Ownership Principles for Distributed Database Design

December 1992 WP #3518-93
CISL WP# 92-07

Marshall Van Alstyne
Erik Brynjolfsson
Stuart E. Madnick

Sloan School of Management, MIT

* see page bottom for complete address

Marshall Van Alstyne E53-308
Erik Brynjolfsson E53-313
Stuart Madnick E53-321

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

Ownership Principles for Distributed Database Design:

Marshall Van Alstyne
Erik Brynjolfsson
Stuart Madnick

MIT Sloan School
Cambridge, MA
marshall@athena.mit.edu

Abstract: This research addresses the problem of assigning control over multiple database systems. Previous research has formalized the importance of setting standards, achieving user transparency, and reducing data inconsistencies in support of the argument for centralized control. In practice, however, many centralization efforts have failed. Such failures typically stem from a lack of departmental incentives and the need for local autonomy. Motivational factors, on the other hand, have typically eluded formal characterization. Using an "incomplete contracts" approach from economics, it is possible to model the costs and benefits of decentralization, including critical intangible factors. This paper presents normative principles of database decentralization; it derives formulas that give the principles a theoretical underpinning; and it illustrates the application of each principle in actual practice.

Motivation

Organizations as diverse as airlines, banks, defense contractors, and universities face a common problem in the ownership of their large scale information systems. For systems involving personnel as varied as end-users, data entry staff, and database administrators the question remains: who should control the final system? Are users better served by a governance structure where one department owns and oversees a monolithic system or by a decentralized control structure where departments coordinate activities? From a technical perspective, there is a considerable body of literature on the communications aspects of connecting distributed systems (Notkin 1987, Shuey 1986, Ceri 1984) while researchers have devoted considerable effort to improving data integration (Chen 1976, Date 1981, Martin 1982). We recognize the difficulty of these technical problems, however, even assuming that they are tractable, there remains the issue of organizational control. This research abstracts from communications costs and other technical constraints of database integration and focuses on the as yet unaddressed organizational and incentive issues.

Arguments about the benefits of centralization or decentralization of information systems have been debated almost since the advent of computers. Recent advances in technology, such as low-cost workstations and high-performance networks, and the movement towards client-server computing have further intensified the issues.

The complexity of these issues is increased by the fact that there are multiple dimensions to the notions of centralization and decentralization. Three major dimensions -- equipment, development, and control -- are depicted in Figure 1. This paper will focus on the control dimension. In principle, each of these dimensions can be independently centralized or decentralized as explained below. In Figure 1, the intercept point represents maximal centralization, whereas moving out along an axis represents increased decentralization.

Equipment: All computing and data storage equipment can be at a central location (e.g., a corporate data center) with world-wide access provided via remote terminals, such as an automatic teller machine (ATM) network. Alternatively, the computing and data storage equipment can be decentralized. For example, a brokerage firm may provide a powerful workstation to each trader all over the world – but each workstation could have exactly the same software that was planned, designed, and programmed by a central group.

Development: Development (i.e., the programming of the system) may be performed by a central group or by each of the individual groups whether the equipment is centralized or decentralized.

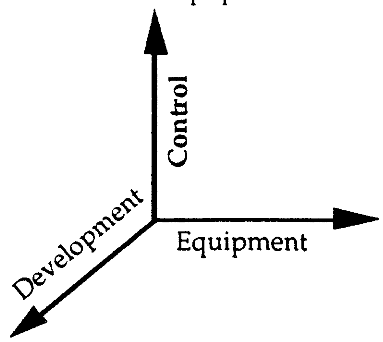

Figure 1 -- We focus on the control aspects of system decentralization.

