Why Not One Big Database?

Principles for Data Ownership

Authors: Marshall Van Alstyne
                                 Erik Brynjolfsson
                                 Stuart Madnick

Affiliation: MIT Sloan School

Correspondence Address: Marshall Van Alstyne
                                 MIT Sloan School
                                 Rm. E53-308
                                 30 Wadsworth Street
                                 Cambridge, MA 02139
                                 marshall@athena.mit.edu

Why Not One Big Database?

Principles for Data Ownership

Abstract: Results of this research concern incentive principles which drive information sharing and affect database value. Many real world centralization and standardization efforts have failed, typically because departments lacked incentives or needed greater local autonomy. While intangible factors such as "ownership" have been described as the key to providing incentives, these soft issues have largely eluded formal characterization. Using an *incomplete contracts* approach from economics, we model the costs and benefits of restructuring organizational control, including critical intangible factors, by explicitly considering the role of data "ownership." There are two principal contributions from the approach taken here. First, it defines mathematically precise terms for analyzing the incentive costs and benefits of changing control. Second, this theoretical framework leads to the development of a concrete model and seven normative principles for improved database management. These principles may be instrumental to designers in a variety of applications such as the decision to decentralize or to outsource information technology and they can be useful in determining the value of standards and translators. Applications of the proposed theory are also illustrated through case histories.

1.1 Introduction: "Why not one big database?"

Information systems designers often argue that centralized control is better control. From a technology standpoint, this is readily defensible in terms of data integrity and enforcing a uniform standard. From an economic standpoint, centralization limits the costs of redundant systems. In addition, stories of confusion sometimes characterize decentralization. One senior executive at Johnson and Johnson waited three weeks for the list of his corporation's top 100 customers world-wide due to problems linking multiple systems. Difficulties with "dis-integrated" systems have led senior staff to inquire "Why not create one big database or at least control them all from one central location?" With optical technology and newer microprocessors, barriers imposed by communications bandwidth and speed-bound central hardware continue to fall. Local data control no longer seems necessary or warranted.

Technical considerations, however, represent only part of a more complex story in which less tangible managerial and incentive issues play a critical role. We present a framework demonstrating that local control can be optimal even when there are no *technical* barriers to complete centralization. This assertion is based on research showing that "ownership" is a critical factor in the success of information systems.

In developing an "interaction theory" of people and systems, Markus observes that problems with a database at a large chemical company arose from changes in control. After implementing a new information system, "all financial transactions were collected into a single database under the control of corporate accountants. The divisional accountants still had to enter data, but they no longer *owned* it." [19 p. 438][1]

---

[1]Emphasis is the original author's.

Similar arguments are put forth by Maxwell [21] and Wang [30]. Of the factors Maxwell considers most important to improving data quality, data ownership and origination are among the most critical. Spirig argues that when data ownership and origination are separated, information systems cannot sustain high levels of data quality. [30 Cited in Wang p. 31]   Ralph Larsen, the CEO of Johnson and Johnson, states unambiguously, "We believe deeply in decentralization because it gives a sense of ownership."[7]

The key reason for the importance of ownership is self-interest: owners have a greater vested interest in system success than non-owners. Just as rental cars are driven less carefully than cars driven by their owners, "feudal" databases -- those not owned by their users -- are maintained less conscientiously than databases used by their owners.

Ignoring ownership is also one possible explanation for IS failures since the impetus for system development is external to the groups being affected. In fact, evidence suggests that most top-down strategic data planning efforts never meet expectations [11]. Orlikowski [23] has observed that employees in a major consulting firm refused to share information despite senior management encouragement, company-wide introduction, and an industry standard group support tool. Culture and incentives opposed the knowledge transfers which the technology was designed to support. In the words of one IS practitioner, "No technology has yet been invented to convince unwilling managers to share information. . ." [9 p. 56]   Information assets have simply become too valuable to give away.

The issues highlighted in these studies [9, 11, 19, 23] are organizational not technical. Prior to deciding on the implementation of features and functionality, it becomes necessary to ask who should have the power to decide? Will an outsourcing contractor decide on system features which are in the strategic interests of the firm?

Will one department sufficiently value the interests of another regarding database integrity?   These questions link technology issues to management concerns at a fundamental level. In response, we develop the concept of data ownership to provide a mechanism for ensuring that key parties receive compensation for their efforts.

This is developed into two separate contributions. First, a rigorous model gives mathematical definitions of non-technical costs and benefits arising from changes in database control. Using the "incomplete contracts" approach pioneered by Grossman and Hart [12] and applied to information assets by Brynjolfsson [5], it formalizes intuitive concepts of independence, ownership, standardization, and other intangibles that affect system design and that have generally eluded precise specification. The results are therefore testable and less ambiguous. Second, we use the model to construct normative database principles that solve problems caused by the separation of ownership from use. This leads us to propose seven database design principles based on ownership to complement existing design principles based on technology.

The remainder of this introduction carefully defines ownership and situates it among the broader issues of database design with references to existing literature. Section two explains the economic model. It defines the mathematical concepts and the assumptions used to construct the database design principles. Following these formulation arguments, section three discusses the role of ownership given complementarities among databases and given critical or indispensable personnel. Section four deals with the effects of ownership in the context of database standards and the decision to outsource design and maintenance. This is followed by section five which examines tradeoffs among conflicting design principles and proposes a solution to a lack of ownership incentives in decentralized systems. Throughout each of these

five sections, case histories provide context and interpretation in order to simplify the application of the model to real world database design.

1.2 Database Architecture and the Definition of Ownership

To place ownership among the technical and non-technical aspects of database architecture, we propose that database design involves at least three major dimensions -- system components, development, and control. These are depicted in Figure 1. The first dimension, components, includes the literal parts of the system hardware, software, and network connections.[2] The second axis, development, concerns procedural aspects of programming and implementation.[3] The third issue, control, describes the rights and responsibilities of the parties involved in the database system. This includes, for example, the authority to set standards and to approve system modifications and hardware acquisition.[4]

One distinguishing design element, that cuts across all axes, is the degree of database concentration. In principle, each dimension can be independently centralized or decentralized.   As shown in the diagram, the origin represents maximal centralization, whereas moving outward along any given axis represents increased decentralization. Since two of these dimensions, components and development, have received attention from several important contributions to the research literature. This paper focuses on unaddressed issues of control.

