Dual Approach to Organizational Data Planning

D. Pascot  D. Kidjanovic  R.W. Mantha  A. Marcel
Laval University  Faculty of Business Administration  Sainte-Foy, Québec, Canada, G1K 7P4  Zero-time, KPMG Peat Marwick  1175 avenue Lavigerie, Suite 090  Sainte-Foy, Québec, Canada, G1K 4P1

ABSTRACT

Two major approaches to data modeling, view modeling and integration, and global data modeling, are compared and their consequences to organizational data planning are presented. The view modeling and integration approach is based on the analysis of application requirements, i.e., it is "needs-driven", while global data modeling is based on the synthesis of primary data elements (data that supports the basic business operations) into entities and relationships, i.e., it is "primary data-driven". On one hand, if the environment changes, the views approach may not produce a stable data model, on the other hand, the global data model, based on the primary data supporting the basic business operations, may not provide a complete support for strategic information systems. Dual approach to organizational data planning combines the two approaches by minimizing their disadvantages and emphasizing their positive features.

Data Modeling Alternatives

There are two major approaches to data modeling: view modeling and integration, and global data modeling. View data models are defined by applications requirements [1]. Even within the same application there may be several views, depending on users, contexts, management objectives, etc. The process of view integration resolves inconsistencies among different views and produces, first, application data models, second, department data models, and last, an organization data model. The advantage of this approach is that it lends itself easily to well recognized partitions, like applications, departments, etc. Its main disadvantage is that the process is time consuming and the scope is difficult to bound—there are many different organizational views that frequently change when application requirements change, and application requirements are often dependent on the environment of an organization.

In global data modeling, view data models are not explicitly recognized and there is no separate process of view integration. A global data model is derived from the data supporting the basic business operations—primary data [2]. Primary data elements are found in forms handled within basic business operations. Entities and relationships are then conceptualized and added to one global data model. The advantage of this approach is that the scope of modeling is clearly defined by the forms used in basic business operations. Its main disadvantage is the lack of convenient partitions that would help in the conceptualization process and the possible omission of some strategically important data.

In summary, the view modeling and integration approach is based on the analysis of application requirements, i.e., it is "needs-driven", while global data modeling is based on the synthesis of primary data elements into entities and relationships, i.e., it is "primary data-driven". On one hand, if the environment changes, the views approach may not produce a stable data model, on the other, the global data model, based on the primary data supporting the basic business operations, may not provide a complete support for strategic information systems.

Dual Approach

An organizational data plan that supports either the view modeling and integration approach or the global data modeling approach would risk providing an incomplete data basis for information systems development activities.
However, supporting both approaches may be more than any organization can afford. A practical solution to the presented dilemma is to combine the two approaches by minimizing their disadvantages and emphasizing their positive features.

The view modeling and integration scope can be bounded by modeling only those views that directly support critical decision making using the process of critical success factors (CSF) analysis. Those views would serve as a basis for decision support systems development and as a validation mechanism for the global data model. The application data models would be validated against the global data model in the process of applications development.

The global data modeling approach lacks a convenient way of partitioning the model building process. Guidelines for classifying sources of primary data, basic business operations and their forms would help in the conceptualization process.

Boundary Issues

As with any large-scale effort, determining the boundary of the ODP task is a crucial step. There are two types of data that are of interest to an organization—internal and external data.

Internal data can be divided into primary data and data views. Data views can be obtained by rearranging primary data (simple views) or by deriving them from primary data (aggregate views). ODP should be concerned only with primary data, because the ratio of primary data over data views is, in general, closer to zero than one, since there can be many more ways of viewing data elements than there are data elements. The assumption is that data views can be derived, in one way or another, from primary data, and that primary data can be derived from the forms handled within basic business operations.

External data are environment dependent and their scope is practically impossible to determine a priori. Since many CSFs support organizational strategic decisions, their decision support indicators and underlying data models would probably include some external data. In this way, the scope of the external data is determined indirectly by CSFs.

Partitioning Mechanisms

There are two major ways of partitioning a problem: levels of abstraction and classifications within the level. CSF data models, application data models, and the global organization data model are at three different levels of abstraction. At every level, objects of interest may be classified according to different criteria. The following are some examples of classification criteria:

- organizational structure
- business functions
- interest groups represented by major actors
- decision making indicators
- products/services
- production processes
- major types of data (primary, derived; primary: stable, dynamic)

- major types of data flows (financial, materials, personnel, assets)
- nature of data flow transformations (time, space, form).

Classification may be used to partition a domain of consideration into meaningful units in order to facilitate the domain comprehension and discovery/conceptualization process. For example, critical success factors may be classified according to business functions and interest groups. Basic business operation forms may be divided into major categories based on products and production processes. Primary data sources may be classified by the types of objects exchanged between the external and internal environments of the organization.