Control: This is the key dimension of interest in this paper. The planning and design of the databases and application programs may be done by a central group that "owns" the system whether the equipment is at a central site or distributed throughout the organization, such as the high-powered trader workstations example above. Furthermore, the development could also be done by this central group or may be done by the individual groups implementing the specifications set by the central group. Regardless of the location of equipment or development, we consider control to be "centralized" if the central group retains the right to make any decision not specifically delegated to others. Following Grossman & Hart (1986), we call this the "residual right of control" and associate it with ownership of the system.

Alternatively, control of the planning and design may be decentralized to the individual groups. In such a decentralized control environment, the equipment could also be decentralized or the groups could all make use of a shared central corporate data center, with each group designing its own databases and applications. Likewise, each group could do its own development or could contract to use the services of a central development staff working under its specifications. In this case, the local groups would have the residual rights of control, and therefore effective "ownership" of the respective systems.

Among information systems, the frequency of decentralized ownership clashes with a common presumption that centralizing control "under one roof" is the preferred design choice. Fidelity Investments reduced its costs and improved its marketing by consolidating customer account data from multiple product lines into a single database. In contrast, a Vice President at Johnson and Johnson waited three weeks for a list of his top 100 customers world-wide due to complications in linking data from distributed systems. Such problems of "dis-integrated" systems have led senior staff to inquire "Why not create one big database?" It is often assumed that distributed systems represent failures in design or, more generously, that designers "did the best they could based on the technical limitations of their day." Two notorious technical limitations are the expensive and speed bound hardware necessary for central processing and the need to keep local data copies due to limited communications bandwidth. We believe that these constraints are at best only part of the story, and present a framework demonstrating that decentralization can be optimal even in their absence.

Our model is consistent with recent empirical studies which show that a substantial fraction of centralized data planning and consolidation efforts end in failure (Goodhue, Kirsch, Quillard, & Wybo 1992). The difficulties are less technical than managerial. Centralization efforts must fit the goals of the multiple departments which they affect. Managers are reluctant, for example, to assign their best technical people to other departments' projects despite it being in the interests of the company as a whole.

Ownership of a system critically affects the motivation of the people necessary for its success, particularly for databases where the value often lies in intangibles. It is these incentive issues which we attempt to model.

There are three basic advantages to the approach taken here. First, the research adapts an economic theory of the incentives created by asset ownership to the specific application of designing composite database systems. Second, the model formalizes non-technical costs and benefits of database decentralization. It captures several intuitive concepts -- for example standardization and intangible effort costs -- that affect system design and that have generally eluded precise specification. Third, normative principles for designing databases emerge from the theory and from extensions developed here. The next section explains the economic model. This is followed by a section which defines mathematical concepts and constructs the main propositions. Interpretations of actual case histories follow each proposition to render the mathematical abstractions more concrete.

Methodology -- The Grossman, Hart & Moore Framework

Incomplete contracts theory, in its original form, considers the problem of asset allocation as a motivator of firm integration. It recommends either integration or divestiture by considering how the ownership of assets affects the creation of value. Assets which create greater value inside a firm should be owned and managed internally whereas assets which create greater value outside a firm are better contracted for than purchased. In the present context, the incomplete contracts model is useful in deciding which distribution of database control maximizes database value. A database partition which creates greater value in the hands of one group should be owned by that group rather than another.

An example serves to illustrate. The Chemical Abstracts Society (CAS) produces a database of chemical compounds with a sophisticated capability for matching one related compound with another. CAS, however, has a small user base, a relatively unsophisticated marketing capability, and limited resources. In contrast, DIALOG Information Services has an enormous user base, sophisticated marketing, and considerable resources. DIALOG serves as a reseller of CAS data but is reluctant to make asset specific investments which improve the user interface or the marketing because it does not own the data it sells. If DIALOG investments were to substantially increase the value of the CAS database, CAS would be in a position to charge more for the data, capturing much of this value. This hold up problem prevents DIALOG from reaping the full benefits of its investment. As a consequence, DIALOG will be less likely to invest than if it owned the data and did not need to share the benefits. In this case, asset value would be maximized if DIALOG were to own the CAS database[1]. On the other hand, if it were important for CAS's uniquely skilled technical staff to make updates and enhancements, then asset value would be maximized by remaining separate.

