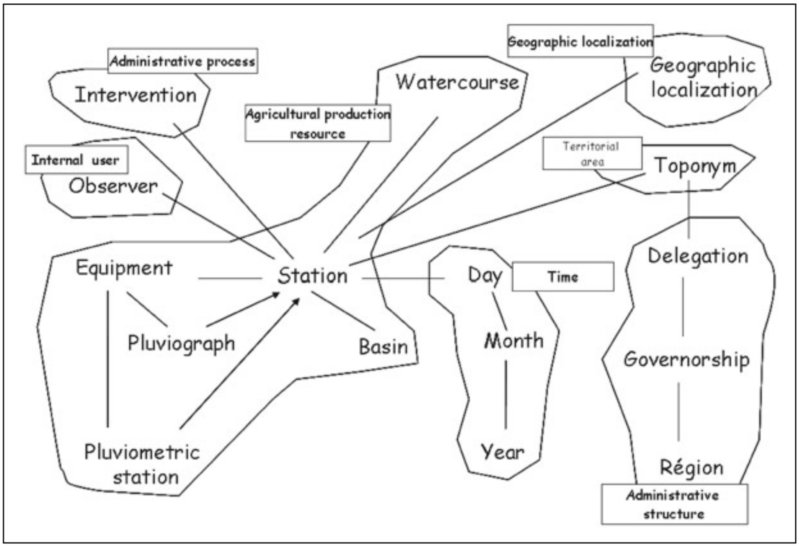

Fig. 13.8 Abstractions for the water resources schema

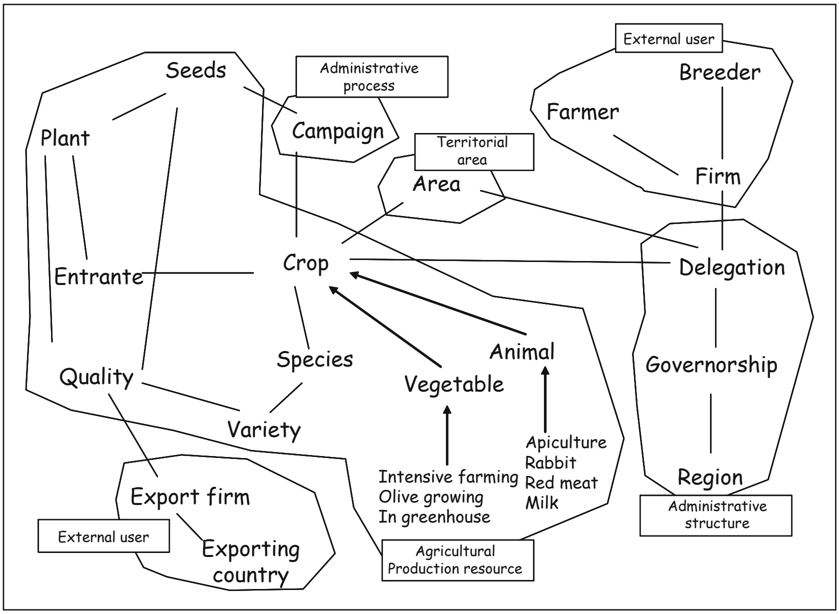

Fig. 13.9 Abstractions for the agricultural production schema

To conclude, we have to note that there are no fixed rules for the abstraction process and that experience is helpful in uncertain contexts.

Another issue that deserves a comment concerns the size of clusters of concepts to be abstracted. Coming back to the previous example, the clusters associated with Agricultural production resource are wider than the clusters associated with Administrative structure. This is not surprising since the number of resources used or produced in agriculture is wide while (fortunately) the number of different types of administrative structures is usually small. Other times the difference in size could be an indication of a different "abstraction speed" in the abstraction process. In these cases we have to fragment large clusters into smaller ones, in order to achieve better balancing in the abstraction process.

13.5 The Repository of Schemas

We are now ready to produce the upper levels of the repository of schemas. We obtain the abstract schemas corresponding to the four initial schemas by substituting each cluster of concepts with the unique abstract concept chosen. In doing so, we have also to inherit the connections of clusters of concepts with other clusters. The process leads to produce the four abstract schemas shown in Figs. 13.10 and 13.11.