Business data can be broadly classified as stable (reference), dynamic (transaction), and derived (management reporting). Stable data represent the relatively stable objects of the business, its resources, products, and assets (e.g., employees). Dynamic data represents the business transactions (e.g., orders). Stable and dynamic data are two major categories of primary data. Derived data usually represents measurements (e.g., cost per unit).

Stable data has a relatively long life and low frequency of updates, high read-only activity, high participation in user views, and an identifiable organizational owner, i.e., a single update responsibility.

Dynamic data has a relatively short life and high frequency of updates, often multiple reference relationships, usually multiple owners, and is the basis for derived data.

Derived data is time dependent. Its life is relatively short and its characteristics are determined by the derivation or reporting rules.

Reference Points

The assumption that organizational data reflects what an organization does, as opposed to how it does it, and that ways of doing things change more frequently than the basic nature of the organization's business, provides an opportunity to create generic, business type data models. A generic data model would represent a state-of-the-art summary of practical experiences and theoretical advances in a particular type of business or industry, and serve as a good point of reference in the process of data planning and data modeling.

Another important point of reference is the existing collection of organizational files. A problem with organizational files is their high degree of data redundancy and unnormalized structures, in addition to nonstandard and often ambiguous naming conventions (if such conventions exist at all). A lengthy list of existing data elements may serve as a good point of reference in the process of data modeling. An important task of comparing the global data model and existing file descriptions should be done in order to assess the degree of difficulty in moving from a file to a database environment.
Organizational Aspects

If data is considered as an organizational resource, its organizational importance cannot be overemphasized. Many businesses have made the mistake of taking the narrow view that data is solely the domain of the information systems department. Recently, many companies have realized that the data resource may provide that long desired competitive edge.

ODP must establish strong links with organization strategic planning, information systems strategic planning, and data technology support. One of the ways to acknowledge the importance of the data resource management is to create a data administration function that is clearly distinct from a database administration function [5]. The data administration function should be ideally located outside of the information systems department and be closely related to the top management. In this way, a crucial top management support would be provided, and the technology side of the data administration function would be properly deemphasized within the function.

Data resource management requires the cooperation of many departments that often have competing interests. Even with the top management commitment, cooperation may be difficult to obtain. Also, the benefits of ODP may be difficult to project. The explicit payoffs usually come later when new application development efforts take advantage of the global data model.

Another barrier to making data resource management work is the credibility of information systems departments. A typical information systems department would have a technology, not a business orientation. It would be unresponsive to demands for new applications and it would resist changes.

Data administrators are charged with the critical responsibility of planning, establishing and maintaining control over the organizational data resource. The major product of their work is the organization global data model and CSF data models, that must be continuously maintained. Both types of data models should be technology independent. Specific application data models should be the responsibility of the application users and the application development function. Technology support for all data models should be the responsibility of the database administration function. The data administration function should establish positive working relationships with the application development and database administration functions.

Technological Aspects

Data resource management requires good technology support. The expertise in data modeling is crucial, but without the technology support the expertise can be wasted on paper work. There are two major types of tools important for data modeling productivity--graphical support for data modeling and data dictionary support for data resource management.

A graphical data modeling tool should support the entity-relationship type of data modeling, be capable of representing large data models, be helpful in improving quality of data representations, and provide the raw data for a data dictionary.

A data dictionary should support organization data standards, provide a comprehensive collection of data resource management reports, and be an active link between the data administration, database administration and application development functions.

Sequencing of Steps

Most data planning approaches are "needs-driven". Determining information needs of major business functions is one of their priorities. The following steps outline, in a general way, the major steps of the "needs-driven" approaches.

1. Determine information needs of the major business functions:
   1.1 identify the major business functions;
   1.2 identify the major stakeholders in each function;
   1.3 organize interviews with the major stakeholders to determine data classes created and used by the business functions (questions asked are related to business problems, potential solutions, critical success factors, efficiency, effectiveness, current and projected products, services and markets);
   1.4 create a business function/data class matrix;
   1.5 package data used or created together into subject databases.
2. Define the applications portfolio:
   2.1 determine priorities for the development of subject databases and applications;
   2.2 determine business functions responsible for the planned development.

Figure 1 presents major outcomes of the "needs-driven" data planning approaches.
A major problem with "needs-driven" data planning approaches is the quality of the data classes determined by the process of analyzing business function needs. Data classes are in general too abstract to be useful for database development. That is one of the reasons why most data planning documents end up sitting on a shelf. If the decision is made later on to refine data classes into entities and relationships, this is usually done within the view modeling and integration framework, involving many people. Consequently, projects drag on, and often do not get completed.