Grossman and Hart (1986) and Hart and Moore (1990) consider the effects of ownership on investment behavior and define ownership as the residual right to control access to an asset. The "residual" control rights become important to the extent that specific rights have not been contractually assigned to other parties. If a contract were to *completely* specify all uses to which an asset could be put, its maintenance schedules, its operating procedures, associated liabilities, etc. then residual rights of control would have no meaning. All control rights would have been circumscribed by the contract. If, on the other hand, an "incomplete" contract were to fail to anticipate every possible contingency -- a much more plausible situation -- then the residual control provided for by ownership would determine the assets' disposition.

---

[1] In fact, DIALOG attempted to improve certain elements of the user interface only to find that CAS changed the underlying format to render this impossible. CAS feared losing its more profitable core business to its less profitable resale business while it also feared becoming dependent on a single major distributor. The case is currently under litigation. Under these circumstances, DIALOG has little interest in investing further in the database.

Consider, as a further illustration, DIALOG did not anticipate CAS's reaction to increased competition. CAS was prohibited from withdrawing its database completely but exercised a residual right as owner to modify the underlying structure. Whether this violated the spirit if not the letter of an existing contract is open to question, but it has definite implications for investment. Ownership matters when firms must make asset specific investments. The more specific the assets, the more firms prefer to own the assets they use. DIALOG would not necessarily underinvest, for example, if interface improvements benefited all its databases. If the benefits of investment are subject to hold up problems arising from unforeseen events, non-owners will underinvest.

Formally, Hart & Moore 1990 characterize this model in the following manner. Let $V(S, A | X)$ denote the value created by a set (or coalition) S composed of I individuals or agents. A single agent is $i = 1 \ldots I$ and makes an investment x_i. The coalition S also controls assets $a_1, a_2, \ldots a_n \in A$ and makes collective investments $X = (x_1, x_2, \ldots x_I)$ at a cost $C(X)$. The model covers two consecutive periods; in the first period agents choose their investment levels and in the second period they realize the benefits accruing from their investments. Having invested in the first period, value is determined in the second as a function of the agents in the coalition S and the assets A they control given their prior decision to invest X, hence the notation $V(S, A | X)$. Among others, the model includes the following assumptions, letting $(\partial/\partial x_i)V(.) = V^i(.)$:

Assumption 1: $V(S, A | X) \geq 0$, $V(.)$ is twice differentiable and concave in x.
Assumption 2: $C_i(x_i) \geq 0$, $C(.)$ is twice differentiable and convex in x.
Assumption 3: For all subsets $s \subseteq S$, $a \subseteq A$, $V^i(S, A | X) \geq V^i(s, a | X)$

The first two assumptions are standard in economics implying that value per dollar is decreasing while costs are increasing. This guarantees that a solution exists. Assumption three states that the marginal return on investment increases with the number of other agents and new assets in the coalition. The optimal investment levels would then be determined according to the globally efficient levels:

$$\max_{\vec{X}} V(S, A | X) - \sum_{i=1}^{I} C(x_i)$$

The level of compensation granted each member of the coalition, however, is not the total value V(.) but some fractional share $f(S)$ of V(.) based on the size and membership of the coalition. This share is restricted to the range $0 \leq f(S) \leq 1$. For clarity, most subsequent examples will assume that[2] $f(S) = 1/2$. Individual coalition members do not share their respective costs[3]. Continuing the earlier analogy, this is equivalent to assuming that if DIALOG can create $100,000 by investing $x (of uncompensated effort) in marketing, for instance, then it will be able to keep $50,000 (i.e. $f(S)V(\{s_1, s_2\}, A | X) = (1/2)\$100{,}000$). CAS retains the other $50,000 share. Realizing this, DIALOG will only incur expenses up to a maximum of $50,000 even though any amount less than $100,000 generates a profit. Assuming it must share any proceeds, the same argument applies to CAS's investment as well. Formally, each agent, acting in his own interest, chooses to invest according to:

$$\max_{x_i} f(S)V(S, A | X) - \sum_{i=1}^{I} C(x_i)$$

---

[2]Hart Moore (1990) sets $f(S) = (S-1)!(I-S)!/I!$ which is the Shapley value. Using $f(S) = 1/2$ is equivalent to assuming that there are two members in the coalition and each receives half the total benefit. Other values are possible and the results are completely generalizable so long as $f(S) \in [0,1]$.

[3]Ensuing results may be extended to include both shared (contractible) and unshared (non contractible) costs with minor modifications. It does not, however, materially impact the results while it significantly complicates the exposition.

which, by taking first order conditions and letting $f(S) = 1/2$ is:

$$\frac{1}{2}V^i(S, A \mid X) = C^i(X_i)$$

indicating that each agent underinvests. At an intuitive level, the model assumes that today's actions or investments affect tomorrow's payoffs, that asset specificity raises the importance of ownership in light of hold up problems, and that not all actions can be explicitly measured or anticipated. This leads to a realization that persons will underinvest when they are not asset owners and that altering the ownership structure can improve their incentives to invest while correspondingly improving total value.

In the context of database systems, the inability to verify data quality, adequate standardization, usefulness of interfaces and desirable skill sets makes it difficult to specify these features in any meaningful fashion to developers or system administrators. Intangible, unverifiable and non-measurable phenomena are endemic to information and to information systems. Brynjolfsson (1991) argued that these properties make the insights of an "incomplete" contracts approach particularly appealing and derived a number of properties for information ownership by applying the Hart Moore framework. In this paper, we focus on applying the framework specifically to distributed databases.

If one allows that individual departments act to minimize their own costs and effort levels while maximizing their own interests, then the model and its insights may be used to describe the departmental behaviors. It is the incentive principles which govern this behavior, irrespective of technical feasibility, which most concerns the desirability of decentralizing control. The motivations of a central IS organization differ substantially from those of a local department depending on who uses, supplies, and maintains the database system. Different ownership structures generate different behavior and only one structure is likely to maximize database value. The benefit of this research is a coherent set of principles governing centralized, distributed, shared, and joint control over a composite database system. This paper examines, in depth, a number of cases using this framework and formally proves the applicable design principles. Citations, references, and case histories will be used to render the theory concrete.

Effects of Independence and Indispensability

For concreteness, we consider the following case history. This case and others following each represent samples of composite or distributed database systems in actual use.

Case I: In 1990, local branches of a national post office forwarded their operating data to a central office for storage and processing. Managers at the local branch submitted requests for operating reports to the central office. Although both the primary users and suppliers of data were local, this centralized arrangement reduced local equipment costs, it facilitated standardization and in many ways it was consistent with Strategic Data Planning (SDP). It also provided the central office with accounting information to use in controlling local office operations. The central office, however, had little incentive to supply reports to local branches in a timely manner and, being unable to effectively use the delayed reports, branch offices had little incentive to supply accurate or complete data. As a result, neither office received sufficiently useful data for its accounting purposes. As a further disincentive, branches also learned of local problems after the head office had learned of them[4].

This case motivates several normative design principles. The first four parallel Hart & Moore after which we build further upon their framework.

Define "value independence" as a marginal product which is unaffected by access to other groups or assets i.e. for all coalitions $s_i, s_j \subseteq S$ and for all sets of assets $a_i, a_j \subseteq A$

[4]This case is based in interviews conducted in May-July 1991.

$$V^i(\{s_i\}\cup\{s_j\}, \{a_i\}\cup\{a_j\}, X) \equiv V^i(\{s_i\}, \{a_i\}, X)$$

where V^i represents the marginal value contributed by group i.