The final step concerns the integration of the four abstract schemas. We follow the methodology described in Sect. 2.4. Synonymies and homonymies in the four initial schemas have been analyzed and solved before; we have also checked in the abstraction steps that no new name heterogeneity has been introduced. Thus, we can simply superimpose the four schemas obtaining the integrated schema shown in Fig. 13.12. In this regard, we have to add the *interschema properties*, corresponding to the semantic relationships among concepts in different schemas which were hidden previously, but have to be added now for concepts that have been merged. An example of interschema property is the generalization hierarchy between External user and Internal user that leads to the introduction of a common ancestor between them, namely User.

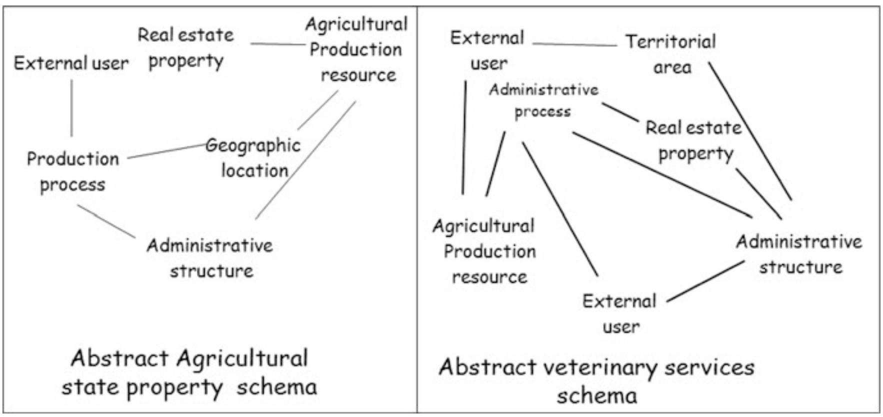

Fig. 13.10 The abstract agricultural state property and veterinary services schemas

