

**Integration Technology: The Reinvention of the Linkage
Between Information Systems and Computer Science**

February 1993

WP # 3542-93
CISL WP# 92-09

Stuart E. Madnick

Stuart Madnick E53-321

**Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 01239**

To appear in a forthcoming issue of *Decision Support Systems*

Integration Technology: The Reinvention of the Linkage Between Information Systems and Computer Science

Stuart E. Madnick

John Norris Maguire Professor of Information Technology and
Leaders for Manufacturing Professor of Management Science
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
(617) 253-6671
smadnick@mit.edu

Abstract: The Computer Science (CS) and Information Systems (IS) fields are facing considerable challenges. At the same time, the Information Technology & Systems (ITS) community, which lies at the intersection of CS and IS, is uniquely positioned to address and solve important problems that are outside of the traditional domains of either CS or IS. An important development is the increasing criticality of Integration Technologies (IT)¹ and the need for Integration Specialists (IS)¹. Examples are given of applications-driven integration technology-intensive research needed by business to support globalization, increased productivity, and rapid adaptation.

Inventing a New Future

Computer Science (CS) has served as an important reference discipline for the emergence and growth of the field of Information Systems (IS). We all can take delight in the latest technical advances: the newest high-speed chip, the increasingly exotic multi-media workstations, the incredibly sophisticated software, etc. Although this relationship will continue to be important, it is undergoing important changes that must be recognized. The CS and IS fields have for years been major forces for reexamination and change in many fields and industries, such as financial services, manufacturing, transportation, etc. These "paradigm shifts" are almost always difficult and often resisted but do usually lead to new and more productive forms. It is now time to reexamine the CS/IS relationship and its transformation.

The Need for a Reexamination

Although I will be presenting my own personal views on these issues, the need for a reexamination has attracted the attention of others in recent years. One of the most recent and most comprehensive studies was reported in the book, *Computing the Future: A Broader Agenda for*

¹ These are not the normal uses of the terms IT and IS.

Computer Science and Engineering by the National Research Council [HL92]² which starts its Executive Summary as follows:

As an academic discipline, computer science and engineering (CS&E) has been remarkably successful in its first decades of existence. But both the intellectual focus of academic CS&E and the environment in which the field is embedded are today in the midst of significant change. ...

Further into the Executive Summary, the committee outlines its recommendations and priorities including:

Given the many intellectual opportunities available at the intersection of CS&E and other problem domains and a solid and vigorous core effort in CS&E, the committee believes that academic CS&E is well positioned to broaden its self-concept. Such broadening will also result in new insights with wide applicability, thereby enriching the core. Furthermore, given the pressing economic and social needs of the nation and the changing environment for industry and academia, the committee believes that academic CS&E *must* broaden its self-concept or risk becoming increasingly irrelevant to computing practice.

We have already acknowledged Computer Science as a reference discipline for Information Systems. Paradoxically, Information Systems, and its coupling to business needs and uses, is becoming an increasingly important "reference discipline" for Computer Science. This notion also has been hinted at by others. For example, the Association of Computing Machinery report on "The Scope and Directions of Computer Science: Computing, Applications, and Computational Science" [A91] advised:

A close interaction between computer researchers and others is essential so that the questions under investigation remain connected to real concerns. Otherwise computing research can drift into irrelevance and cease to earn public support. For this reason it is in the best interests of the computing profession for computer researchers to engage with applications.

Speaking even more recently and urgently, Professor Peter Denning, a former President of ACM, starts his article "Educating A New Engineer" [D92] with the observation that:

University education is experiencing an enormous breakdown. An increasing number of students, employers, faculty, business executives, management specialists, public officials, and taxpayers have declared their dissatisfaction with the education and research that is available in most of our universities.

He proceeds to note points such as: "Finding new answers to these questions is a task of special urgency" and "The connection between much of the research in universities and the concerns of

² This paper makes use of some material from that study and the author's participation in related panel discussions.

people in business firms or general public concerns is not apparant either to the outside observer or to the researcher."

These observations are not limited to the purified atmosphere of academic journals and studies, a recent issue of *USA Today* (December 15, 1992) included quotes from individuals such as Mr. David Glass, President of Wal-Mart who noted, "I don't think that most Americans fully realize how important it will be to compete effectively in a global environment going forward or what it actually means or how quickly that will be upon us" and Nobel Prize-winner Professor Robert Solow who concluded, "The only solution to the productivity problem is investment in modern industrial equipment, in research and development, in transportation and communication infrastructure, and in educating and training our people."