Design principle 1: Organizations using databases which are value independent should dispense with joint control[5].

Proof Sketch: Consider group s_i and assume that it must share the value it creates but cannot measure its intangible costs to the satisfaction of other groups. Then s_i chooses

$$\max \quad 1/2[V_i(\{s_i, s_j\}, \{a_i, a_j\}, X) + V_j(\{s_i, s_j\}, \{a_i, a_j\}, X)] - C_i(\{s_i, s_j\}, \{a_i, a_j\}, X)$$

which, after first order conditions, is equivalent to

$$1/2V_i^i(\{s_i\}, \{a_i\}, X) = C_i^i(\{s_i\}, \{a_i\}, X)$$

by the definition of value independence. Therefore, each group underinvests. Under independent control, $V_i^i(.) = C_i^i(.)$ and there is no underinvestment.

Interpretation: Design principle 1 suggests that the reason the national post office database system is substandard is that the group responsible for local operations does not own the data it uses. It also supports established research suggesting that data should be stored closest to its most frequent users (Ceri, 1984). In identifying the cause of the problem, design principle 1 also suggests a solution: local branches should control their database partitions. This would both motivate them to supply more accurate and timely data; it would also eliminate the hold up problem of the central office supplying tardy reports. Note that while the local branch is independent of the central office, the central office depends on the local branch. Design principle 2 handles this aspect below.

Define an "indispensable" agent, s_i, as one who is critical to project success in the sense that some asset a_i is nonfunctional without the agent. The marginal product of any group without the indispensable agent is unaffected by whether or not they own the relevant asset. Mathematically, $V^j(S, A, X) \equiv V^j(S, A\setminus\{a_i\}, X)$ if $s_i \notin S$. Assume $V^j(S, A, X) > V^j(S, A\setminus\{a_i\}, X)$ if $s_i \in S$.

Design principle 2: Persons or organizations which are indispensable to the functioning of a database partition should control that partition.[6]

Proof Sketch: If no one controls the asset, then it cannot be used to create value. This case is suboptimal and is not considered further. If someone other than the indispensable agent controls the asset, then in order to make use of it, they must share benefits with the indispensable party. Both agents face $1/2V^j(\{s_i, s_j\}, \{a_i\}, X) = C^j(\{s_i, s_j\}, \{a_i\}, X)$ and again there is underinvestment. Control by the indispensable agent leads to optimal investment $V^i(\{s_i\}, \{a_i\}, X) = C^i(\{s_i\}, \{a_i\}, X)$ by the indispensable party without affecting the incentives of others. This indicates that the indispensable party should own the asset.

Interpretation: In fact, the local branch data is used to support two distinct functions (1) local operations (as in the previous example) and (2) central office cost accounting. For the cost accounting, the

[5]This can be understood as an application of Hart Moore propositions 8 & 10.

[6]This applies Hart Moore proposition 6.

local office is indispensable and design principle 2 indicates that the local office should own this specific partition. The central office may need to pay for the data it requires[7]. This supports research which finds that internal rather than external pressure leads to more active user participation and superior database performance (Osborn, Madnick, & Wang, 1989). In general, agents should assume control of distributed functions for which they are indispensable.

Effects of Cooperative Payoffs and Complementary Assets

Case II: A major midwestern hospital communicates directly with its independent physicians' clinics via a composite information system. The system includes database partitions for patient records at the doctors' offices, pharmaceutical data on inventories and treatment suggestions at the hospital, laboratory test results, and operating room scheduling at the hospital. Additionally, the hospital maintains a database of specialty practitioners for doctor to doctor, hospital to doctor, and doctor to hospital referrals. Parties trade information in both directions.

Define "cooperative payoff" as a marginal product which increases with the inclusion of outside agents[8]. In this case

$$V^i(\{s_1\} \cup \{s_2\}, A, X) > V^i(\{s_1\}, A, X).$$

Design principle 3: Cooperative payoffs are necessary but not sufficient to justify sharing database control.