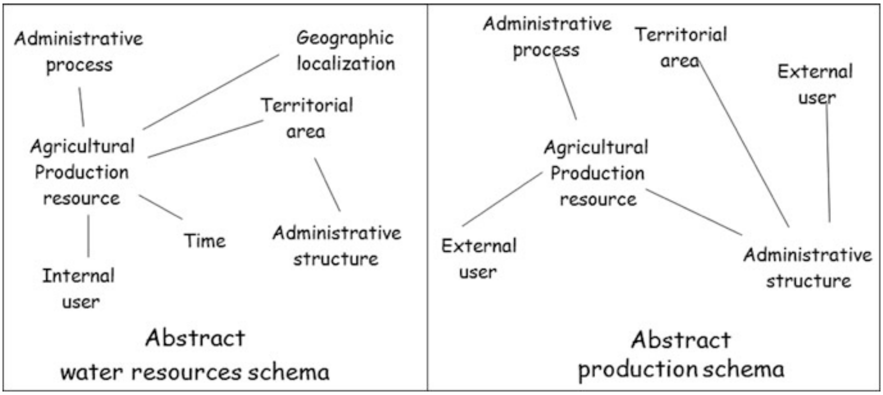

Fig. 13.11 The abstract water and production schemas

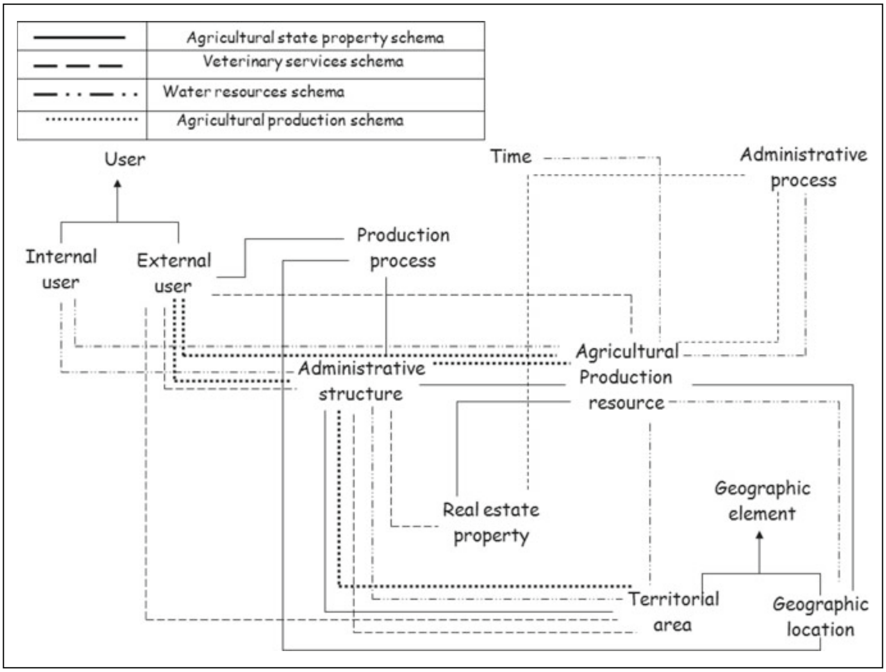

Fig. 13.12 The abstract integrated schema with the four local schemas in evidence

13.6 Analysis of the Schema Repository to Achieve Effective Strategic Planning Decisions

The analysis of the repository of schemas can be performed at different levels of abstraction. Looking at the entire documentation we highlight the following problems.

- Different entities exist which are common to several schemas. Looking at the schema of Fig. 13.12 the multiplicity of schemas in which an entity is involved is graphically highlighted by the number of incoming/outcoming edges. This situation usually leads to redundant representation of data and high risks of misalignment among the different copies of the same data. In the choice of future projects we may consider the possibility to establish a publish – subscribe layer among schemas that disciplines and aligns the updates of the entity instances in different databases.
- As natural for any ministry of agriculture, the MAHR jurisdiction is in practice on all the territory of the country. While all databases need to represent the territory, only one database, the `Water resources` database, explicitly represents the *geographic localization*. This choice results in heterogeneities in the representation of the territory and leads to limited integration in the representation of the different entities grounded in the territory, e.g., the state property resources cannot be correlated with the cadastral databases, which another ministry is in charge of. A design choice to be evaluated in more detail in a feasibility study should be to move to a homogeneous representation of the territory in the four databases, by means of a spatial DBMS, namely a DBMS that allows to represent spatial objects (such as points, lines, polylines, areas) natively in the logical model.
- Looking at the different types of `Internal` and `External users`, an incompleteness in the representation of users is perceived. This observation may lead to looking at services provided by the ministry in such a way to find the set of all possible types of users, so as to discover in case new services are to be provided effectively.

The participatory analysis activities put in evidence other problems. Once designed the re-engineered conceptual schema for the department of agricultural production, one of the managers attending the workshop pointed out that the schema was wrong, due to the lack of a relationship between the entity `Farmer` and the entity `Agricultural product` (a kind of strategic information in the case of MAHR).

We first checked with the help of other attendees that the schemas produced exactly match with the one produced by the external company involved in the production and maintenance of the current database, the result was negative. At this point all the attendees became aware of the fact that in order to maintain the control over the outsourcer companies, the methodology and the participatory work can be very helpful in providing an effective instrument for managing information and knowledge of interest for the administration.

Furthermore, the awareness arose that in strategic and operational activities on the information system of administration it may be worthwhile to leave the deployment and maintenance activities to the market (namely, private vendors), while the public administration has to retain the planning, design, and quality evaluation of the eGovernment projects, with particular attention paid to data governance decisions.

We finally applied the decision process on the optimal evolution of the architecture described in Appendix A to the four databases (we suggest to read the appendix before reading this part). We first collected information on the application load and other variables considered in the decision process.

Figure 13.13 shows a decision table for the participatory choice of technologies suitable to support the exploitation of the integrated architecture. The figure results from the usage of the decision table shown in Fig. 13.14 and discussed in detail in Appendix A.

Figure 13.13 shows that publish – subscribe is a suitable technology for all the departments, these latter fulfilling requirements such as high autonomy and low relevance of querying with respect to update operations; whereas a data warehouse solution is needed for some of the involved departments, i.e., the department in charge of the restructuring of agricultural state-owned domains, the department of veterinary services, and the department of agricultural production. The conclusion is that

1. all databases fit the criteria shown in the upper gray line, thus leading to the proposal to adopt a publish – subscribe architecture on common entities for all of them;