---

[2]Technical issues of network protocols covering modular design and layering of abstraction levels are summarized in [28] and [29]. Additional issues of concurrency control covering serializability, record locking, and recovery are also described in [2] and [3].

[3]For a reference on software measurement issues see [10] and for assessing project risk and complexity [4, 16]. Specific issues of relational database design and data manipulation are covered in [8] and [6], Issues of cooperative software development are covered in [15]. Improving development through software reuse is described by [17].

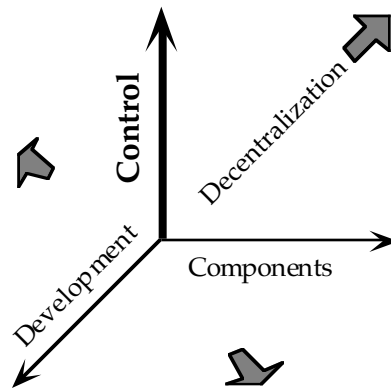[4]Control aspects of strategic data planning appear in [20].

Figure 1 -- Of the three main axes to decentralization, we focus on control.

Components: All computing and data storage equipment can be centralized at one location, with world-wide access provided via remote terminals. An automatic teller machine (ATM) network is an example. Alternatively, the computing and data storage equipment can be decentralized. For instance, a global brokerage firm might provide a workstation to each of its traders – but each workstation might run software developed by a central group.

Development: Development may be performed by a central group or by each local department regardless of equipment location. "A decision to use one central computer, for example, does not necessarily imply centralizing systems development. Conversely, a decision to centralize all development ... does not compel the organization to use one ... computer." [26 p. 16] Individual departments might even contract for development from the central group but then own the finished products.

Control: Control of the databases, planning, and application programs may be *centralized* to a corporate data center that "owns" the system irrespective of equipment location. Traditionally, this has been the finance department or a corporate resource center. Local divisions would then defer to this central authority for all IS functions. Alternatively, control might be *decentralized* to local divisions. Under decentralized

control, divisions might contract via a "chargeback" system for data center resources or they might assume completely independent responsibility for their IS resources. Each of these options has been observed in practice.

We consider control to be *centralized* if a corporate data center retains the right to make any decision not explicitly and specifically delegated to others. Adopting Grossman and Hart's [12] use of terminology, we refer to this as the "residual right of control" and associate it with ownership of the system.

For databases, "ownership" and "use" are easily confused as both connote privileges ranging from read and query access to creation and modification rights. By *usage* rights, we mean the ability to access, create, standardize, and modify data as well as all intervening privileges. Usage, however, is not what is meant by ownership. We use *ownership* and the *residual right of control* to mean the right to determine these privileges for others. The ownership archetype is a single database controlled and operated by a single department with no outside access. This group, which exercises control over format, access, standards, etc., is the exclusive owner. It may then grant successively more permissive access to outsiders until the effective usage privileges of outsiders resemble the usage rights of the owner. It is the authority, however, to subsequently alter or retract these privileges that distinguishes the owner from a non-owner. If the ability to alter others' access is interfered with or vetoed, perhaps by a central authority, then the original owner is not, by our definition, the sole owner of the database. Subsequent design principles answer the important managerial question: "Who should own the data?"

2.1 Background: Incomplete Contracts in a Database Context

Incomplete contracts theory, considers asset allocation as a cause for firms' integration. Firms should either acquire or divest assets by considering how ownership of these assets affects incentives for the creation of value. When owning an asset induces higher investment and higher realized value, a company should purchase that asset and manage it internally. However, when an asset creates greater value in the hands of others, a company is better off contracting for that asset from the market and then it should not own that asset. Although Hart and Moore consider residual rights to be synonymous with firm boundaries, we follow Brynjolfsson [5] and argue that the concept can also apply to intra-firm database transactions. This is because effective ownership of information rarely accrues solely to its nominal legal owners, the stockholders of the firm. More realistically, various groups within the firm are the *de facto* owners with residual rights of control that can be transferred by changes in organizational structure or management edict. In the present context, the incomplete contracts model is useful in deciding which distribution of database control maximizes database value.

Grossman and Hart [12] and Hart and Moore[13] 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 determined 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' use under circumstances where control had been left unspecified.

Ownership issues, in fact, arise with considerable frequency as illustrated by the conflicting interests of two vendors of database search services. The Chemical Abstracts Society (CAS) produces a database of chemical compounds with a sophisticated capability for matching one related compound with another. CAS, however, initially had a smaller user base, a less sophisticated marketing capability, and limited resources. In contrast, DIALOG Information Services had an enormous user base, sophisticated marketing, and considerable resources. As a value added reseller, DIALOG can repackage CAS data but is reluctant to make asset-specific investments which might improve the user interface or the marketing of the chemical database because it cannot claim ownership of the data it sells. If DIALOG investments were to substantially increase the value of the CAS database, CAS would be in a position to extract a sizable portion of any increased profits. As owner, CAS could restrict access to the database unless DIALOG agreed to share the incremental profits even if DIALOG were the sole investor in any new project This is the classic "hold-up" problem. As a consequence, DIALOG is less likely to invest than if it owned the data and had no need of sharing its profits.

Under these circumstances, total asset value would be increased if DIALOG were to own the chemical database. DIALOG would invest up to the product's full potential. On the other hand, there might also be reasons not to transfer ownership. If it were true that only CAS's chemically sophisticated staff were capable of making enhancements or that transfer foreclosed other resellers' investments, then asset value would be maximized by leaving ownership with CAS, thereby preserving existing incentives. The point is that different incentive requirements lead to different ownership results.

Our model captures these and other tradeoffs for databases inside a company where such allocation decisions are more easily made.

There is a further complication, however, relating to the verification of DIALOG's investment. If DIALOG's contributions were easily and completely documented, then DIALOG could be fully compensated. But what if these contributions are intangible or difficult to measure such as brand name equity, executive expertise, strategic positioning, or interface quality? Then DIALOG can never be certain that deploying its assets to benefit CAS products will be in DIALOG's own best interests. DIALOG would be unable to document its contribution and would instead be required to expend resources in costly negotiation -- a situation that changes if DIALOG were to own the database.