As one colleague noted, developing CS/IS-based information-intensive solutions to complex global transportation management problems, such as intransit visibility, can be just as challenging research as developing new query optimization algorithms – but can have impacts that can be measured in billions of dollars of savings. Computer Science and Information Systems are becoming increasingly and inexorably linked. They can either work together proactively to define the future or they will be inevitably dragged into it.

The Information Systems & Technology Core

There is ambiguity and disagreements regarding the precise definitions and boundaries of fields such as Computer Science and Information Systems. Furthermore, given the rapid changes, the definitions are likely to continue to evolve over time. The CS/IS research agendas overlap with each other as well as with other fields in science and engineering as well as with various disciplines of management, as depicted in Figure 1. We will refer to the intersection region between Computer Science and Information Systems as the Information Systems & Technology (ITS) core (following the convention established by the Workshop on Information Systems & Technology – WITS).

Although there are many benefits to the ITS overlap, it also can create problems because some research issues have been largely neglected because IS researchers felt that they were primarily technical CS problems and thereby avoided them; at the same time, the CS researchers felt that they were primarily management IS problems and also avoided them; thus, nobody focused on them! In many ways, this center segment of Figure 1 represents the unique opportunities of the ITS research agenda. Research in this area requires a deep and enlightened understanding of both the management problems and the technologies.

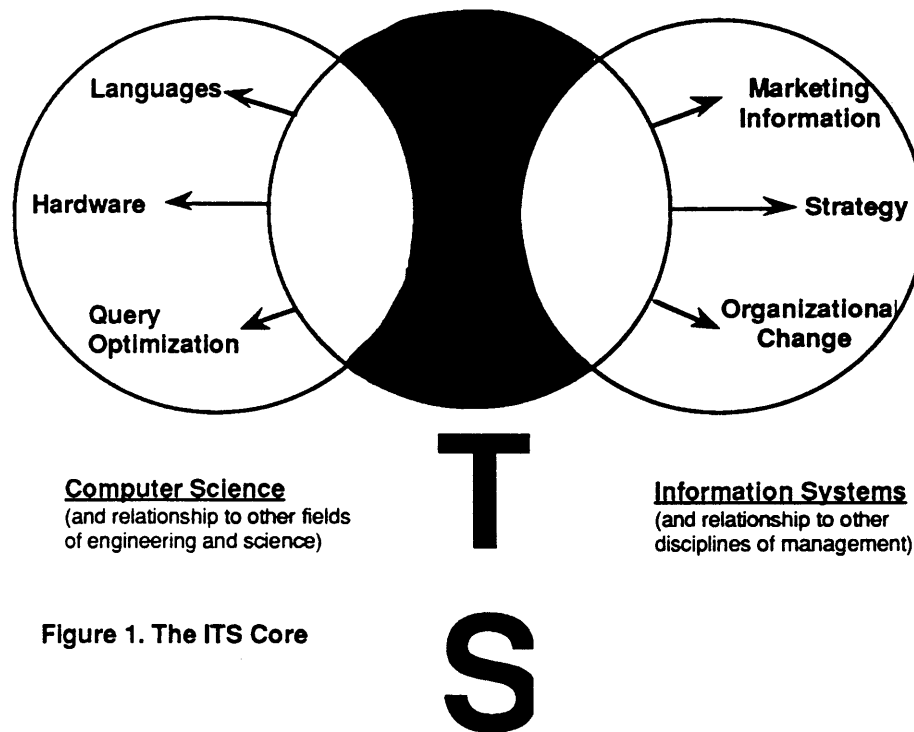


Figure 1. The ITS Core

Some researchers in the field of Information Systems, such as Prof. Warren McFarlan of Harvard, have argued that although IS-related research may continue to expand, it will (or has) moved into the research agendas of the related functional areas, such as strategy (IT strategy), organizational studies (IT-based organizational change), and marketing (creation of enormous data bases), leaving the IS field empty. Without debating the position that some IS research will diffuse into other existing management disciplines, the topics within the ITS core are unlikely to be "sized" by any other other management discipline since they require significant technical research expertise.