> *Proof Sketch*: By design principle 1, unless there is a cooperative payoff, control should be independent. But given a cooperative payoff, it may be the case that $1/2V^i(\{s_1\} \cup \{s_2\}, A, X) > C^i(\{s_1\}, A, X)$ implying that even though a payoff is shared, the marginal benefit of working together is worth the trouble. This is not necessarily sufficient, however, since this requires both $1/2V^i(\{s_1\} \cup \{s_2\}, A, X) = C^i(\{s_1\} \cup \{s_2\}, A, X) > C^i(\{s_1\}, A, X)$ and $V(\{s_1\} \cup \{s_2\}, A, X) > V(\{s_1\}, A, X)$ over some useful range of investment.

Interpretation: In case II, both the independent doctors and the hospital realize benefits unattainable by either party acting independently. Doctors receive patient referrals, information on new treatment methods, and guidance on under-served neighborhoods where they might establish new clinics. The hospitals in turn receive surgical referrals, improved patient scheduling, a broad sampling of treatment success rates, and improved access to specialty services. The database design issue is to seek and to surpass a given threshold of cooperative payoffs. Without sufficient mutual benefit, the coordination costs of a distributed system may not justify multiple party control. Cooperative payoffs are a prerequisite to shared ownership. Consider now complementary assets.

Define "complementary assets" as assets which have considerable value together but which have little or no value apart. Mathematically, suppose $a_m, a_n \in A$, then

$$V^j(S, A\setminus\{a_m\}, X) \equiv V^j(S, \{A\setminus\{a_n\}, X) \equiv V^j(S, A\setminus\{a_m, a_n\}, X) < V^j(S, A\, X)$$

Design principle 4: Database partitions which are complementary should be controlled together.[9]

---

[7]Design principle 5 considers the case where the central office is also indispensable.

[8]This definition is consistent with assumption 3.

[9]This applies Hart Moore proposition 8.

Proof Sketch: Let $\underline{A} \equiv A \setminus \{a_m, a_n\}$. If the assets are controlled by separate groups which do not cooperate (asset a_m is not used with asset a_n) then individual group s_i faces $V^i(\{s_i\}, \underline{A} \cup \{a_m\}, X) = C^i(\{s_i\}, \underline{A} \cup \{a_m\}, X) = V^i(\{s_i\}, \underline{A}, X)$ and each is just as well off as in the case where they do not own either asset. But if the assets are controlled by different groups which do cooperate then each faces $1/2V^i(\{s_i, s_j\}, \underline{A} \cup \{a_m, a_n\}, X) = C^i(\{s_i, s_j\}, \underline{A} \cup \{a_m, a_n\}, X)$. This is less than the case where there are no hold up problems and only agent s_i controls the asset pair i.e. $V^i(\{s_i\}, \underline{A} \cup \{a_m, a_n\}, X) = C^i(\{s_i\}, \underline{A} \cup \{a_m, a_n\}, X)$. As before, the maximum incentive is centralized control indicating that the benefits are undivided.

Interpretation: Consider the pharmaceuticals database. It includes data on inventories and data on treatment methods, two databases which are strictly complementary. There is little merit to prescribing treatments which are unavailable or to stocking drugs which are outdated treatments. To provide the maximum practical incentive, the data should be controlled by the same agent rather than distributed among multiple agents.

Situations of Suboptimality, Standards, and the Need for New Strategies

Design principle 5: If assets are strictly complementary and agents are indispensable, then the presence of private cost information implies that there is no distribution of database control which is globally optimal.