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 advance in any meaningful fashion to developers or system administrators. Intangible, unverifiable and non-measurable phenomena are endemic to information and to information systems. Deprived of measurement instruments, technology solutions handle intangible issues poorly. Brynjolfsson [5] argued that these properties make the insights of an "incomplete" contracts approach particularly appealing in this domain and derived a number of properties for information ownership by applying the Hart-Moore framework.

In fact, DIALOG did attempt to improve certain elements of its own version of the user interface despite CAS's control of key unspecified parameters of the database. Shortly thereafter, 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 with DIALOG suing precisely over denial of access [22]. CAS was prohibited by contract from withdrawing its database completely, but exercised a residual right as owner to modify the underlying structure. This did not violate the letter of the existing contract, but it has definite implications for investment incentives. Ownership matters when firms must make asset-specific investments. The more specific the assets, the more firms prefer to own the assets in which they invest. If the benefits of investment are subject to hold-up problems by owners -- problems which arise from unforeseen events -- non-owners will underinvest.

2.2 Methodology: The Grossman, Hart & Moore Model

Formally, Hart and Moore [13] model ownership in the following manner. Let $V(S, A | X)$ denote the total value created by the full set (or grand coalition) S of agents who control assets A and have previously chosen to invest X. The grand coalition S of all individuals I can be broken into any subset s. A single agent is indexed by $i = 1 \dots I$ and makes an investment x_i. The coalition s also controls assets $a_1, a_2, \dots a_n \in A$ and makes collective investments $X = (x_1, x_2, \dots x_I)$ at a cost $C(X)$. An ownership map α describes the control s exercises over its assets written as $\alpha(s) = \{a_1, a_2, \dots a_n\}$.

The model covers two consecutive periods. In the first period agents choose their investment levels; in the second period they realize the benefits accruing from their investments and divide the benefits in proportion to their bargaining power. Having invested in the first period, value is determined in the second as a function of the agents in the coalition $s \subseteq S$ and the assets $a \subseteq A$ they control given their prior decision to invest $x = (x_{i1}, x_{i2}, \dots x_{in})$, hence for a single coalition the notation is $V(s, a | x)$. The Hart-Moore model includes the following assumptions, letting $V^i(.) \equiv (\partial / \partial x_i) V(.)$:

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: $V^i(s, a \mid x) = 0$ if $i \notin s$.

Assumption 4: $(\partial / \partial x_j)V^i(S, A \mid x) \geq 0$ for all $j \neq i$.

Assumption 5: For all subsets[5] $a \subseteq A$, $s \subseteq S$, $V(S, A \mid x) \geq V(s, a \mid x) + V(S \backslash s, A \backslash a \mid x)$

Assumption 6: For all subsets $a \subseteq A$, $s \subseteq S$, $V^i(S, A \mid x) \geq V^i(s, a \mid x)$

The first two assumptions are standard in economics implying that marginal value per dollar is decreasing while marginal costs are increasing. Together, these assumptions permit the use of first order conditions to locate a unique solution. The third assumption implies that an agent's marginal investment affects only coalitions to which he belongs and no other. In assumption four, one agent's investments are complementary at the margin with those of another. Assumption five implies that groups working together create at least as much value as working apart, while assumption six states that the marginal return on investment increases with the number of other agents and new assets in the coalition. Together, assumptions five and six imply that marginal and total values correlate with one another. The optimal investment levels would then be determined according to the globally efficient levels:

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

Additionally, the model allows for substitution of the governance structure α for assets from A that the coalition controls. This leads to rewriting the value function as $V(s, \alpha(s) \mid x)$. The level of compensation granted each individual member of the coalition, however, is not the total value V(.) but some portion p(s) of V(.) based on the members in the group s. Following Hart and Moore [13] the subsequent examples will

---

[5]Reader's Note: The notation $s \backslash \{i\}$, from set theory, is used to designate the removal of element i from the set s. If i is a set then $s \backslash i$ will be used and if i was not originally contained in s then this represents a null operation. In conjunction with the ownership map α, for example, the expression "$\alpha(s \backslash \{i\})$" means the collection of assets owned by a group of which i is not a member.

assume that p(s) is the reduced form probability term from the Shapley value[6]. The intuition behind the Shapley value is that it represents each agent's bargaining power in terms of a percentage of the total value created. Bargaining power varies with value contributed and with assets controlled. Persons who contribute more or who control more assets receive a higher percentage of the benefits.

Despite sharing total value, individual coalition members do not share all their respective costs. Due to a lack of verifiability, certain intangible costs are not contractible. Unreliable software metrics, unknown training requirements, disputed opportunity costs, and spent political capital might fall into this category. Lack of agreement and verifiability means that these costs cannot be directly compensated and therefore group members will not incur them unless receipts exceed them. Costs that are verifiable can be directly compensated according to terms set forth in a contract. Ownership will not affect such costs and so initially we focus only on unverifiable costs. We explicitly reintroduce verifiable costs with Design Principle Four. Continuing the earlier example, these cost conditions imply that if DIALOG can create $100,000 by investing $x of unverified effort in marketing the database owned by CAS, then it will have no recourse for being directly compensated for the $x of investment. However, it will be able to bargain ex post for half the $100,000 of benefits[7] or $50,000. CAS has the bargaining power to insist on the other $50,000 share. Realizing this, DIALOG will only incur expenses up to a maximum of $50,000 even though any investment less than $100,000 would generate a profit. This result holds so long as DIALOG and CAS cannot

---

[6]The full function is actually a fractional share $f(i, s, \alpha)$ which is based on the individual, the membership and on the assets each member controls. For specifics of this function, see the Appendix. Any monotonic decision rule will leave the following propositions unaffected, however, so long as payoff is increasing in the control of additional assets and in contributed value.