In the remainder of this paper, we will identify some of the increasingly important business needs that are driving ITS core research.

The Challenge: Addressing Key Business Problems

Much of the content of this section emerged from this author's participation in the "Management in the 1990s" research project at MIT. This research project culminated with the publication of the book, *The Corporation of the 1990s: Information Technology and Organizational Transformation* (Edited by Michael Scott Morton) [M91, S91].

Business today is characterized by dramatically increased geographic scope due to globalization and world-wide competition, severe time and productivity pressures, and a rapidly changing environment that often requires the restructuring of a company's operations in a very short time [M92, M93]. For example, top management may need to restructure the organization from a company in which each site performs all of the functions relevant to the business, requiring only summary information from other offices, to a company in which each office is responsible for only specific functions, requiring much more information exchange from other offices. Movements toward consolidation (e.g., through bank mergers) are occurring at the same time as movements towards decentralization (e.g., the disintegration of IBM into multiple "Baby Blues").

Often the movement toward consolidation of previously independent companies or divisions into a single unit is motivated by the goal of gaining economy of scale. Thus, a rapid integration of separately maintained databases, processes, and networks would be necessary. Automated data conversion, interoperable network protocols, and transportable software systems are some of the major technological features necessary in such an environment.

Over and over the issue of time comes up. Popular phrases include "just-in-time", "continuous flow of information", "time-based functionality", and "time-to-market". Businesses are trying to compress the time from product concept to start of production, the time from product order to product ship, and the time to respond to a competitor's action. As an extreme case, many innovations in the financial services industry have a life span, from product concept to first deployment to final abandonment, of less than a month. In the business community, computer systems are often viewed as a major obstacle to time compression rather than a facilitator (one executive has a sign in his office that reads: "Once upon a time I thought that computers were the solution, now I realize that they are my major problem.")

One example that illustrates these needs and problems involves several ships full of goods being sent from the USA to help people in a foreign country suffering from a political disruption. While the ships were underway, the political situation improved in that country but a natural disaster struck another nearby country. The question posed was: Could one or more of the ships be diverted to help deal with the new disaster? This involved knowing what goods were supplied (they had come from multiple sources, each with their own computer systems), in which containers had those goods been stored, which containers were on which ships, and where was each of the ships currently located. Even though all of the necessary information existed in some computer system, there were so many disparate systems involved that a manual inspection of the contents of

all the containers on each ship was ordered. Thus, a delay of several days occurred before the appropriate ships could be redeployed.

Although the above example may seem to be an interesting but isolated case, it represents much more the norm than the exception in dealing with complex global transportation situations. For example, during the Gulf War, of the 40,000 containers of material shipped to the Gulf, about 28,000 had to be manually unloaded and inspected in order to determine their contents. In general, the physical movement of material was faster than the movement of the supporting information. Transportation and logistic systems, in general, represent major challenges to the effective utilization of information technology [Q91].

The Characteristics of Integration Technology

Some of the key characteristics of the emerging field of *Integration Technology (IT)*³ are summarized in Table 1 and explained in the following paragraphs. These characteristics also highlight the skills that are needed for the emerging *Integration Specialists (IS)*³.

- | |
|--|
| <ul style="list-style-type: none">• Technology-enabled• Continuous rapid change and innovation• Applies in all functional areas and industries• Requires cross-functional perspective• Infrastructure-focused• Force for enterprise-wide transformation |
|--|

Table 1. Characteristics of Integration Technology

Technology-enabled. Although there are many ways to support organizational integration, we are focusing on the new opportunities being provided through technology – especially through more intelligent and easy-to-use shared databases and communication networks. Thus, integration specialists must be well-versed in these technologies.

Continuous rapid change and innovation of information-based technologies. The automobile and the railroad also could be viewed as integration technologies but their current use is fairly well understood and has been absorbed into normal operating procedures. The information-based technologies that we are considering, such as portable computers, high-performance world-wide

³ Note that this represents a change from the normal abbreviations where *IT* meant *Information Technology* and *IS* meant *Information Systems*.

communication networks, and intelligent database systems, are undergoing continuous rapid change and innovation. Thus, we do not fully understand how to best use these technologies as they currently exist (because they are so new) nor do we know that such uses will remain best (because of the changes yet to occur). Integration specialists must be able to function effectively in and adapt to such a volatile environment.