Proof Sketch: Given that assets are complementary and agents are indispensable, the only optimal solution is to include both assets and both agents in a coalition such that $V^i(\{s_i, s_j\}, \{a_m, a_n\}, X) = C^i(\{s_i, s_j\}, \{a_m, a_n\}, X)$. This requires the agents to collectively choose investment X. However, each party shares only half of the total value created while each fully bears his uncompensated costs x_i. At the optimum, marginal costs equal marginal value. Each agent can therefore reduce his investment and save himself Δx while only reducing his share of the benefit by $1/2\Delta v \approx 1/2\Delta x$. In other words, each agent may make money by investing less. This implies that at the optimum, agents collectively invest less than the optimum which is a contradiction.

Interpretation: At the national post office, high level accounting functions require a central organizational perspective and specialized resources. The central office is indispensable to these resources so by design principle 2, the central office should control them. The complementary data, however, are required from the local branches and so by design principle 4, the local branches should control these resources. Since there is minimal cooperative payoff -- data flows in one direction -- design principle 3 does not apply. Accordingly, there may not exist a globally optimal solution.

Define a "standardized" relative to a non-standardized database as one which has lower marginal costs of integrating new resources into the system. Formally, letting $\underline{C}$ and $\overline{C}$ represent the costs of the minimally standardized and highly standardized systems respectively, this is:

$$\overline{C}^i(S, \{A \cup a\}, X) < \underline{C}^i(S, \{A \cup a\}, X)$$

This enables us to derive a new principle and a corollary of particular interest to database systems.

Design principle 6: Increasing standardization leads to more efficient use of *shared* resources.

Proof Sketch: To illustrate the proof, we divide investments X into those with hidden or private costs χ and those with costs visible to outside departments K such that $X = \chi + K$. Thus $V(.) = V(S, A \mid \chi + K)$ and $C(.) = C(\chi + K) \approx C(\chi) + C(K)$. Assuming that if both groups support the same standard, then information asymmetries are reduced and each group can reasonably expect to be reimbursed for certain mutually recognized costs, individual departments will solve for $1/2(\partial/\partial\chi)V(S, \{a_m \cup a_n\} \mid \chi) = (\partial/\partial\chi)C(\chi)$ and $(\partial/\partial K)V(S, \{a_m \cup a_n\} \mid K) = (\partial/\partial K)C(K)$ and departments underinvest due to uncompensated costs χ^i. If costs are monotonic and χ accounts for the nonstandard costs then costs are reduced for the highly standardized system $\overline{C}^i < \underline{C}^i$ and the distortion due to asymmetric information falls. The implication for ownership is that standardization increases the potential for efficient decentralized control.

Corollary : Increasing standardization does *not* lead to more efficient use of *un*shared resources. In this case $V^i(S, \{a_m \cup a_n\}, X) = C^i(S, \{a_m \cup a_n\}, X)$ and there is no distortion due to sharing. The decision to standardize will depend on the expected integration of future assets within a single department.

Interpretation: Design principle 6 suggests that the benefits of standardization are only partially due to the decreased costs of connecting multiple resources in an information system --the usual reason given for standardizing platforms and software. An alternate explanation is that standardization increases the shared knowledge of the participants enabling them to determine costs more accurately. Design principle 6 and its corollary predict that the patient records which are shared between clinics and the hospital would be standardized. Individual clinics' billing information, however, which doctors do not share with one another (and which they may be legally barred from sharing) ought not to be standardized.

The foregoing principles concern the issue of when to centralize and when to decentralize the ownership of a database system. They may be summarized in the following manner.

Local Owner		Central Owner
←	Local Group Skills Indispensable to Database	
←	Local Group's Investment Critical	
	Central Group Skills Indispensable to Database	→
	Central Group's Investment Critical	→
←	Local Asset a_l Independent of Centrally Controlled Asset a_c	
	Local Asset a_l Complements Centrally Controlled Asset a_c	→
←	Database Standardization Increasing (enables greater decentralization)	

Figure 2 – Distributed database design needs to consider these issues regarding the ownership of a local group database project.