[7] Letting DIALOG = i, CAS = j, and investment = x, assets controlled by DIALOG and CAS are $\alpha(i, j)$ and the full functional form for DIALOG is $f(i, \{i, j\}, \alpha(i, j)) = \Sigma p(s)V(\{i, j\}, \alpha(i, j) \mid x) = (1/2)\$100,000 = \$50,000$.

write a contract based on the size of DIALOG's investment. Formally, an agent acting in his own self interest will choose to invest according to:

(2) $$\max_{X_i} \sum_{s|i \in s} p(s)[V(s, \alpha(s)\,|\,x) - V(s\backslash\{i\}, \alpha(s\backslash\{i\})\,|\,x)] - C(x_i)$$

This states that individuals profit according to their value added, i.e., the difference in value created with and without their participation net of costs. Their share of total returns increases as their inputs and assets contribute to the group's output. We also assume that individuals will invest only to the point at which private marginal cost equals private marginal benefit (MC = MB) which is not the same as the group's marginal value (MC ≠ MV). After taking first order derivatives and using assumption three to reduce the second term, this becomes:

(2a) $$\sum_{s|i \in s} p(s)V^i(s, \alpha(s)\,|\,x) = C^i(x_i)$$

Because $\sum_{s|i \in s} p(s) \leq 1$, this result indicates that the lefthand side is at most V^i(s, a | x) and therefore each agent underinvests. At an intuitive level, the model combines three key insights. First, today's actions or investments should affect tomorrow's payoffs, i.e., V depends on x. Second, since share rises with assets controlled, asset ownership matters as an investment incentive. This means that i will invest a smaller x_i if j controls critical asset a_i which is essential to i's final product. Third, since not all actions can be explicitly measured or anticipated and costs $C(x_i)$ are sunk before V is realized, transferring ownership beforehand can alter and improve investment incentives. In sum, altering ownership structure can improve total value. This simple rule leads to our subsequent propositions.

In this paper, we focus on applying the model specifically to decentralized databases. Of the following design principles, the first three are direct applications of propositions that were proven by Hart and Moore [13], which consider only intangible

costs. Building upon this basic framework, we subsequently relax the assumption of no tangible costs, and the relaxed program of equations leads to design principles four through seven.

3.1 Effects of Independence and Indispensability

For concreteness, we consider a pair of case histories. The following case represents a system whose ownership is concentrated in the hands of a central authority while its input operations are decentralized to satellite groups. The inherent conflict in this organizational structure serves to illustrate several issues of control. Each case describes an operational database system. This one is based on interviews conducted in May-July 1991.

Case I: In 1990, local branches of a national post office forwarded their operating data to a central office for storage and processing. Needing data for their own operations, local managers submitted requests for summary reports to the central office. Differences in data requirements emerged, however, since financial and management accounting needs diverged. 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 financial accounting information to use in gauging postal efficiency. The central office, however, had little incentive to supply management accounting 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. Consequently, neither office received sufficiently useful data for its accounting purposes. Also, as a further disincentive to supply accurate data, local branches learned of their internal problems only after the head office had learned of them.

One of the main issues of this case is that the central office provides negligible value to the branch offices in exchange for their operating data. In effect, branches have simply been ordered to produce data according to a given set of standards. This independence of value leads to the proposition below.

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

$$V^i(\{i\} \cup s, \{a_i\} \cup a_s \mid X) \equiv V^i(\{i\}, \{a_i\} \mid X)$$

where V^i represents the marginal value contributed by agent i. This may be interpreted to mean that marginal value is the same regardless of participation or non-participation by other agents.

Design Principle 1: Organizations using databases which are value independent should dispense with joint control.

Proof:[8] Consider group i and assume that it must share the value it creates but cannot measure its intangible costs to the satisfaction of other groups. Then i chooses

(3) $$\max_{X_i} \sum_{s \mid i \in s} p(s)[V(s, \alpha(s) \mid X) - V(s \setminus \{i\}, \alpha(s \setminus \{i\}) \mid X)] - C(x_i)$$

which, after applying first order conditions, reduces to

(3a) $$\sum_{s \mid i \in s} p(s)V^i(\{i\}, \{a_i\} \mid X) = C^i(x_i)$$

by the definition of value independence. The lefthand side is at most $V^i(\{i\}, \{a_i\} \mid X)$ and therefore group i, who must share its assets, will underinvest. By assumption 3, i's investments have no effect on the investments of any other group j when they are not in the same coalition so j's incentives are no worse under independent control and value independence. Under independent control, however, i retains his benefits since $\sum_{s \mid i \in s} p(s)V^i(.) = V^i(.)$ and there is no underinvestment.

---

[8] After Hart-Moore proposition 10.

Interpretation: Design Principle One requires that there be a cooperative payoff for joint control to be beneficial. The reason the post office database system performs badly is that the group responsible for local operations does not own the data it uses. The solution is to pass control of local partitions to local branches.   This would both motivate them to populate their database with more accurate and timely data; it would also eliminate the hold-up problem of the central office supplying tardy reports. Design Principle One also supports established research suggesting that data should be stored closest to its most frequent users [6]. Note that while the local branch is independent of the central office, the central office depends on the local branch. Design Principle Two handles this aspect below.

Define an "indispensable" agent, 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, $V^j(s, a \mid x) \equiv V^j(s, a \setminus \{a_i\} \mid x)$ if $i \notin s$.

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

> *Proof:*[9]   Consider giving ownership of asset a_i to i.   As new owner, i's incentives are at least as great as before. For any j the change in incentives is the difference between the new and old control structures with the asset transferred:
>
> (4)
> $$\sum_{\substack{s \mid i,j \in s \\ \wedge\ a_i \notin \alpha(s)}} p(s)[V^j(s, \alpha(s) \cup \{a_i\})\ -\ V^j(s, \alpha(s))] - \sum_{\substack{s \mid i \notin s, j \in s \\ \wedge\ a_i \in \alpha(s)}} p(s)[V^j(s, \alpha(s))\ -\ V^j(s, \alpha(s) \setminus \{a_i\})]$$
>
> As a_i is useless to j without i, and by assumption 6, the second summation is zero. Group j only benefits from working with both i and a_i. If a_i were owned by a third party k, however, then j would have had to work with {i, j, k} but this introduces an additional hold-up, lowering j's incentives.

---

[9] After Hart-Moore proposition 8.

