The Impact of Data Integration on the Costs and Benefits of Information Systems

By: Dale L. Goodhue
    Information and Decision Sciences
    Carlson School of Management
    University of Minnesota
    271 19th Avenue South
    Minneapolis, Minnesota 55455
       U.S.A.

    Michael D. Wybo
    Faculty of Management
    McGill University
    1001 Sherbrooke Street West
    Montréal, Québec
    Canada H3A 1G5

    Laurie J. Kirsch
    Information and Design Sciences
    Carlson School of Management
    University of Minnesota
    271 19th Avenue South
    Minneapolis, Minnesota 55455
       U.S.A.

Abstract

For many organizations, the ability to make coordinated, organization-wide responses to today's business problems is thwarted by the lack of data integration or commonly defined data elements and codes across different information systems. Though many researchers and practitioners have implicitly assumed that data integration always results in net benefits to an organization, this article questions that view. Based on theories of organizational information processing, a model of the impact of data integration is developed that includes gains in organization-wide coordination and organization-wide decision making, as well as losses in local autonomy and flexibility, and changes in system design and implementation costs. The importance of each of these impacts is defended by theoretical arguments and illustrated by case examples. This model suggests that the benefits of data integration will outweigh costs only under certain situations, and probably not for all the data the organization uses. Therefore, MIS researchers and practitioners should consider the need for better conceptualization and methods for implementing "partial integration" in organizations.

Introduction

A common language for communicating about business events is a prerequisite for coordinating diverse and far-flung units of organizations. In the realm of computer information systems, one form this common language can take is data integration, or data elements with standard definitions and codes. Researchers studying the impact of information technology on organizations often assume that the common language provided by data integration exists, or that it will be developed because of the benefits of increased communication within (or across) organizations (e.g., Huber, 1990; Malone, et al., 1987).

However, there is evidence that this common language of logically compatible data does not exist in a great many large organizations today (Goodhue, et al., 1988). Within a single company there are often different identifiers for key business entities, such as customer or product, different schemes for aggregating key indicators, such as sales or expenses, or different ways of calculating key concepts, such as profit or return on investments. These inconsistencies cause major problems when firms ask questions that span multiple systems or multiple subunits, thwarting their ability to make coordinated, organization-wide responses to today's business problems.

Researchers and practitioners have responded to this problem by developing ways to increase data integration. For example, network and relational data structures that could accommodate the varied needs of many users in a single integrated data structure have been described

(Bonczek, et al., 1978; Date, 1981). The entity relationship model has been proposed as a means of conceptually modeling all of a corporation's data, providing the basis for integrated computer systems (Chen, 1976; McCarthy, 1982). Other researchers have recognized that organizations do not start from a clean slate, and existing systems typically are not integrated. Thus, theoretical approaches for integrating existing, disparate database schemas have been developed (Batini, et al., 1986). Still others have focused on developing practical methods for immediate use in actual organizations, including information engineering methodologies that would provide an information architecture by which organizations could transform their systems from non-integrated to integrated forms (Finkelstein, 1990; IBM, 1981; Martin, 1982; 1986). In general, the presumption has been that greater data integration is universally desirable and leads to greater benefits for organizations.

In spite of the conceptual appeal of these methods for achieving data integration, many organizations that have attempted them have failed or experienced major difficulties (Goodhue, et al., 1988; 1992; Hoffer, et al., 1989; Lederer and Sethi 1988; 1991). This raises an important question: If data integration is universally desirable, why is it so difficult to implement in practice? Existing research has offered two general explanations. The first considers implementation pitfalls of data integration, such as the difficulty of obtaining top management support or the need for managing expectations (Goodhue, et al., 1988; Hoffer, et al., 1989; Lederer and Sethi 1988; 1991). A second explanation is the possibility of shortcomings in the methodologies used (Goodhue, et al., 1992).

This article considers a third possible explanation. Counter to the usual presumption, data integration efforts may fail because in certain organizational contexts they do not provide sufficient benefits to offset their costs. This conclusion emerged gradually as we pondered the evidence of over 35 case studies of data management efforts conducted at the Massachusetts Institute of Technology (MIT) and the University of Minnesota over the past eight years.[1] The

cumulative experience represented by those cases strongly suggests that the net benefits of increased data integration will not always be positive but will depend upon three main organizational factors: (1) the interdependence of subunits, (2) the need for locally unique or flexible action by subunits, and (3) the difficulty of designing and implementing systems with integrated data.

First, theories of organizational information processing (Daft and Lengel, 1986; Galbraith, 1973; Tushman and Nadler, 1978) are discussed to show that the same three factors that emerged from the case-study evidence can be derived from an existing theory base. Next, each of the three factors is explored further. They are defended by additional theoretical argument and illustrated with selected anecdotal evidence drawn from the case studies. This leads to several propositions for empirically testing the main tenets of the new model in future empirical work. The discussion begins with an overview of the data integration problem in the context of large organizations.

What Is Data Integration?

Data integration generally means the standardization of data definitions and structures through the use of a common conceptual schema[2] across a collection of data sources (Heimbigner and McLeod, 1985; Litwin, et al., 1990). Data integration ensures that data have the same meaning and use across time and across users, making the data in different systems or databases consistent or logically compatible (Martin, 1986).

The operational definition used here focuses on logical compatibility at the level of the individual data element. We define data integration as the use of common field definitions and codes across different parts of the organization.[3] This defini-

---

[1] The case examples come from large ($500 million and above) U.S. or Canadian firms. All company names are disguised. For more details see Goodhue, et al. (1988) and Goodhue, et al. (1992).

[2] A conceptual schema specifies field and record definitions, structures, and rules for updating data values.

[3] An alternate approach to hammering out agreement on definitions and codes across all systems would be to allow each subunit to design its systems autonomously, and then meld these disparate databases into a federated database system (Heimbigner and McLeod, 1985; Sheth and Larson, 1990) where presumably at least some data could be translated or mapped to permit limited sharing. This leaves the sharability of data somewhat up to chance, because without central planning and discipline the actions of any single subunit could negate the sharability of any data element.

tion lets us easily conceptualize the degree of data integration in a firm. Beyond none at all, the least amount of data integration is a common definition and set of codes for a single field across two databases or information systems. Data integration can be increased along one or both of two dimensions: (1) the number of fields with common definitions and codes, or (2) the number of systems or databases that adhere to these standards.

Table 1 shows three examples of integrated and non-integrated data for two divisions of a single company. In the first example, both divisions in the integrated environment have the same PART _ NUMBER code in all systems for every part used in their products. In the non-integrated environment, the two divisions have not coordinated the assignment of PART _ NUMBER codes. Though it is possible to translate from one division's part numbers to another's, if there are many part numbers it is quite difficult to determine the mapping (and keep it up to date!).

In the second example, both divisions in the integrated environment use a single code to identify customers. In the non-integrated environment, Division B has a two-level structure for customers, so separate locations that order and pay for their own purchases can be given separate customer identifiers, even if they are really part of the same large company. With this arrangement it would not be possible to translate data from Division A into the form used by Division B. Finally, the third example shows that different divisions can have different ways of handling adjustment (e.g., returns or discounts) in their sales figures. In such situations, translation of the data to a common form is often possible, but what might be a mere inconvenience when a single data element is involved becomes a serious problem when there are tens or hundreds of such data elements that need to be shared. The difficulties are magnified when the individuals making the translation are not aware of the differences in the definitions or cannot determine exactly what has been included or excluded in calculating a particular field.