The six design principles lead to the realization that new technical strategies may be necessary. The Composite Information System (CIS) approach (Madnick, Siegel, & Wang, 1990), for example, allows for the possibility of the central group accessing the local group's data in a non-intrusive manner. This results in minimal cost or disruption to the local group who continues to accrue the full benefits of local control. The central group incurs only the costs of linking the CIS technology using context interchange technology (Siegel & Madnick, 1991). Data update and maintenance costs are not duplicated. Linkage expenses, however, are generally much lower than the benefit of direct access to updated and accurate data. The ability to share information permits multiple users and beneficiaries of a distributed database without necessarily multiplying the costs.

Conclusion

This research details several normative design principles governing the distribution of control rights over a distributed database system. It also provides cases and interpretations indicating that an incomplete contracts paradigm is applicable to the problems of an IS organization. We have adapted a general theory to a specific application, interpreting economic constructs as variables in database systems. Additionally, the model captures intangible control rights and other features which have proven elusive to formalize in the past.

Above all else, this research points out that in addition to technical issues, there are ownership and control issues that systems designers need to confront when building actual systems. The interests and motivations of people working with and managing databases will affect realized performance levels. Certain basic ideas, which may be implicit in emergent systems, are here formally captured and made explicit.

The model is also amenable to other interpretations and we are continuing to develop new principles and to adapt it to related areas within information technology. Progress has been made in applying the framework to the problem of outsourcing, for example. Since it considers intangibles, the framework can be used to address qualitative issues of the decision to outsource over and above the hardware, software, and bandwidth concerns of working through independent organizations. We hope to propose new theories in this area in future research.

Bibliography

Brynjolfsson, E. (1991). An Incomplete Contracts Theory of Information, Technology, and Organization (Center for Coordination Science, MIT Sloan School of Management.

Ceri, S. & Pelagatti, G. (1984). Distributed Databases Principles and Systems. New York: McGraw-Hill.

Chen, P. (1976) The Entity Relationship Model -- Toward a Unified View of Data. T.O.D.S 1(1); 166-193.

Date, C. (1981) An Introduction to Database Systems, Reading, MA: Addison-Wesley.

Goodhue, D. L.,Kirsch, L. J.,Quillard, J. A., & Wybo, M. D. (1992). Strategic Data Planning: Lessons from the Field. MIS Quarterly(March), 11-33.

Grossman, S. J., & Hart, O. D. (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. Journal of Political Economy, 94(4), 691-719.

Hart, O., & Moore, J. (1990). Property Rights and the Nature of the Firm. Journal of Political Economy, 98(6), 1119-1157.

Madnick, S. E.,Siegel, M., & Wang, Y. R. (1990). The Composite Information Systems Laboratory (CISL) Project at MIT. IEEE Data Engineering, 13(2), 10-15.

Martin, J. (1982) Strategic Data Planning Methodologies. Englewood Cliffs, NJ: Prentice-Hall.

Notkin, D, Hutchinson, N., Sansilo, J., & Schwartz, M. (1987) Heterogeneous Computing Environments: Report on the ACM SIGOPS Workshop on Accommodating Heterogeneity. Communications of the ACM, 30(2) 33-43.

Osborn, C. S.,Madnick, S. E., & Wang, Y. R. (1989). Motivating Strategic Alliance for Composite Information Systems: The Case of a Major Regional Hospital. In Proceedings of the Twenty Second Annual Hawaii International Conference on System Sciences, . Honolulu:

Richmond, W. B.,Seidmann, A., & Whinston, A. B. (1991). Incomplete Contracting Issues In Information Systems Development Outsourcing. Decision Support Systems Journal (forthcoming).

Shuey, Richard & Gio Wiederhold, (1986) Data Engineering and Information Systems. IEEE Computer, Jan. 18-30

Siegel, M., & Madnick, S. E. (1991). Context Interchange: Sharing the Meaning of Data. SIGMOD Record, 20(4), 77-79.