Interpretation: In fact, the local branch data is used to support two distinct functions: (1) local operations and (2) central office cost accounting. In both cases, the local office is indispensable and Design Principle Two indicates that the local office should own this specific partition. If the central office were also indispensable, there would be a conflict -- a possibility which is addressed in Design Principles Six and Seven. The effect of transferring ownership to the local office also supports research which finds that internal rather than external pressure leads to more active user participation and superior database performance [24]. In general, agents should assume control of decentralized functions for which they are indispensable.

3.2 Effects of Complementary Assets

Case II: A major midwestern hospital communicates directly with its independent physicians' clinics via a decentralized 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 "complementary assets" as assets which have great value together but which have negligible value apart. Mathematically, suppose $a_m, a_n \in A$, then

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

Design Principle 3: Database partitions which are complementary should be controlled together.

Proof:[10] Again, consider the transfer of asset a_n to a group that already owns complementary asset a_m . The increase in value is given by

---

[10]After Hart-Moore proposition 6.

$$(5) \sum_{\substack{s| \; i \in s \; \wedge \; a_m \in \alpha(s) \\ \wedge \; a_n \notin \alpha(s)}} p(s)[V^i(s, \alpha(s) \cup \{a_n\}) \; - \; V^i(s, \alpha(s))] - \sum_{\substack{s| \; i \in s \; \wedge \; a_m \in \alpha(S \backslash s) \\ \wedge \; a_n \in \alpha(s)}} p(s)[V^i(s, \alpha(s)) \; - \; V^i(s, \alpha(s) \backslash \{a_n\})]$$

As the assets are complementary and considering assumption three, the second summation is zero. There is no loss of investment incentives to the present owner of a_n due to the transfer. The receiving party, in contrast, has strictly higher incentives to invest indicating a net gain in total welfare. Equivalently, a_m could have been transferred in the other direction thereby increasing the other party's incentives.

Interpretation: Consider the pharmaceuticals database. It includes partitions both for inventories and for treatment methods, two databases which are strictly complementary. There is little merit in prescribing treatments which are unavailable or in 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. The hospital does, in fact, control both databases in this more successful system.

4 The Use of Standards and Outsourcing as Control Issues

4.1 Considering both Contractible and Non-Contractible Costs

For several of the principles which follow, we relax assumptions of the basic model to extend its scope and to generalize the insights from more theoretical to more applied tasks. Specifically, while the preceding propositions assume that all costs are non-contractible, subsequent propositions allow costs to be divided into verifiable, contractible or tangible costs τ (tau) and into unverifiable, non-contractible or intangible costs $\Im$ (iota). Earlier design principles hold given contractible costs in addition to uncontractible costs, but the exposition and proofs become more complex. In this context, costs become $C(x_i, x_t) \equiv \tau(x_t) + \Im(x_i)$ where the subscript refers to the tangible or intangible choice of investment and the standard convexity assumptions from section

two apply to both τ and $\Im$. Intuitively, this equation captures the idea that any group can independently choose its investment behavior regarding actions x_t, which are open to public scrutiny, and regarding actions x_i, which are obscured from view. In effect, tangible and intangible action choices may be separated as may be decisions regarding equipment purchases and emphasis on data quality respectively.

4.2 Standardized Systems

Occasionally, computer standards can be used to simplify or even to circumvent data sharing problems. If data formats and management methods are standardized, it may be possible to communicate more of the associated collection and maintenance costs. We address the use of standards below.

Define a "standardized" relative to a "non-standardized" database as one which has lower marginal costs with respect to intangible aspects of the system. Formally, letting $\underline{\Im}$ and $\overline{\Im}$ represent the lower intangible costs of the standardized and higher intangible costs of the non-standardized systems respectively, this leads to:

$$\underline{\Im}^i(x_i) < \overline{\Im}^i(x_i)$$

The definition implies that increased standardization between groups causes costs from information asymmetry to grow less quickly. Each group can more reasonably anticipate another's costs of working with the system by virtue of system familiarity. Unfamiliarity and complexity may be alternate interpretations of this phenomenon in which communicating detailed knowledge of a product is difficult but developing knowledge through experience is easier. Standards lower this barrier. By increasing transferable knowledge and the associated level of database certainty,

standards allow hidden, unverifiable, and intangible costs $\Im(x_i)$ to increase more slowly. This enables us to derive the following statements relative to database systems.

Design Principle 4:   Increasing standardization leads to investment nearer the optimum given *shared* databases.

Proof: Given that some tangible portion of the investments remain fully observable, the maximization problem becomes

(6) $$\max \sum_{s|i \in s} p(s)[V(s, \alpha(s)\,|\,x_t, x_i) - V(s\backslash\{i\}, \alpha(s\backslash\{i\})\,|\,x_t, x_i) - \tau(x_t)] - \Im(x_i)$$

Equating marginal costs and marginal benefits; then noting that only tangible costs may be included in an ex post contract gives:

(6a) $$\sum_{s|i \in s} (\partial/\partial x_t)p(s)V(s, \alpha(s)\,|\,x_i, x_t) = \sum_{s|i \in s} (\partial/\partial x_t)p(s)\tau(x_t) \text{ for tangibles and}$$

(6b) $$\sum_{s|i \in s} (\partial/\partial x_i)p(s)V(s, \alpha(s)\,|\,x_i, x_t) = (\partial/\partial x_i)\Im(x_i) \text{ for intangibles.}$$

where, by equation (6a), any given group i has optimal investment in tangible effort x_t while, by equation (6b), the distortion in intangible investment x_i is proportional to $(1 - \sum_{s|i \in s} p(s))\Im^i(x_i)$.   According to the definition of standardization, this distortion is less for a standardized system than for a non-standardized system. By assumption two it follows that $\underline{x}^* - \underline{x}_i \le \overline{x}^* - \overline{x}_i$ and investment in the standardized system is nearer the optimum. Note that for a perfectly standardized system, $\Im^i(x_i)$ approaches zero implying that it has no distortion. The implication for ownership is that standardization increases the potential for efficient decentralized control.

Corollary 4A: Increasing standardization does *not* lead to investment nearer the optimum given *un*shared resources. In this case $\sum_{s|i \in s} p(s) = 1$ and there is no distortion regardless of standardization. The decision to standardize will depend on the expected integration of future assets within a single department.