A concrete example will assist us in understanding what data integration or the lack of it might mean in an organization.

The Van Buren Bank has felt the effects of deregulation, which has made the once stable banking industry highly competitive. With the decreased spread between borrowing and lending rates, profits on loans have dwindled, making profit on services to customers critical. In the corporate banking group, account managers who in the past could concentrate only on loan volumes must now focus on customer and product profitability. This means they must make decisions differently and need different kinds of information. For example, if a long-time customer threatens to take his or her business elsewhere unless he or she is given an unusually low interest loan, the account manager must decide whether this is an otherwise profitable customer in terms of his or her total business with the bank.

In order to determine how profitable a customer is, the account manager must gather information about the various products and services the

Table 1. Examples and Counter Examples of Data Integration

	Integrated Environment		Non-Integrated Environment	
	Division A	Division B	Division A	Division B
1. PART _ NUMBER Codes for a 3/4" SCREW	115899	115899	115899	337189
2. CUSTOMER _ ID Codes for IBM	42765	42765	42765	42675, 49345, 51349, etc.
3. Definition for SALES	Not adjusted for returns	Not adjusted for returns	Not adjusted for returns	Adjusted for returns

customer buys and the profitability of each. Conceptually this could be done by communicating with other account managers around the world who also do business with this customer (or communicating with their electronically readable records) and consolidating the information.

However, it may or may not be true that the customer is identified in the same way in each account manager's records; or that the various products and services the bank sells are identified the same way or grouped into the same categories; or that the size of each account is recorded in the same currency; or that discounts, refunds, and so forth are handled in the same way. The account manager must translate all this information from many sources into a common form and determine how profitable this customer is.

> Unfortunately at the Van Buren Bank, all customer identifiers were typically assigned by the branches and were not standard across the bank. Therefore, there was no way to identify all the business of a given customer short of phoning up every branch and asking. It was clear that the Van Buren Bank required much more data integration than was currently built into its information systems.

Few large international banks would be willing to require every account manager (regardless of product type, customer type, or country) to record all his or her information in exactly the same format. On the other hand, neither would such banks be likely to permit each local account manager to record all information to be most helpful to him or her individually, with no effort at standardization. Most banks would likely choose a middle road. This "partial integration" might be achieved by standardizing on certain data elements (perhaps key identifiers, such as customer and product) and allowing flexibility on certain other data elements. Such a "partial integration" solution has the advantage that the most frequently used and troublesome data could be standardized, while leaving some flexibility to local account managers in defining other data less critical for cross-bank questions. But determining how many and which data elements to standardize then becomes an important and difficult data management question. One way to address this question is to consider the impact of increasing levels of data integration on the costs and benefits of information systems.

A Model for the Impact of Data Integration

The model to be developed here depends upon a very *rational* view of organizations and organizational design. It assumes that benefits and costs will be summed at the level of the organization and that everyone in the organization shares the goal of maximizing overall benefits minus costs. In a later section of this article several factors are considered that do not fit within this very rational view, including power and politics, and the interdependence between data integration, strategy, and structure. In spite of these caveats, the rational model provides a powerful conceptual framework for thinking about the issue of data integration.

Costs and benefits of IS

The idea that a careful conceptualization of costs and benefits can lead to useful insights about design choices for information systems (IS) is not new. As early as 1959, Marschak conceptualized the differences in information system designs as different ways of selecting and partitioning the available data from the environment. Information systems can be seen as having costs (design and implementation costs to the provider of the system) and benefits (improved decision making and more informed actions for the user of the information), which can vary quite independently. Organizations face an array of possible information systems, each with certain implementation costs and different expected decision-making benefits. Rational organizations should choose a design that maximizes the benefits minus the costs (Emery, 1982; Marschak, 1959; 1968; 1971; Mendelson and Saharia, 1986).

To look at data integration in this light, a clear conceptualization is needed of how integration affects both the benefits and the costs to users and implementers. Organizational information processing theory (Daft and Lengel, 1986; Galbraith, 1973; Tushman and Nadler, 1978) gives us a basis for such a conceptualization.

Organizational information processing

Organizational information processing theory is concerned with the design of organizations, in

particular the design of structures or mechanisms to deal with information processing requirements. Uncertainty, a central concept in the theory, drives the need for information processing. Galbraith (1973) proposes that when uncertainty in an organization is low, three different mechanisms can be used to resolve it: rules and procedures, hierarchies, and goals. However, when an organization faces greater uncertainty than can be handled by these three mechanisms, it must either reduce the need for information processing (through slack or self-contained tasks), or it must increase information processing capacity (through computerized information systems[4] or lateral relations).

While recognizing that uncertainty can be dealt with by any of Galbraith's options, this paper focuses primarily on his option of computerized information systems. Such systems differ along several dimensions, two of which parallel nicely the concept of data integration discussed in this paper: the degree of formalization of the data and the scope of the database (Galbraith, 1973). Data integration (common field definitions and codes) is an example of a highly *formalized language* for describing the events occurring in an organization's domain. The *scope* of data integration is the extent to which that formal language is used across multiple functions or subunits of the organization. Thus, the degree of data integration of a computerized information system can be seen as an important design characteristic in dealing with organizational uncertainty. To understand when data integration is especially desirable, we must delve deeper into the sources of uncertainty facing the organization.
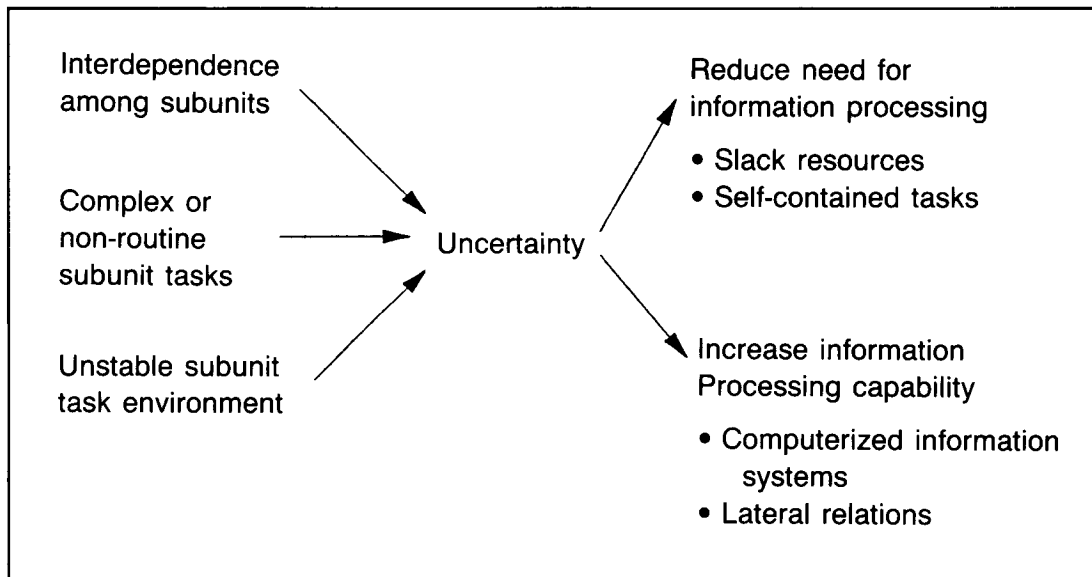
Sources of uncertainty

For Galbraith, the uncertainty that motivates choices among information processing mecha-