Condition of use (advantage)	Condition of use (inconvenience)	Department	Technology
▪HIGH AUTONOMY OF MANAGEMENT ▪HIGH ECONOMIC VALUE OF THE INTEGRATION ▪LOW RELATIVE IMPORTANCE OF THE QUERYING WITH RESPECT TO THE UPDATE	▪LOW OVERLAPPING OF THE DATA BETWEEN THE DIFFERENT DATABASES	1) The department of animal production; 2) The department of the restructuring of agricultural state-owned domains; 3) The department of veterinary services; 4) The department of water resources management	Enterprise Application Integration/Publish & Subscribe
▪LOW AUTONOMY OF MANAGEMENT ▪HIGH COST OF HETEROGENEITY ▪HIGH ECONOMIC VALUE OF THE INTEGRATION ▪HIGH COMPLEXITY OF MANAGEMENT	▪HIGH AUTONOMY OF MANAGEMENT	-	Central Database
▪LOW AUTONOMY OF MANAGEMENT	▪HIGH NUMBER OF LOCAL QUERIES	-	Distributed Database
▪HIGH AUTONOMY OF MANAGEMENT ▪LOW VOLATILITY OF SOURCES ▪HIGH IMPORTANCE FOR THE ACCESS TO HISTORICAL DATA ▪HIGH RELATIVE IMPORTANCE OF THE QUERYING WITH RESPECT TO THE UPDATE ▪HIGH ECONOMICAL VALUE OF THE INTEGRATION ▪HIGH COMPLEXITY OF THE QUERYING	▪LOW RELATIVE IMPORTANCE OF THE QUERYING WITH RESPECT TO THE UPDATE ▪HIGH VOLATILITY OF SOURCES	1) The department of animal production; 2) The department of the restructuring of agricultural state-owned domains; 3) The department of veterinary services	Datawarehouse
▪HIGH AUTONOMY OF MANAGEMENT ▪HIGH VOLATILITY OF THE QUERYING ▪HIGH RELATIVE IMPORTANCE OF THE QUERYING WITH RESPECT TO THE UPDATE ▪HIGH COST OF HETEROGENEITY ▪HIGH ECONOMIC VALUE OF THE INTEGRATION	▪LOW RELATIVE IMPORTANCE OF THE QUERYING WITH RESPECT TO THE UPDATE ▪HIGH VOLATILITY OF SOURCES	-	Enterprise Information Integration

Fig. 13.13 Decision table for the participatory choice of suitable technologies

Decision criteria									Suggested solution
Autonomy	Relevance of historical data	Query complexity	Relevance of currency in queries	Economic value of integration	Relevance of queries wrt transactions	Volatility of queries	Management complexity	Costs of heterogeneities	Preferred solution
-	High	-	-	-	-	-	-	-	Data Warehouse
High	-	-	Low	High	-	-	-	-	Publish & Subscribe
Low	-	-	-	High	-	-	High	High	Consolidation
Low	-	-	-	High	-	-	High	High	Consolidation
Low	-	-	-	High	-	-	High	High	Consolidation
High	-	-	High	High	High	High	-	High	Data integration
High	-	-	-	-	High	-	-	-	Data integration
High	-	High	-	High	-	-	-	-	Data Warehouse

Fig. 13.14 Decision table for choosing the optimal data integration architectural solution

2. the three databases related to
    - agricultural state property
    - agricultural production
    - veterinary services

    all need a common data warehouse.

13.7 Summary and Conclusion of the Book

The overall activity discussed in this chapter has shown that the eG4M methodology is a good point of equilibrium between simplicity of application and effectiveness of decisions. We also have to underline that, as said previously, before moving to production, feasibility studies are needed whose goal is to deepen the technical analysis and evaluate costs and benefits of the project more precisely.

Finally, at learning process level the eG4M methodology-related workshops held in Tunisia fulfilled the share of IT skills among the participants, improving knowledge and computer literacy of some attendees with no previous IT capability. In terms of absorptive capacity, the eG4M methodology has successfully supported the assimilation by the organization of IT-related knowledge [82, 233].

In conclusion of the book we may say that eG4M, as expressed in the title of the book, is a methodology suitable to be used in a wide spectrum of strategic and operational planning activities for service-oriented information systems. Its modular approach makes easy an adaptation of the methodology to specific and more focused needs of the public administration, as shown by the case study discussed in this chapter. Clearly, every methodology needs a suite of tools in order to be effective. For we are currently developing, experimenting, and using such a suite of tools as a business activity of NexTTLab, a spin-off of the University of Milano Bicocca (Italy).