Interpretation: Design Principle Four suggests that the benefits of standardization are only partially due to the decreased tangible 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 assess costs and workload more accurately. Design Principle Four accurately predicts that the patient records which are shared between clinics and the hospital will be standardized. Individual clinics' billing information, however, which doctors do not share with one another -- and which they may be legally barred from pooling -- need not be standardized. Individual billing systems may differ from doctor to doctor.

The increased efficiency of standardizing resources has led many companies to insist on compatibility as a prelude to large scale knowledge sharing initiatives. The premise supporting this requirement is that superior knowledge will allow corporate managers to make better decisions and that standards ensure the availability of superior knowledge. One of the insights of this research model, however, is that the presence of hidden or intangible costs -- effort levels, reduced political influence, or the opportunity cost of deploying the best staff -- will alter managers' local decisions relative to the global optimum. A common misconception is that good standards and a sound technical design will ensure a successful system. This leads to the following proposition.

> **Corollary 4B**: Standardization does not guarantee optimal data sharing. According to Design Principle 4, so long as intangible costs are non-zero, the investment distortion for any given group is proportional to $(1- \sum_{s|i \in s} p(s))\Im^i(x_i)$.
> Standardization does not necessarily eliminate intangible costs, only reduce them. It follows that realized investment choice x_i, call it maintenance or data gathering, falls below the optimum. Reduced investment results in reduced realizable value such as data availability, accuracy, or recency. Thus, all else being equal, increased standardization cannot be said to necessarily induce optimal data sharing.

Interpretation: The IS literature provides strong support for this observation. Technology solutions alone do not provide the local compensation necessary to

motivate data sharing. One IS consultant points out that, "As information has become the key organizational 'currency,' it has become too valuable for most managers to just give away.", [9 p. 53]. A representative sample of this phenomenon is provided in [23]. In this field study, an international consulting firm is observed installing collaborative work software following the realization by senior management that they were not making effective use of information technology to leverage existing knowledge. The established reward system, however, hinged on client billable hours leading to a disincentive to sacrifice one's own billable hours to support those of another or even to learn the software. According to Orlikowski, "... where there are few incentives or norms for cooperating or sharing expertise, groupware technology alone cannot engender [them]." [23 p. 363]. The industry standard product has no mechanism for compensating employees either for their opportunity costs of learning the system or for the political costs of divulging their private information. These intangible expenses are left completely unreimbursed.

4.3 Outsourcing Systems

One option for improving system-wide cost effectiveness is to outsource system maintenance. The incomplete contracts framework has important implications for asset ownership here as well.[11] The primary reason to outsource is to realize savings from lower cost technologies, lower overhead, or increased economies of scale with declining unit costs. Let the existing cost structure be given by $\tau(x_t) + \Im(x_i)$ while the subcontractor, with whom the group outsources, enjoys lower observable costs for a total of $\underline{\tau}(x_t) + \Im(x_i)$. Lower observable costs produce measurable savings $\tau(x_t) - \underline{\tau}(x_t) > 0$. These definitions form the basis of Design Principle Five.

---

[11]Alternative frameworks for outsourcing are considered in [25] and [1].

Design Principle 5:   Lower cost technologies and reduced overhead are insufficient to justify outsourcing.

Proof: Prior to outsourcing, a group faces net benefit function

(7) $$\sum_{s|i \in s} p(s)[V(s, \alpha(s)|x_t, x_i) - V(s \setminus \{i\}, \alpha(s \setminus \{i\})|x_t, x_i) - \tau(x_t)] - \Im(x_i)$$

Following outsourcing, however, the net benefit function using cheaper technology is

(7a) $$\sum_{s|i \in s} p(s)[V(s, \alpha(s)|x_t, x_i) - V(s \setminus \{i\}, \alpha(s \setminus \{i\})|x_t, x_i) - \underline{\tau}(x_t)]$$

Importantly, the intangible efforts cannot be contracted. Since the contractor does not own the project, and has no ex post bargaining power, he will try to minimize his hidden costs. This he can do easily by setting $\Im(x_i) = 0$ or equivalently $x_i = 0$. It follows that for all value functions such that $V(s, \alpha(s)|0, \varepsilon) > V(s, \alpha(s)|\varepsilon, 0)$ where $\varepsilon > 0$ and $V^i > V^t$ and for all cost functions such that $\Im^i(x_i) < \underline{\tau}^t(x_t)$ the first equation provides greater net benefit than the second. This represents the case where the subcontractor captures none of the incremental value. Using an alternative Nash bargaining solution, the subcontractor retains some bargaining power and receives half the incremental value, yielding a maximization function of:

(7b) $$(1/2)\sum_{s|i \in s} p(s)[V(s, \alpha(s)|x_t, x_i) - V(s \setminus \{i\}, \alpha(s \setminus \{i\})|x_t, x_i) - \underline{\tau}(x_t)] - \Im(x_i)$$

which is closer to but still far from optimal for exactly the same reasons as in earlier proofs. The precise interpretation of these conditions yields a decision rule for retaining project ownership.

Corollary 5A: Organizations should retain ownership of projects in which a majority and increasing share of the benefit derives from intangibles. Conversely, if a contractor has lower cost technology, organizations should outsource projects in which a minority and decreasing share of the benefit derives from intangibles. Precise parameters are governed by benefit function specifics.

Interpretation: In the case of the post office, the central office might consider outsourcing local office functions on the basis of lowering cost. The outsourcing contractor would then need to perform local data gathering and to assume responsibility for functions previously performed by the local office. The central office, however, would be in no better position to enforce quality data gathering than before

since the intangible aspects of this process are not observable. In fact, since the outsourcing contractor does not make use of the data for its own operations, the contractor might be less interested in data quality than the local office. Cost savings alone may not justify outsourcing. Although the details may be open to question, this interpretation confirms the basic premise that ownership is an important incentive as noted in [7, 19, 21].

5 Tradeoffs, Control Problems, and Data Translation as a Solution Alternative

As a matter of practical design, principles may not always agree and designers must balance the most important features of each. Occasions arise when design constraints interact or even contradict one another.   One of the points of the model, however, is that disregarding any design principle carries a cost.   If principles oppose one another then any design choice must bear the costs of the violated design principle. This point is captured in the following proposition.