---

[4] Galbraith uses the term *vertical information systems* rather than computerized information systems, and his examples typically refer to systems providing information vertically up the hierarchy to managers responsible for all units involved. However, he makes it clear that the essential aspect of this alternative is the provision of information to those who need it: "It should be pointed out that taking information from its points of origin and collecting it into a global data base to be presented at a level high in the hierarchy is only one method of operationalizing this dimension. . . . However, the policy dimension is to increase the scope of the data base available to the decision mechanism independent of what it is or where in the hierarchy it is located" (Galbraith, 1973, p. 33).

nisms is defined in general terms for the whole organization, without any detailed description of where it comes from. Tushman and Nadler (1978) pushed the analysis down to the level of the subunit and distinguished between three different sources of uncertainty: complex or non-routine subunit tasks, unstable subunit task environments, and interdependence between subunits. For example, a manufacturing unit employing a temperamental, thin film manufacturing technology might face uncertainty in the operation of this highly variable process (complex tasks), uncertainty from the environment where customer demands were constantly changing (unstable environment), and uncertainty in dealing with a procurement office that did not always understand the urgency of its requests (interdependence between subunits). Figure 1 combines the ideas of Tushman and Nadler (1978) with those of Galbraith (1973), showing the three sources of uncertainty.

Tushman and Nadler argue that subunits in a single organization may face quite different task characteristics and task environments and therefore might need quite different information processing mechanisms. For example, the thin film technology manufacturing unit might need quite different information processing capabilities from a more traditional assembly line subunit facing a stable environment, even though both are in the same firm. In addition, the most appropriate information processing mechanism for a given subunit could change over time if task characteristics or task environments change. Thus, individual subunits might desire the autonomy and flexibility to design and change their information systems independently.

If we look at data integration in this light, we can see that its impact might differ depending on which *source* of uncertainty dominated. Consider the three subunits mentioned above: the thin film manufacturing unit, the assembly line unit, and the procurement office. Where uncertainty comes mainly from interdependence between subunits (for example, when the two manufacturing subunits must deal with the procurement unit to purchase material), data integration would be highly desirable because it provides a standardized, formalized language shared by all subunits and facilitates communication between the interdependent subunits. On the other hand, when uncertainty comes mainly from task complexity or

**Figure 1. A Model of Organizational Information Processing
(Combining Galbraith, 1973, and Tushman and Nadler, 1978)**

environmental instability unique to individual subunits (for example, the temperamental process and changing demand faced by the thin film manufacturing unit but not faced by the other units), mandatory data integration might reduce the flexibility of an individual subunit to redesign its information systems to address its unique needs. Here a high level of data integration could be undesirable.

Uncertainty and equivocality

Recent additions to organizational information processing theory take the concept of uncertainty as used by Galbraith (1973) and Tushman and Nadler (1978) and break it into information requirements of two types: uncertainty and equivocality (Daft and Lengel, 1986). This distinction can help us understand when computerized information systems (whether with data integration or not) may be of little assistance in meeting a firm's most pressing information needs. *Uncertainty* by this new definition is the absence of specific, needed information. For example, a manager might want to know *whether* sales of widgets dropped in the southern region last month. The question is clear, and given data on specific variables, the uncertainty would be removed.

Equivocality means there are multiple, conflicting interpretations of a situation. When equivocality is present, the precise information needed to resolve the situation is not clear. For example, the same manager might want to know *why* sales dropped in the southern region last month. For such a question, a number of types of information might be appropriate, including the attitude and behavior of the sales force, the actions of competitors, the economic situation, trends in customer satisfaction, the weather, etc. Competing perspectives might be as useful to the manager as factual data.

In general, uncertainty can be reduced by a sufficient *amount* of information, while equivocality can be reduced by sufficiently *rich* information (Daft and Lengel, 1986). Certain types of information processing mechanisms (such as computerized information systems and, by implication, systems with integrated data) provide large amounts of information and can thus help reduce uncertainty. However they are not as rich a source as other information processing mechanisms (such as face-to-face meetings) and thus are not as effective in reducing equivocality.[5]

---

[5] Note that in general it is not necessary to choose between computerized information systems and rich informal systems—both exist side by side and, ideally, support each other.

Therefore, the appropriateness of information systems with integrated data might depend on the degree to which unmet information needs are characterized by uncertainty as opposed to equivocality. Using the framework provided by Tushman and Nadler, Daft and Lengel carry their analysis down a level to look at each of the three sources of information requirements: interactions between subunits, complex subunit tasks, and unstable subunit task environments.

First, consider information requirements that result from interaction between subunits. Interdependence between subunits leads to a need for greater amounts of information, and differentiation between subunits leads to a need for richer information (Daft and Lengel, 1986). Applying these ideas to data integration, if interdependence is high (for example between the two manufacturing units and the procurement office) the formalized language of integrated data will allow large amounts of information to be easily shared. But if subunits are highly differentiated (for example the thin film technology unit and the assembly line unit), different perspectives, values, or frames of reference may lead to high equivocality and make it difficult to implement data integration because it would be more difficult to agree on common definitions for data items and the events they describe. This suggests that data integration would probably be most useful where subunits were very interdependent and most easily implemented where subunits were not highly differentiated.

Now consider the other two sources of uncertainty: complexity of the subunit's tasks and stability of the subunit's task environment. In general a lack of analyzable cause-and-effect relations leads to a need for more "rich" information processing mechanisms (such as face-to-face meetings) to reduce equivocality, while variety and constant change lead to a need for greater amounts of information processing (using mechanisms such as computerized information systems) to reduce uncertainty (Daft and Lengel, 1986). Where organizations face great equivocality as opposed to uncertainty, data integration and the formalized language it implies (as well as computerized information systems in general) may not be as useful in addressing information processing needs.

Using this logic we might expect that information systems with integrated data would be quite useful in, for example, organizations in the snack food industry in the U.S., where new products and changing consumer tastes make for variety and constant change but the basic ground rules are well-understood. On the other hand, systems with integrated data might be less useful (and face-to-face meetings more useful) in organizations trying to gain a foothold in the new markets of Eastern Europe, where cause and effect relationships are ambiguous and critical success factors are unclear and constantly shifting.

Balancing benefits against implementation costs

Finally, both Galbraith (1973) and Tushman and Nadler (1978) emphasize the importance of balancing greater effectiveness of more complex information processing mechanisms against their greater costs: "[T]he basic design problem is to balance the costs of information processing capacity against the needs of the subunit's work—too much capacity will be redundant and costly; too little capacity will not get the job done" (Tushman and Nadler, 1978, p. 619).

While organizational information processing theory does not give much detailed help in conceptualizing the impact of data integration on the costs of design and implementation of systems, we can resort to arguments on design difficulty in general (Simon, 1981) and the design and implementation of information systems in particular (Banker and Kemerer, 1989; Brooks 1975; Martin, 1982). These suggest that data integration could have a positive impact on reducing costs by reducing redundant design efforts. However, because multiple subunits would be involved, data integration could also have a negative impact on costs by increasing the size and complexity of the design problem or increasing the difficulty in getting agreement from all concerned parties.

If we consider only those situations where uncertainty dominates and information systems in general are very appropriate choices, this suggests that the impact of data integration on the costs and benefits of information systems will come primarily through three potential factors: (1) increased ability to share information to address subunit interdependence, (2) reduced ability to meet unique subunit information requirements, and (3) changes in the costs of in-

formation system design and implementation. This is illustrated in Figure 2. The next three sections of this paper look in more depth at each of these potential impacts.