Applies in all functional areas and industries. Integration technologies are needed in essentially all functional areas and industries, ranging from transportation (e.g., intransit visibility) to marketing (cross-marketing) and from financial services (e.g., global risk management) to manufacturing (e.g., concurrent engineering), etc. Thus, it makes sense to view it as a separate research area whose output and practitioners are utilized in many fields of endeavor.

Requires cross-functional perspective. Not only is integration technology needed in every function, its most critical value arises from its use across functions – integrating the various activities of the organization. Thus, integration specialists must have an enterprise-level view of their organizations.

Infrastructure-focused. In spite of all the rapid changes, both in technology and applications, there is a need for a base capability, an infrastructure, on which the integration technologies can be leveraged. This may include corporate communication networks, data standards, and software packages. Integration specialists must help to define, maintain, and evolve these infrastructures. This role is becoming increasingly important as many of the roles of "traditional information systems personnel", such as software development, are disappearing either through outsourcing, purchased software, or user-developed software – which may actually increase the difficulty of integration without careful consideration of the supporting infrastructure.

Enterprise-wide transformation. Although much of the focus in this section has been on the technologies, these technologies will only be effective if they are appropriate to the organization and the organization is appropriate to the technologies. Integration specialists must help the organization to define its goals and necessary changes that must occur to attain these goals. Often considerable enterprise-wide changes, even transformations, are needed.

Although much of what has been noted in this section is already occurring, with examples that may go back decades, there is an increasing intensity and urgency to these issues. It is time to acknowledge that the era of Integration Technology is upon us. There must be an explicit recognition

of and training for Integration Specialists. Furthermore, critical research issues must be addressed and resolved in order for these integration technologies to be fully effective.

Examples of Key Integration Technology Research Issues

Effectively integrating information from multiple sources both within organizations and between organizations represent both an important solution to many critical business needs [M91] and a key challenge for integration technology research [MSW90, MW91, SM89a]. An organization can be simultaneously "data rich" and "information poor" if they do not know how to identify, categorize, summarize, and organize the data. Although there are many important integration technology research needs in this area, three particular challenges will be highlighted as examples: data semantics acquisition, data quality, and data semantics evolution.

Data semantics acquisition. As business operations become increasingly dispersed geographically and functionally, differences in work processes at each site performed by people trained for each site will become more critical, leading to data incompatibilities and inconsistencies when these differing sites must interact. For example, a certain insurance company has 17 different definitions of the term "net written premium" which is their primary measure of sales. Which definition is used depends on which office and functional group within the company is using the term and for what purposes the definition is being used. A useful integration of this company's databases would need to reconcile conflicting definitions of terms when necessary. Before these differences could be reconciled, we would need to be able to represent the semantics of the data as used in each environment, this is sometimes called the *context* of the data [SM91b]. Further research on using metadata to represent *context definitions* would provide the basis for capturing and disseminating knowledge about data meanings and facilitate the data reconciliation and integration process [SM89b, SM91a].

Data quality. Organizations have become very concerned about quality in areas ranging from manufacturing quality to software program quality. Data quality, in comparison, has received relatively little attention. Issues relating to data quality are becoming increasingly important as information is moved through the organization. To a large extent, data quality considerations in the past were handled through personal familiarity; the user knew the characteristics of the data used in his or her organization and informally took this into account when using the data. This approach is not feasible as increasing numbers of information sources are used, many not well known to the user. They are increasingly exposed to data with various levels of quality for which they do not have first-hand familiarity. Furthermore, many currently automated processes for converting, merging, and manipulating the data renders inaccessible information about the original data that might

have conveyed information about its quality. For example, the source of a given piece of information is often a key element in judgements about its credibility and quality [WM90]. There are many additional data quality attributes that may be important, such as accuracy, completeness, timeliness, and stability. Defining and measuring the important data quality attributes, which we refer to as *context characteristics*, and properly maintaining this quality-related information as data moves through and between systems represents a significant research challenge. With this quality information, decision makers would be better able to make more effective use of the data.