Design Principle 6: If databases are strictly complementary and more than one agent is indispensable, then the presence of private cost information implies that there is no distribution of database control which induces first best investment[12].

> *Proof*: Given that assets are complementary, Design Principle Two proves underinvestment will result unless assets are concentrated in the hands of a single agent. Given that agents are indispensable, Design Principle Three proves underinvestment unless the assets are controlled by both agents. The contradiction follows immediately.

> *Interpretation*: At the national post office, high level financial accounting functions require a central organizational perspective and universal access. The central office is

---

[12]Hart and Moore [12] prove a stronger version of this statement for firms but use more restrictive assumptions.

indispensable to these resources, implying that the central office should control them. The complementary data, however, are required from the local branches whose indispensability indicates that they should control these resources. Accordingly, there may exist no globally optimal solution.

The potential frequency of conflicting design principles makes it desirable to characterize the tradeoffs between them. This too, however, is difficult.

> **Corollary 6A**: If databases are complementary and more than one agent is indispensable, then the presence of private cost information implies that precedence ordering of design principles is not possible. By construction it is possible to show that the gains from combining complementary assets (equation 5) may be greater than or less than the losses from depriving an indispensable agent of his asset (equation 4). It follows that transferring one of the complementary assets is, a priori, no more desirable than not transferring it.

The absence of an obvious precedence ordering is mitigated by two important considerations. The first is that the need to balance conflicting principles is a fact of normal database design. Capturing the essential conflicts and subtleties of real design dilemmas increases the credibility of any model over others which do not admit to such concerns. The second important point is that data and information are unlike traditional assets insofar as copies are virtually free. Giving data to a second owner does not imply that its original owner must forego its use. This leads us to Design Principle Seven below which describes one method for circumventing the problems introduced by conflicting ownership principles. For most assets, it would be reasonable to follow the design principle which weighed most heavily in inducing investments while compromising the others as little as possible. Without considering the impact of Design Principle Seven, a reasonable heuristic for balancing principles is to consider which ownership structure contributes most to marginal value and to total value. If this is the same structure then it represents the best choice.

As databases possess unique properties regarding duplication, we relax the assumption of indivisibility of assets for Design Principle Seven. Given that data can be copied at a negligible cost, we define a (perfect) "translator" as software which not only copies data from one owner to another but which also translates from the database format native in one group to the database format native in another. A translator may be thought of as a low cost method of providing a duplicate asset. It may be as simple as a disk copy or as complex as a translation between different vendor's formats. In practical terms, it has two essential features, namely, that it provide near on-demand read access -- a short delay can simply be factored into its cost -- and that it not materially interfere with the operation of the database by its owner. The purpose of allowing different formats is to permit each group to manifest its needs and skills in updates to the format that it uses. Let the price of a translator be K. Under the proposed definition, each group invests according to

$$(8) \qquad \sum_{s|i \in s} p(s)[V(s, A \,|\, X) - V(s \backslash \{i\}, A \,|\, X)] - C(x_i) - K$$

This formulation is identical to the one given in equation two save that the ownership map $\alpha(s) \subseteq A$ has been replaced with the entire set of available assets A and the group incurs the expense of the translator. It is as if each department "owned" *all* assets for the price of K. Translators can restore the incentives provided by better access as indicated by the following proposition:

Design Principle 7: The use of a (perfect) translator leads to first best levels of investment.

Proof: The foregoing equation measures total value accruing to a single group. The value to all groups is correspondingly

$$(9) \qquad \sum_{i=1}^{I} [\sum_{s|i \in s} p(s)[V(s, A \,|\, X) - V(s \backslash \{i\}, A \,|\, X)] - C(x_i) - K]$$

for each proper subset $s \subseteq S$ such that $|s| < |S|$, the inner terms cancel and the residual fractions on the grand coalition S sum to one leaving only

(9a)
$$V(S, A \mid X) - \sum_{i=1}^{I} C(x_i) - IK$$

which is the maximum possible value given in the introduction net of the constant term IK. The additional ownership privileges afforded by the translator motivates each group, acting in its own interest, to choose the optimal investment level without direction from a central department.

Interpretation: For each combination of assets, $\alpha(s)$ has given way to A and no group suffers a hold-up problem due to inaccessible assets. A straightforward solution to the problem faced by the post office is for the central office to use an immediate access copy or "translation" of branch office data. This returns local incentives to the branch office, inducing higher data quality and superior maintenance. It also enables the central office to perform necessary cost accounting and general ledger functions which require centralization in order to paint a global picture of fiscal health. A translator performs the function of increasing resource availability. The importance of the preceding definition of a translator is that it gives a theoretical limit as to how much one is worth. More specifically, this value is given by corollary 7A.

Corollary 7A: The value of a translator is

(10)
$$V(S, A \mid X) - \sum_{i=1}^{I} [\sum_{s \mid i \in s} p(s)[V(s, \alpha(s) \mid X') - V(s \setminus \{i\}, \alpha(s \setminus \{i\}) \mid X')]$$

where X represents first best levels of investment and X' represents second best levels. This amount must exceed IK, for purchase or development to be worthwhile.

Developing intra-organizational translators can increase value. Fidelity First is an example of such a system which integrates all customer account information from multiple products across all Fidelity's divisions. It may be possible, however, even to develop *inter*-organizational translators. The Composite Information Systems Tool Kit (CISTK) project at MIT [18] integrates data from such independent sources as Reuters, the MIT alumni database, Dataline, Disclosure, and Finsbury. Queries which require access to different data sets in different native formats can be answered using CISTK.

Within an organization, the Composite Information System (CIS) approach allows for the possibility of the central group accessing the local group's data in a non-intrusive manner. This results in minimal cost and disruption to the local group, which continues to accrue the full benefits of local control, while the central group incurs only the costs of linking the system [27]. 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 decentralized database without necessarily multiplying the costs.

The emergence of multiple views of a database does introduce the potential for divergent growth. We assume that local groups can modify their local copy but not the original and that a translator may be used to reconcile differences between the two at subsequent dates. Database reconciliation is, by itself, an interesting and difficult problem and is an active area of research in the CISTK project. "Context Interchange," which facilitates database transformations and addresses the problem of declaring source data meaning, is especially difficult [27]. What we attempt to provide with Design Principle Seven is a mechanism to measure a translator's added value.