Factor 1: Benefits From Integrated, Sharable Data

In his discussion of the application of information technology to organizational design, Simon (1973) suggests that "there is no magic in comprehensiveness....The mere existence of a mass of data is not sufficient reason for collecting it into a single comprehensive information system" (p. 271). The major benefits of "a single comprehensive information system" come from the ability to share or aggregate information across many divisions or functions or units of the organization. The need for sharing information is greatest where actions and events across those units are interdependent, that is, where the actions of one subunit affect the actions or outcomes of another subunit (McCann and Ferry, 1979).

When there is interdependence but not too much difference in perspective between the subunits,

a formalized information system (including data integration) will be an appropriate mechanism (Daft and Lengel, 1986). We can distinguish at least two different impacts data integration might have on an organization: (1) improved managerial information for organization-wide communication and (2) operational coordination between interdependent parts of the organization (Goodhue, et al., 1988).

Improved communication across subunits

Data integration is necessary for data to serve as a common language for communicating about events the organization faces (Galbraith, 1973). Without data integration there will be increased processing costs and ambiguity of meaning as messages between subunits are processed. The result should be delays, decreases in communication, reductions in the amount of summarization, and greater distortion of meaning (Huber, 1982). Thus, data integration facilitates the collection, comparison, and aggregation of data from various parts of the organization, leading to better understanding and decision making when there are complex, interdependent
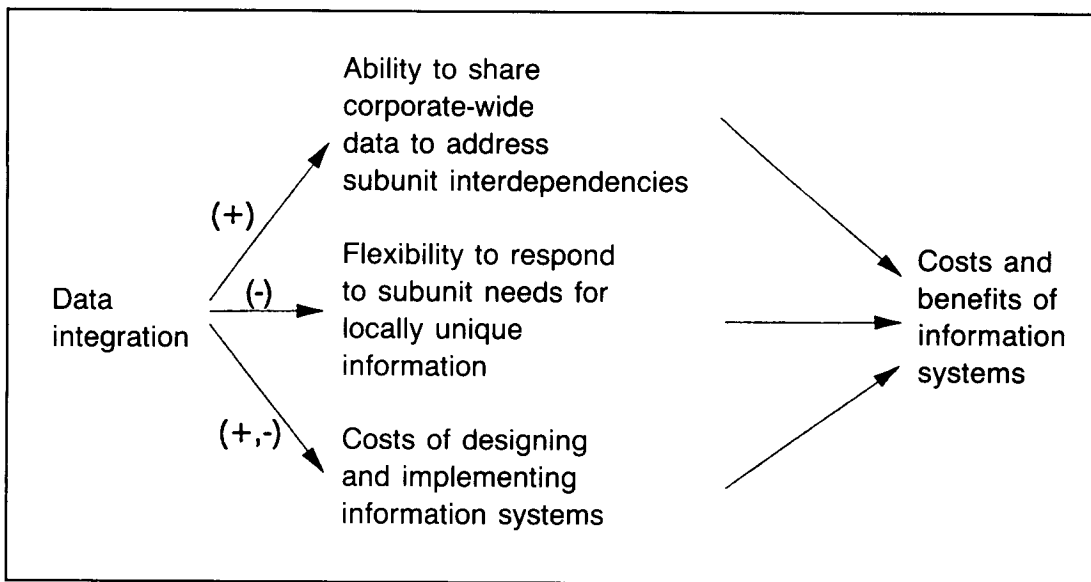

Figure 2. The Impact of Data Integration on the Costs and Benefits of Information Systems

problems. Two examples from our case studies illustrate this.

> When on-time deliveries (a critical competitive issue in the semiconductor market) fell to only 70 percent, a multi-disciplinary team at Devlin Electronics used organization-wide integrated scheduling data to track how production schedules were made, changed, and adjusted by the many different groups involved. They found a number of interlocking problems: some plants were not properly updating their inventory levels and equipment conditions, marketing was overriding the schedule without regard to plant capabilities, and plants were overriding the system without regard to critical order requirements. Organization-wide integrated data allowed Devlin to understand its problems and take corrective action. On-time delivery improved from 70 percent to 98 percent.
>
> At Southern Cross, Inc., the executive vice president expressed concern at hearing that in spite of many new accounts, sales were up by only one-half percent. He complained that, "It's obvious we are bleeding somewhere, but nothing in these standard reports gives me any indication of what the problem is." He demanded an analysis across all regions, all customers, and all products to determine the cause. Because the information was contained on several different (unintegrated) systems, and because products had over time been grouped into various different (and shifting) categories for reporting purposes, there was no way to do the analysis in an automated fashion. Using spreadsheet programs and manually backing out products that had switched categories and adjusting other inconsistencies between the different systems, top-level analysts were able to assemble a compatible base of information to answer the EVP's questions after 40 person-hours of effort. The analysis indicated the sales slump was occurring primarily in the old, established, long-time distributorships (an insight that was not apparent from analysis of the non-integrated data). With the problem somewhat pinpointed, managers could take appropriate further action.

One possibility is that the Southern Cross systems were not well-designed in the first place. However, Southern Cross has for years enjoyed great strategic advantage from its systems. Since the need for this type of cross-product line analysis had not surfaced in the past, its systems were not designed to address it.

The temptation might be to design enough data integration into all our systems to be able to answer any question that might come up in the future. However, as seen later in this paper, the very real benefits of data integration must be balanced against equally real costs.

Improved operational coordination across subunits

The actions of separate subunits in an organization can be coordinated using a number of different approaches (Galbraith, 1973; Malone, 1987). For example, organizational hierarchies can be designed such that the most interdependent units are closely linked (Thompson, 1967). However, in order to pass large amounts of information between subunits, a formalized, standardized language is required. Thus, some ways of coordinating subunits will require integrated data so that unambiguous messages may be sent from one subunit to another. Top management must recognize that approaches used to coordinate subunit action may be constrained by the degree to which a common language and a database of integrated, compatible data exist.

> Each of Greenfields Products' five divisions has its own sales force and distributes its own product lines in separate trucks. Top management wanted the ability to present a single face to the customer by having only one (not five) salesperson call on each customer and only one (not five) truck back up to the customer's delivery dock. They realized that without a single, consistent base of customer and order data, coordinating the actions of the five divisions to create a single customer interface would be impossible.
>
> Burton Trucking Company completely overhauled its information processing systems based on a single logical data model for the entire corporation. This allowed them to link not only across geography but also across functions. By using common, sharable data, they found their dispatch systems (the responsibility of operations) could be expanded with little effort to give them a much better shipment tracking system (the responsibility of marketing). Thus, data integration allowed them to capitalize on previously unrecognized interdependencies between dispatching and shipment tracking.

Our discussion of the impacts of data integration on the sharability of information leads to a first proposition. Note that we assume that information needs are characterized by uncertainty as opposed to equivocality, so that information systems in general are appropriate mechanisms for organizational information processing.

Proposition 1: All other things being equal, as the interdependence between subunits increases, the benefits of data integration will increase, and the amount of

data integration in rational organizations should also increase.