On the other hand, the "primary data-driven" planning approach has a clear objective—to produce the global data model that includes only the primary data, without elaborating on business function needs. The following steps outline, in a general way, the major steps of the "primary data-driven" approach.

1. Identify organizational primary data:
   1.1 find external entities by examining the interface between the organization and its environment (a context data flow diagram is usually helpful);
   1.2 determine data flows to and from the external entities (a classification of data on financial, material, personnel and asset flows helps in partitioning the problem);
   1.3 determine the basic business operations that generate transaction data (recording the exchanges of flows in 1.2);
   1.4 collect forms of the basic business operations;
   1.5 consider discarding forms not used for some period of time;
   1.6 select most frequently used forms (the old 20%-80% rule usually holds: only 20% of the forms generate 80% of the primary data);
   1.7 make a list of the primary data items from the forms selected in 1.6.

2. Conceptualize an entity-relationship diagram of the global data model:
   2.1 create potential entities by selecting candidate identifiers;
   2.2 focus efforts on stable entities and their relationships (a generic data model may be very helpful);
   2.3 include dynamic entities and their relationships;
   2.4 assign all primary data left either to entities or relationships;
   2.5 validate the data model against the forms used for modeling and a sample of additional forms initially discarded.

3. Define the applications portfolio:
   3.1 create a business functions/global data model matrix;
   3.2 package data used or created together into subject databases;
   3.3 determine priorities for the development of subject databases and applications (e.g., based on data creation/usage precedence);
   3.4 determine business functions responsible for the planned development.

Figure 2 presents major outcomes of the "primary data-driven" data planning approach.

A problem with the purely "primary data-driven" approach is a possible omission of some strategically important primary data which is not in current transactions. In order to include strategically important data, a dual approach to organizational data planning would add the "needs-driven" part of the CSF analysis to the "primary data-driven" approach. CSFs are the limited number of areas in which results, if they are satisfactory, will insure successful competitive performance for the organization. In other words, they are the "short list" of the few things which must go absolutely right. CSFs are obtained by interviewing senior managers, starting with their goals. Once CSFs are determined, they must be related to information needed to monitor the desired performance, i.e., decision making indicators must be established. Indicators are almost always derived data whose roots must be traced to the primary business data. CSF data models include all the primary data and derivation rules necessary for the reconstruction of indicators. In that way, CSF data models serve as a strategic validation of the global data model. The primary data needed to support CSFs often cross functional/departmental boundaries. Therefore, a validation of the global data model will indicate how the data partitioning should not be done if CSF support is to be adequately provided.

Figure 3 presents major outcomes of the dual approach to organizational data planning.
The very general objective of Information Resource Management (IRM) is managing information as an organizational resource. As such, IRM encompasses organizational data planning, strategic information systems planning, and systems development (applications development and database administration). Figure 4 shows the place of ODP within the IRM framework. Arrows represent major influences of the IRM building blocks on each other. ODP, with its global data model, may provide an excellent basis for information systems planning. On the other hand, CSF analysis, done in the strategic information systems planning, is a necessary precondition for ODP.

**Organizational Data Planning Within Information Resource Management**

In summary, an organizational data planning should provide a sound basis for future competitive informations systems, for reports that cross departments, for business functions and applications boundaries, and for the effective control of data resources. CSF data models would hopefully support future actions, while the global data model of primary data would be the essential basis for the integration of existing systems.

**Conclusion**

Most organizations will not be able to survive without strategic information systems. Strategic information systems cannot be developed without strategic information systems planning. Strategic information systems must rely on internal and external organizational data. In most organizations data are handled at the physical, file level and...
Strategic Information Systems Planning

- objectives, critical success factors (CSF), indicators, decisions
- problems, solutions, effectiveness, efficiency, products, services

Organizational Data Planning

- CSF data models

Applications Development

- business functions
- reports, calculations
- updates
- data views

Database Administration

- integrity constraints maintenance
- logical data model
- database

primary data sources, data flows, basic business operations, forms

Figure 4. Information Resource Management

are not well understood at the conceptual, organizational level. A global data model is a mechanism to represent organizational data at the planning level without introducing unnecessary, computer based details. A global data model may be obtained from the analysis of applications requirements, application data modeling and integration. However, this approach is so time consuming that most organizations cannot afford to apply it. Another approach, based on the business primary data, can produce a global data model in much shorter period of time. Coupled with the derived data that supports critical success factors, the approach provides efficiently the abstract representation of internal and external organizational data needed in the process of strategic information systems planning and development.

REFERENCES