Evolving semantics. It must be realized that autonomous databases are independently evolving in semantics as well as in content (i.e., values). For example, consider the situation of stock exchanges around the world. Not only are the stock prices changing continuously, but the definition of the stock price also can change. At some time in the future, the Paris stock exchange will probably change from being measured in French francs to ECUs (European Currency Units). The normal "ticker tape" data feeds do not explicitly report the currency, it is implicit in the context of the source. More subtle examples include changes from reporting "latest nominal price" to "latest closing price" or from a percentage based pricing to actual prices, as currently happening at the Madrid stock exchange. Furthermore, in a historical database of stock prices, it must be recognized that the meanings had changed over time especially when doing a longitudinal analysis.

What is needed is not only a way to capture the meaning of each of these sources but also a way to represent the desired (or assumed) meaning of the receiver, which may be a human, an application, or another database. Then it would be possible to develop a capability, which we refer to as a *context mediator*, to formally and automatically compare the semantics of the source and the receiver to determine if they are compatible, partially compatible, convertible, or incomparable. Research on the source/receiver model [SM91a] represents a direction towards solving the more general problem, which we call *context interchange* [SM91b].

Conclusions

It has been observed that the evolution of society goes through periodic discontinuities. At these transition points, new paradigms and fields of endeavor emerge. The ITS community is uniquely positioned to be a leader in helping to make this happen by effectively combining skills from Computer Science and Information Systems to the development of integration technologies needed to address these critical problem areas. But, before we can help transform society, we have to be prepared to transform ourselves.

Acknowledgements

Work reported herein has been supported, in part, by the National Science Foundation, MIT's International Financial Services Research Center (IFSRC), MIT's Leaders For Manufacturing (LFM) Program, MIT's PROductivity From Information Technology (PROFIT) Program, and MIT's Total Data Quality Management (TDQM) Program.

REFERENCES

- A91 Association for Computing Machinery, "The Scope and Directions of Computer Science: Computing, Applications, and Computational Science", *Communications of the ACM*, Volume 34, October 1991, p. 131.
- HL92 Hartmanis, J. and Lin, H. (Editors), *Computing the Future: A Broader Agenda for Computer Science and Engineering*, National Academy Press, Washington, D.C., 1992.
- D92 Denning, Peter, "Educating A New Engineer", *Communications of the ACM*, Vol. 35, No. 12, December 1992, pp. 83-97.
- M91 Madnick, S., "Chapter 2: The Information Technology Platform", in *The Corporation of the 1990s: Information Technology and Organizational Transformation*, M. S. Scott-Morton (Editor), Oxford University Press, 1991.
- M92 Madnick, S., "The Challenge: To Be Part of the Solution Instead of Being the Problem", *Proceedings of the Workshop on Information Technology and Systems (WITS'92)*, December 1992, Dallas, Texas.
- M93 Madnick, S., "Chapter 16: Putting IT All Together Before it Falls Apart", in *Information Technology in Action: Trends and Perspectives*, Richard Y. Wang (Editor), Prentice-Hall, 1993.
- MSW90 Madnick, S., Siegel, M. and Wang, Y.R., "The Composite Information Systems Laboratory (CISL) Project at MIT", *IEEE Data Engineering*, June 1990, pp. 10-15.
- MW91 Madnick, S. and Wang, R., "Logical Connectivity: Applications, Requirements, and an Architecture", *Proceedings of the 1991 Hawaii International Conference on Systems Sciences*.
- Q91 Quiddington, P., "Cruising Along the Information Highway", *MIT Management Magazine*, Fall 1991.
- S91 Scott-Morton, M. S., (Editor), *The Corporation of the 1990s: Information Technology and Organizational Transformation*, Oxford University Press, 1991.
- SM89a Siegel, M., Madnick, S. et al, "CISL: Composing Answers from Disparate Information Systems", *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases*, December 1989, Evanston, IL.
- SM89b Siegel, M. and Madnick, S., "Schema Integration Using Metadata", *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases*, December 1989, Evanston, IL.
- SM91a Siegel, M. and Madnick, S., "A Metadata Approach to Resolving Semantic Conflicts", *Proceedings of the VLDB Conference (Barcelona, Spain)*, September 1991.
- SM91b Siegel, M. and Madnick, S., "Context Interchange: Sharing the Meaning of Data", *SIGMOD Record*, December 1991, pp. 77-79.
- WM90 Wang, R. and Madnick, S., "A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective", *Proceedings of the VLDB Conference (Brisbane, Australia)*, August 1990, pp. 519-538.