Factor 2: Flexibility to Meet Unique Subunit Information Requirements

King (1983) notes that organizations wrestling with issues of control over computing face "the dilemma of deciding between organizational standardization and departmental autonomy," and that local autonomy over the design of systems "makes uniform collection of data for upward reporting more difficult, whereas the imposition of organization-wide standards... diminishes departmental autonomy" (p. 329). Organization-wide data integration, as one aspect of the design of organizational computing resources, also implies some degree of "logical centralization" and some central authority with control over the logical aspects of data (Heimbigner and McLeod, 1985). Thus, it may result in a loss of local autonomy in the design and use of data (Sheth and Larson, 1990).

Data integration may involve not only a loss of local autonomy but also a loss of local effectiveness. Over time, different subunits may face different inherent task complexity and different environmental challenges. In order to effectively meet the challenges of unanticipated local events, subunits may need the flexibility to change their information systems on a local, unilateral basis (Tushman and Nadler, 1978). Gupta and Govindarajan (1986) argue that the scale advantages of resource sharing between strategic business units do not come cost-free and that, in particular, resource sharing may lead to losses of flexibility in responding to unanticipated local events for individual SBUs. For example, IS champions may see their ability to move rapidly to achieve first-mover advantages as being hampered by IS standards and infrastructure goals, such as data integration (Beath, 1991).

Thus, constraints that forced a subunit facing locally unique demands to conform to corporate-wide data integration standards might not be optimal from the total organization perspective. Choosing the appropriate level of data integration in an organization may require trading off improved organization-wide coordination against increased local flexibility and local effectiveness. The next section looks at two ways local flexibility can be reduced by data integration: compromise and bureaucratic delay.

Compromises in meeting local information needs

By forcing the whole organization to conform to a single logical data design, data integration can reduce a local subunit's knowledge about its unique local environment and diminish its sensitivity to local environmental changes (Ashby, 1959; Weick, 1976). Consider an organization with several product divisions, each with its own products. Each has its own different approach to manufacturing and marketing those products based on the different characteristics of the products and the markets for them. Top management could allow each division to design and implement its own systems, based entirely on best serving its local operational and information needs. The result would be systems that were locally optimal but not integrated across the divisions, with different definitions, identifiers, and calculations in each division.

On the other hand, top management could require the divisions to cooperate and design a single logical data design, which could be used by all of them. This "common" logical design would allow easy access to compatible data across all the divisions but might not meet the local needs of the separate divisions as effectively well as separately designed logical data designs would.

> At Burton Trucking, IS and operations had cooperated to develop a common dispatching system for its freight terminals across the country. The new system contained integrated data about all customers, equipment, and shipments. Salespeople at each local terminal argued that for the information to be valuable to them, they needed to add additional fields such as permissible delivery hours, after-hours phone numbers, and special instruction for drivers. But the salespeople couldn't agree on exactly which additional fields should be added. Terminals with close-in satellites (trucks every 2-3 hours) had very different needs from those with distant satellites (trucks every 5 hours). IS decided that trying to standardize at this level didn't make sense. They designed 10 extra fields that the local people could use as they saw fit and gave them search capabilities and screens to update and query whatever data they needed in those fields.

At Ventura Products, the Automotive Products Division undertook a major strategic planning effort to identify ways in which its information systems could assist in meeting its strategic goals. The result was a consensus within the division that it should move toward a tighter link with its customers and should design a new order-entry system as part of that effort. However, there was an historic precedent at Ventura of using a common-order entry system across all divisions. The corporate IS department fought Automotive Product's proposal on the basis that it would reduce corporate ability to present a single face to customers and would make it harder to maintain compatible data on sales across the total corporation.

The concept of requisite variety in systems theory suggests that in order to successfully respond to its environment, a system must be a good model of that environment (Ashby, 1956; Conant and Ashby, 1970). That is, if a division wishes to guide action to respond to its environment, it must have an information system (and a data model) at least as complex as the environment on all the essential dimensions that determine what the best response should be. The Burton Trucking scenario is an example of that principle: salespeople needing to respond to different types of events in different environments require different data models keyed to their unique environments.

Proposition 2a: All other things being equal, as the differentiation between subunits increases, data integration will impose more and more compromise costs on local units; therefore, the amount of data integration in rational firms should decrease.

Bureaucratic delays that reduce local flexibility

Even when subunits can agree on a single data model and data definitions without undue compromise, data integration can slow local flexibility. When changing local business conditions suggest a change to the data collected by one subunit, all other subunits would have to study the impacts of those changes on their own operations, and all would have to agree or the changes could not be made. Requests for changes in the use and semantics of data will necessarily involve bureaucratic delays as those requests are channeled through a remote authority (Leveson and Wasserman, 1982).

At Superior Manufacturing, a large global corporation, a major effort to integrate corporate data is underway. As part of this effort, procedures have been put in place specifying how changes to the corporate data model will take place. Requests first go to the data resource management office, which decides which subject areas are affected, and are then passed to the relevant subject area data stewards who review, analyze, and recommend action. This recommendation will then be reviewed by affected division data administrators and their data user groups, who will approve or modify the recommendation. Finally, the modified recommendation will be presented to the corporate data resource management policy and steering committee for review and approval. All together, five groups of players will conduct four separate reviews of the request.

At Burton Trucking, when the operations group wanted to automate freight transfer recording using bar codes, it took six months to get the request through the data administrator. According to one manager, the data administration group didn't understand what they were trying to do and "couldn't square it with their data model." After a delay of six months, the project was allowed to proceed and was quite successful.

Proposition 2b: All other things being equal, firms with increased data integration will experience greater bureaucratic delay in getting approval for changes affecting the data models used by individual subunits.

Factor 3: IS Design and Implementation Costs

Thus far two factors organizations must consider in weighing the advisability of increased data integration have been discussed: gains due to improved ability to manage interdependence, and losses in local flexibility. Both of these factors focus on the impact of data integration once it has been implemented. Even when these potential gains outweigh the losses, organizations must also consider the cost of designing and implementing data integration. If the additional cost is too large, rational firms will make do without integrated data. Thus, the impact of data integration on design and development costs is also a critical factor.

What is the impact on these costs? The typical argument in the literature is that the higher costs of more expensive initial design and implemen-

tation are compensated by lower costs for subsequent modifications of the systems, with a net increase in productivity (see, for example, Martin, 1982). This is certainly the situation in the following example.

> At Dobbs Insurance, over a period of about 10 years, the pensions business unit has developed extensive shared databases used for all pensions processes. The IS group readily acknowledges that developing integrated databases has cost more than would have a series of stand-alone systems with stand-alone databases. The biggest problem has been getting all parties to agree on design issues that have important implications to many parts of the organization. Dobbs justifies these extra costs based on additional benefits from the integrated data. For example, new systems that require no changes to the existing integrated databases can be implemented very rapidly. This includes new reports for cross-functional information requested by top management.

Though this example supports the typical argument, in general the impact of data integration on design and implementation costs depends heavily on the situation. To understand this we need to look closely at why and under what circumstances up-front costs are higher, and why and under what circumstances long-term costs are lower.

Why are up-front costs higher?

There is little argument that in practice, building integrated databases based on a wide-scope data model is more expensive in the short term than building separate stand-alone systems. A major reason for this up-front cost is the greater complexity of the design problem: arriving at a single logical design for use across multiple organizational groups can be a difficult problem (Litwin and Abdellatif, 1986). In general, the more subunits involved and the more heterogeneous their needs, the more difficult it will be to develop a single design that meets all needs.

A critical factor in the speed with which complex systems can be designed is the degree to which the overall systems are "decomposable hierarchies"—that is, where collections of components can be grouped into subsystems with rather intense interaction within any subsystem and rather light interaction between subsystems (Simon, 1981). This idea is commonly implemented in structured systems design through the concepts

of cohesion and coupling (Page-Jones, 1980). The need for a common data model moves designers away from a decomposable hierarchies approach, forcing them to deal with the full complexity of the data needs for the whole organization. Whether or not the common data model is required by the interdependency needs of the organization, large-scope data integration will result in a more complex and more expensive design and implementation process. The following example shows the difficulty of linking two different conceptualizations of the data an organization needs.

> After having identified the data requirements for the first of two critical new systems, the Support and Service Division of Ventura Products spent many person-weeks of effort coordinating its new data definitions with the several thousand data standards on the corporate data dictionary. Because of the number of corporate standards and the possibility of subtle differences in meaning, it was quite difficult to determine which, if any, of the division's desired data elements were exactly the same as those in the corporate data dictionary. When the division began to design and develop its second new system, it decided the time and costs of coordinating with the corporate standards were not justified by the benefits to them.

Though empirical research has not focused specifically on the effect of size and the number of participants in data design, there is good evidence for diseconomies of scale in the development of large systems in general. Seemingly disparate empirical findings on the economies of scale for systems development are consistent with increasing economies for small projects and decreasing economies for large projects (Banker and Kemerer, 1989). Further, the effort to build large systems may increase with the square of the size of the effort (Banker and Kaufmann, 1991). This suggests that the costs of designing and implementing data integration could increase rapidly with the size and scope of the effort.

Part of these diseconomies of scale may stem from the fact that as the size of a project increases, there is an increased difficulty in achieving conceptual integrity across different perspectives and different design requirements (Brooks, 1975). This is consistent with Daft and Lengel's (1986) assertion that when there is great differentiation between subunits, much equivo-

cality must be reduced before common understandings can be reached. If a system were to be used by only a single subunit, or if all subunits involved were identical in their needs, a common data model could be designed with input from only one subunit. This would simplify the design problem.

However, when multiple subunits with differing needs are involved, the organization may face a trade-off between imposing a compromise solution that does not meet all local needs or engaging in a more demanding design effort to truly understand and meet each subunit's unique needs. Even if that ideal creative solution can be found, it will usually require better analysts, take longer, and cost more than if fewer distinct requirements need to be met.

> At Dobbs Insurance, even though the pensions business unit has integrated its own data with great success and many benefits, it is moving only very slowly to consider integrating data with the group health unit (a similar market and similar business needs) and has no plans to integrate with the life and casualty business units (a very different market and very different business needs).

We propose that the number of subunits involved and the heterogeneity of their data needs should make a major difference in the cost of developing an agreed-upon design for integrated data; disparate information requirements would complicate the design. Where there are numerous subunits with quite different products, approaches to marketing, or competitive challenges, efforts at data integration could involve long and difficult searches for acceptable compromises.

Proposition 3a: All other things being equal, as the number and heterogeneity of subunit information needs increase, the difficulty of arriving at acceptable design compromises increases, and the cost of the resulting design will increase more than linearly. Thus, rational firms will integrate less extensively when there are many heterogeneous subunits involved.

What about long-term costs?

Once integrated data has been put in place, firms must respond to turbulence in the organization's environment, with the resulting unanticipated information requirements. Where market conditions are rapidly changing, competitors are taking non-routine actions, or new technologies are becoming available that change products or processes, the need for flexibility will be high. If the changes needed to the information systems are completely consistent with the existing data models and databases, modifications to these systems can be designed and implemented very rapidly because all the data design work will have been already completed.

Unfortunately, there are many examples where the data needs of the organization change as well as the business processes. When this happens, system designers must re-enter the difficult and time-consuming process of redesigning (renegotiating) a new data model that meets all parties' changed needs. The more often the data model needs to be redesigned, the higher the average annual costs. Thus, an important question is whether business environments really change enough to require changes to the organization's data model, and how often this might occur. Two examples demonstrate that changing business environments can indeed force changes to the data model.

> At Rolling Freight Incorporated, the business has for years been oriented around managing the equipment needed to continue smooth railroad operations. Under this view of the business, customers and shipments were a secondary focus. With hard competition from trucking and other railroads, a new perspective is forming that puts customers and shipments on the center stage. This will require a major rethinking of the way in which data is conceptualized, captured, and managed.
>
> In the pensions business unit of Dobbs Insurance, the customer that is the focus of marketing activity has changed from an organization with many employees to the individual employee who makes his or her own decisions about what companies and what products he or she wishes to invest in. This has forced a major change in the data structure needed by this business unit.

How often such changes occur might vary from business to business and from one time in history to another. Following Huber's (1984) prediction about the nature of the post-industrial environment, we might guess that such changes will take place increasingly often. In more turbulent environments, firms will be faced with frequently redesigning their data models. This will heighten the average annual cost of information system

design and implementation, *whether or not firms have extensive data integration.*

When subunits are fairly homogeneous and face the same environments and same challenges, an integrated data environment will probably be advantageous because a single redesign will address all new needs. Further, it is clear that the alternative (subunit-specific data models) would involve redundant design efforts. Thus, for organizations with homogeneous subunits, turbulence may make data integration more attractive.

On the other hand, when subunits are quite different and the environment exhibits many degrees of freedom, with different types of new challenges facing each subunit, less data integration and more local flexibility may be preferred so that local data models can be highly tuned to nuances in the local environment. If they did try to integrate, the subunits' heterogeneous needs would make each separate effort at redesigning a common data model quite difficult. In addition, even though only selected subunits would face changes in their information needs, all subunits would have to engage in the redesign effort because all would be affected by the changed design. Thus, organizations with heterogeneous subunits and turbulent environments will be frequently faced with the short-term higher costs of data integration and will seldom have the ability to enjoy its long-term benefits.

Proposition 3b: As organizations face greater instability in their environments and their information requirements, the importance of proposition 3a will increase. In turbulent environments, firms with many heterogeneous subunits will be even less likely to integrate extensively, and firms with homogeneous subunits will be more likely to integrate extensively.

Why the "Rational" Model Is Only Part of the Picture

The model of the impact of data integration shown in Figure 2 is consistent with a rational perspective on organizations. Though much understanding and guidance can come from the rational model, there are some important complications that this model does not explicitly address.

Interdependence with strategy and structure

In an organization, the technologies employed, the structure, the strategy, individual roles, and management processes are all tightly interdependent (Leavitt, 1965; Rockart and Scott Morton, 1984). No one of these can be changed without having an impact on the others. Changes in data integration (a technology) may have powerful interactions with other organizational variables. For example, data integration might be a necessary prerequisite for a shift in strategy or structure. It might create the potential for changes in information flows, thus affecting individual roles and organizational structure. Decisions on the appropriate amount of data integration are intimately tied to decisions about other key corporate variables and to corporate strategy in general.

> At the Van Buren Bank, after the IS department achieved enough data integration to provide customer and profitability information, it was not immediately clear that the account managers would take advantage of the new information. The culture of the bank, including incentive schemes and understandings about the way in which business was run, ran strongly against account managers using masses of quantitative data in their daily work. At first, only the administrative assistants were heavy users of the system. Account managers claimed the information was "not useful."

In a practical sense, no organization has the resources to pursue all attractive technological possibilities available to it. In this paper it has been asserted that interdependence is the driving force for data integration benefits, but interdependence that is not recognized by top management, or that is not integral to top management's current business thrust, will probably not have a large claim on a firm's discretionary resources. It takes more than a recognition by an IS group that parts of a firm are highly interdependent. Those data resource management efforts (and data integration efforts) that succeed are those driven by business requirements clearly understood and championed by top management (Goodhue, et al., 1988).

Power and politics

The rational model in Figure 2 assumes that everyone in an organization shares the same overall goals and can agree about how to value

various benefits. There is certainly enough precedent in research on organizations in general (e.g., Pfeffer, 1981) and on the development of information systems in particular (e.g., Markus, 1983) to suggest that such a rational assessment of costs and benefits is incomplete. Even with large net benefits to the organization as a whole, data integration may distribute those benefits and costs in an uneven way, reducing the local autonomy of some divisions, changing the level of access to critical information, or changing the power balance in some other way.

The disparity between "who pays the costs" and "who gets the benefits" can lead to less than optimum solutions for everyone involved, as Thorn and Connolly (1987) found when they looked at the motivation for subunits to contribute private data to a public "discretionary database." In addition, even the possibility of changes in the power balance can cause resistance to a data integration effort by those concerned they might lose out, regardless of arguments taking a total organizational benefit point of view. The question of who gets the benefits and who pays the costs is not an idle one.

At Spectrum Electronics, each of nine different plants had separate manufacturing systems, making it difficult to coordinate purchasing, scheduling, and spare parts inventories across plants. Over a period of more than five years, the VP of manufacturing, the VP of materials, and the director of MIS lobbied and positioned themselves to move to a single set of integrated databases and common manufacturing systems to be used by all plants. A few plant managers resisted so strongly that only their early retirement allowed plans to proceed. Once implemented, the new integrated databases and common systems were believed to be a major factor in Spectrum's successful competition with its Asian rivals.

Conclusion

Widespread data integration is an expensive proposition. While much of the literature focuses on some attractive benefits, we have tried to balance the view by looking at both positive and negative impacts. The arguments for this new perspective have been primarily theoretical. Though anecdotal evidence was presented to bolster the plausibility of the theoretical propositions, future empirical work will be needed to validate them. Figure 3 summarizes the proposed impact of data integration on an organization. This new perspective should give managers facing data integration decisions a more realistic basis for understanding its impact.

The major implication of this analysis is that in general it will not be cost-effective to integrate all of an organization's data. If this is true, organizations will need help in their efforts to "partially integrate" to achieve the most important benefits and avoid the most burdensome costs. For the MIS field to provide this help, we need to "unbundle" our concept of data integration. It is not an all-or-nothing choice, nor is it the preferred approach. As researchers and practitioners, we need to change our thinking so we can provide guidance to organizations on implementing "partial integration."

Evidence from our case studies gives some initial suggestions on several ways in which partial integration can be implemented. One obvious approach is to limit the scope to only certain subunits, as was done at Dobbs Insurance. There the pensions group implemented data integration only for its own business unit. Even across a wider scope, there are at least three different partial integration approaches evident in the case studies. One is to require all subunits to use a selection of "global" or organization-wide application systems (such as payroll, order entry, or purchasing), allowing them discretion on other less critical application systems (Van Rensselaer, 1985). These organization-wide application systems include a common data model and, by implication, commonly defined data.

Another approach is to develop selected organization-wide databases (such as those related to customers or products) and require all application systems that use these entities to access and update the common databases. Any non-common databases could be designed with local discretion. A third partial integration approach is to identify a selection of critical data elements and hammer out agreed-upon definitions for those across the entire organization. These standard definitions can then be enforced in all systems development. Firms have experimented with anywhere from 20 to a thousand such standard data elements.

Additional conceptual work and empirical research is needed to suggest which of these or other partial integration approaches will be most

Greater Data Integration	
In the Presence of:	**Will Tend to Lead to:**
Interdependence between subunits	Better communications Better coordination
Differentiation between subunits	More compromise or more design costs More bureaucratic delay
Differentiation and environmental turbulence	Even more compromise, or Even more design costs

Figure 3. The Impact of Data Integration in Uncertain (as Opposed to Equivocal) Environments

effective, and why. A second and equally pressing need is for methods that can help an organization determine which data should be integrated and which should not. Existing information engineering methodologies have overlooked this question because they assumed that all data should be integrated, though some other IS planning approaches, such as critical success factors (Rockart, 1979), may be applicable for the partial integration problem. The choice presumably depends heavily on the strategic direction of the organization. Firms that attempt to integrate an inappropriate subset of the data, or on a wider scope and in more detail than is appropriate given their organizational situation, will probably face stiff resistance to either the large cost involved or the loss of local flexibility of heterogeneous subunits.

The emphasis in this paper has been on centrally controlled data integration for large internal information systems. However, the ideas presented here can be applied in either more macro or more micro settings. For example, partners in electronic data interchange clearly have interdependence interests but must also be cognizant of the needs for local flexibility, especially in the face of industry turbulence. End-user computing at the work group level is a realm where local flexibility is of paramount importance, but there are also reasons to impose some level of common language so that data can be shared between interdependent work groups. Thus, the same issues of interdependence, need for local flexibility, and design costs clearly apply in these additional realms.

Though this paper has pointed out some negative aspects to data integration, many organizations today are striving for a more global consciousness of their businesses and a more global response to customers and markets. Huber (1984) suggests that managers in post-industrial organizations will need to access more information about more aspects of the business, from even more parts of the organization, and in shorter time spans. This implies both greater interdependence and greater reliance on computer-based information. These types of changes in the organizational climate will probably shift the balance toward the need for greater (but not total) data integration in many firms, heightening the practical and academic importance of this area of research.

Acknowledgements

References

Ashby, W.R. An Introduction to Cybernetics, Wiley and Sons, New York, NY, 1956.
Banker, R. and Kaufmann, R.J. "Reuse and Pro-

ductivity in Integrated Computer Aided Software Engineering: An Empirical Study," *MIS Quarterly* (15:3), September 1991, pp. 375-401.

Banker, R. and Kemerer, C.F. "Scale Economies in New Software Development," *IEEE Transactions on Software Engineering* (15:10), October 1989, pp. 1199-1205.

Batini, C., Lenzerini, M., and Navathe, S.B. "A Comparative Analysis of Methodologies for Database Schema Integration," *ACM Computing Surveys* (18:4), December 1986, pp. 323-364.

Beath, C.M. "Supporting the Information Technology Champion," *MIS Quarterly* (15:3), September 1991, pp. 355-372.

Bonczek, R.H., Holsapple, C.W., and Whinston, A.B. "Aiding Decision Makers With a Generalized Data Base Management System: An Application to Inventory Management," *Decision Sciences* (9:2), 1978, pp. 228-245.

Brooks, F.P. *The Mythical Man-Month: Essays on Software Productivity,* Addison-Wesley, Reading, MA, 1975.

Chen, P.P. "The Entity-Relationship Model—Toward a Unified View of Data," *ACM Transactions on Database Systems* (1:1), March 1976, pp. 9-36.

Conant, R.C. and Ashby, W.R. "Every Good Regulator of a System Must be a Model of That System," *International Journal of Systems Sciences* (1:2), 1970, pp. 89-97.

Daft, R.L. and Lengel, R.H. "Organizational Information Requirements, Media Richness, and Structural Design," *Management Science* (32:5), May 1986, pp. 554-571.

Date, C.J. *An Introduction to Database Systems,* Addison-Wesley, Reading, MA, 1981.

Emery, J.C. "Cost/Benefit Analysis of Information Systems," in *Advanced System Development/Feasibility Techniques,* J.D. Couger, M.A. Colter, and R.W. Knapp (eds.), Wiley, New York, NY, 1982.

Finkelstein, C. *An Introduction to Information Engineering,* Addison-Wesley, Sydney, Australia, 1990.

Galbraith, J. *Designing Complex Organizations,* Addison-Wesley, Reading, MA, 1973.

Goodhue, D.L., Quillard, J.A., and Rockart, J.F. "Managing the Data Resource: A Contingency Approach," *MIS Quarterly* (12:3), September 1988, pp. 373-392.

Goodhue, D.L., Kirsch, L.J., Quilllard, J.A., and

Wybo, M.D. "Strategic Data Planning: Lessons from the Field," *MIS Quarterly* (16:1), March 1992, pp. 11-34.

Gupta, A.K. and Govindarajan, V. "Resource Sharing Among SBU's: Strategic Antecedents and Administrative Implications," *Academy of Management Journal* (29:4), December 1986, pp. 695-714.

Heimbigner, D. and McLeod, D. "A Federated Architecture for Information Management," *ACM Transactions on Office Information Systems* (3:3), July 1985, pp. 253-278.

Hoffer, J.A., Michaele, S.J., and Carroll, J.J. "The Pitfalls of Strategic Data and Systems Planning," *Proceedings of the Twenty-Second Annual Hawaii International Conference on Systems Sciences,* 1989, pp. 348-356.

Huber, G. "Organizational Information Systems: Determinants of Their Performance and Behavior," *Management Science* (28:2), February 1982, pp. 138-155.

Huber, G. "The Nature and Design of Post-Industrial Organizations," *Management Science* (30:8), August 1984, pp. 928-951.

Huber, G. "A Theory of the Effects of Advanced Information Technologies on Organizational Design, Intelligence, and Decision Making," *Academy of Management Review* (15:1), 1990, pp. 47-71.

IBM. *Business Systems Planning,* IBM Manual #GE20-0527-3, White Plains, NY, July 1981.

King, J.L. "Centralized versus Decentralized Computing: Organizational Considerations and Management Options," *ACM Computing Surveys* (15:4), December 1983, pp. 319-349.

Leavitt, H.J. "Applied Organizational Change in Industry," in *Handbook of Organizations,* J. March (ed.), Rand McNally, Chicago, IL, 1965, pp. 1144-1170.

Lederer, A.L. and Sethi, V. "The Implementation of Strategic Information Systems Planning Methodologies," *MIS Quarterly* (12:3), September 1988, pp. 445-461.

Lederer, A.L. and Sethi, V. "Critical Dimensions of Strategic Information Systems Planning," *Decision Sciences* (22:1), Winter 1991, pp. 104-119.

Leveson, N.G. and Wasserman, A.I. "Logical Decentralization and Semantic Integrity in a Distributed Information System," in *Distributed Data Sharing Systems,* R.P. van de Riet and W. Litwin (eds.), North Holland, Amsterdam, 1982, pp. 243-253.

Litwin, W. and Abdellatif, A. "Multidatabase Interoperability," *IEEE Computer,* December 1986, pp. 10-18.

Litwin, W., Mark, L., and Roussopoulos, N. "Interoperability of Multiple Autonomous Databases," *ACM Computing Surveys* (22:3), September 1990, pp. 267-293.

Malone, T.W. "Modeling Coordination in Organizations and Markets," *Management Science* (33:10), October 1987, pp. 1317-1332.

Malone, T.W., Yates, J., and Benjamin, R.I. "Electronic Markets and Electronic Hierarchies," *Communications of the ACM* (30:6), June 1987, pp. 484-497.

Markus, M.L. "Power, Politics and MIS Implementation," *Communications of the ACM* (26:6), June 1983, pp. 430-444.

Marschak, J. "Remarks on the Economics of Information," in *Contributions to Scientific Research in Management,* University of California, Los Angeles, CA, 1959.

Marschak, J. "Economics in Inquiring, Communicating, Deciding," *American Economic Review* (58:2), May 1968, pp. 1-18.

Marschak, J. "Economics of Information Systems," *Journal of the American Statistical Association* (66), March 1971, pp. 192-219.

Martin, J. *Strategic Data-Planning Methodologies,* Prentice Hall, Englewood Cliffs, NJ, 1982.

Martin, J. *Information Engineering,* Savant Research Studies, Carnforth, Lancashire, England, 1986.

McCann, J. and Ferry, D. "An Approach for Assessing and Managing Inter-Unit Interdependence," *Academy of Management Review* (4:1), January 1979, pp. 113-119.

McCarthy, W.E. "The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment," *The Accounting Review* (57:3), July 1982, pp. 554-578.

Mendelson, H. and Saharia, A.S. "Incomplete Information Costs and Database Design," *ACM Transactions on Database Systems* (11:2), June 1986, pp. 159-185.

Page-Jones, M. *The Practical Guide to Structured Systems Design,* Yourdon Press, New York, NY, 1980.

Pfeffer, J. *Power in Organizations,* Ballinger Publishing Company, Cambridge, MA, 1981.

Rockart, J.F. "Chief Executives Define Their Own Data Needs," *Harvard Business Review* (57:2), March-April 1979, pp. 81-93.

Rockart, J.F. and Scott Morton, M.S. "Implications of Changes in Information Technology for Corporate Strategy," *Interfaces* (14:1), January-February 1984, pp. 84-95.

Sheth, A.P. and Larson, J.A. "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computing Surveys* (22:3), September 1990, pp. 184-236.

Simon, H.A. "Applying Information Technology To Organization Design," *Public Administration Review* (33:3), May/June 1973, pp. 268-278.

Simon, H.A. *The Sciences of the Artificial,* MIT Press, Cambridge, MA, 1981.

Thompson, J.D. *Organizations in Action,* McGraw-Hill, New York, NY, 1967.

Thorn, B.K. and Connolly, T. "Discretionary Data Bases: A Theory and Some Experimental Findings," *Communication Research* (14:5), October 1987, pp. 512-528.

Tushman, M. and Nadler, D. "Information Processing as an Integrating Concept in Organizational Design," *Academy of Management Review* (3:3), July 1978, pp. 613-624.

Van Rennselaer, C. "Global, Shared, Local," *Datamation,* March 15, 1985, pp. 105-114.

Weick, K.E. "Educational Organizations as Loosely Coupled Systems," *Administrative Science Quarterly* (21:1), March 1976, pp. 1-19.

About the Authors

Dale L. Goodhue is assistant professor of MIS at the University of Minnesota's Carlson School of Management. He has worked as a business analyst for American Management Systems, and received a B.S. from Brown, an M.B.A. from Carnegie Mellon, and a Ph.D. from M.I.T. He is continuing to study data management by conducting case studies on planning methods, data architectures, and organizational costs and benefits of data integration. Other research interests include measuring the impact of system characteristics, policies, and design methods on user evaluations of IS.

Michael D. Wybo is assistant professor of MIS on the Faculty of Management at McGill University. He received his Ph.D. from the University

of Minnesota and has consulted for a number of years on systems development projects in West Africa. His research interests include the impacts of alternative data management strategies on firm performance and the role of information systems in the design of organizations.

Laurie J. Kirsch is a doctoral candidate in MIS at the University of Minnesota. Her research interests include management of systems development projects and data management. She has 10 years of experience in the information systems field.