Design Principle Seven responds to problems introduced by the foregoing principles. When assets are complementary and agents are indispensable, for example, it may be possible to configure a system via translation in order that assets behave as if they were combined but indispensable agents do not lose control. The end result is a program of design issues spanning value creation, indispensability, complementarity, standards, outsourcing, and translation as they are affected by ownership. Who owns what is a critical concern in the decision to centralize or decentralize a database system. These issues are summarized in Table 1.

DP (1) **Independence**   If a database system is independent of other parts of an organization, a central authority ought not interfere with its operation, i.e., it should dispense with joint control. Stated differently, an outside authority must add value not merely oversight.

DP (2) **Complements**   Complements in database systems should be combined under centralized control wherever possible.

DP (2, 3) **Indispensibility** The most essential or indispensable department should control a database partition. Given the preceding point, this means that certain departments should absorb responsibility for complementary parts of their systems, for example, a critical end user group may need to perform its own data entry.

DP (4) **Standards**   Standards only increase the efficiency of shared systems. They are irrelevant for standalone systems. Increasing standardization shifts costs more towards observability but good technology and good standards alone cannot create ideal data sharing. System designers must provide incentives to support staff. The higher the fraction of unobserved costs, the more important are these incentives.

DP (5) **Outsourcing**   Lower overhead is not by itself a sufficient reason to outsource. Outsourcing creates greater value only when the observable costs are lower *and* when the unobservable costs are low. The higher the fraction of observable costs, the more beneficial outsourcing becomes.

DP (6) **Design Conflicts** Conflicting design principles create unavoidable costs because no one principle will always dominate. In such cases, a reasonable heuristic is to consider the investment motivations of the group which contributes both the greatest marginal and the greatest total value.

DP (7) **Translators**   When systems require shared data, the use of translators can resolve conflicting design principles and mitigate ownership problems. Ownership encourages groups to invest in their systems. Translation effectively increases resource availability and reduces hold-up problems by making possession appear more widely available.

Table 1 -- Principles of data ownership concern a variety of issues from centralization and decentralization to standards, outsourcing, and translation.

6 Conclusion

The fundamental point of this research is that ownership affects incentives. Any group that provides data to other parts of an organization requires compensation for

being the source of that data. When it is impossible to provide an explicit contract that rewards those who create and maintain data, "ownership" will be the best way to provide incentives. Otherwise, and despite the best available technology, an organization has not chosen its best incentives and the subtle intangible costs of low effort will appear as distorted, missing, or unusable data.

Decentralization concerns equipment and development, but it also concerns intangible issues of ownership and control. Effective ownership is defined as the residual right of control and the motivations of a central IS organization differ substantially from those of a local department depending on who controls the database system. Local 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. These motivations, as well as the technology, affect realized performance. As various ownership structures generate different behavior, only one structure out of many is likely to maximize database value. The seven principles described here support management by helping to choose which structure is best.

Enumerating principles for decentralized database design also begins to make explicit certain basic ideas, which may be implicit in emergent systems or which may only be internalized by the best practitioners. Evidence provided by the post office case, the chemical company of the Markus study [19], the groupware implementation project in the Orlikowski study [23], and the Strategic Data Planning project in the Goodhue, Kirsch, Quillard and Wybo [11] study suggest that decentralization problems are complex and by no means isolated. Several carefully designed systems have run afoul of incentives only to fall short of expectations. Building upon the work of Hart and Moore, the contributions of this research have been to reinterpret intuitive concepts of value creation, independence, complementarity, and indispensability in database terms

and to further elaborate mathematically precise definitions of standardization, outsourcing, and translation. These definitions are then related to a design program for ownership and decentralization and used to interpret several relevant cases.

In all cases where database value can be measured and made explicit, an agreement or contract which details each party's responsibilities and compensation performs as well as redistributing ownership. Much of principal-agent theory is predicated on the ability to achieve optimal effort by measuring results. The intangible nature of information, however, frequently renders measurement of results infeasible and a common theme from information systems literature is that technology assessments alone are insufficient to guarantee system functionality. Key personnel must be given proper incentives to provide support and maintenance. In the absence of an adequate basis for ensuring performance through specific measures, ownership provides an instrument, however blunt, for giving participants the proper incentives.

This model is amenable to other interpretations as well 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 aspects of subcontracting, joint ventures, and facilities management. Since it considers intangible issues, the framework can be used to address qualitative problems of linking companies over and above hardware, software, and bandwidth concerns. We hope to propose new theories in this area in future research.

7 Appendix

The Shapley formula provides an allocation of total value which is proportional to each member's marginal contribution and to his control of the asset pool [14]. These factors represent plausible interpretations of the more intuitive concept of "bargaining power." Given a predetermined investment level x, mathematically this yields:

(11) $$f(i, s, \alpha \,|\, x) = \sum_{s \ni i \in s} p(s)[V(s, \alpha(s) \,|\, x) - V(s \backslash \{i\}, \alpha(s \backslash \{i\}) \,|\, x)]$$

where

(12) $$p(s) = \frac{(|s| - 1)!(I - |s|)!}{I!}$$

which says that a person gets the value created by his group net of the value the group creates without his labors or his assets times the likelihood that he joins that group. The Shapley value balances four properties, namely, (1) treatment of all players is symmetric, (2) non-contributors receive nothing, (3) the division is Pareto efficient, and (4) for multiple games, the expected value of the sum is the sum of the expected values. It also accords well with other efficiency concepts such as Nash equilibrium.

8 Bibliography

1.   Bakos, Y. and E. Brynjolfsson, *From Vendors to Partners: Information Technology and Incomplete Contracts in Buyer-Supplier Relationships.* Journal of Organizational Computing, 1993. **3**(3): p. 301-328.

2.   Barghouti, N.S. and G.E. Kaiser, *Concurrency Control in Advanced Database Applications.* ACM Computing Surveys, 1991. **23**(3): p. 269-317.

3.   Bernstein, P.A. and N. Goodman, *Concurrency Control in Distributed Database Systems.* ACM Computing Surveys, 1981. **13**(2): p. 185-221.

4.   Brooks, F.P., *No Silver Bullet: Essence and Accidents of Software Engineering.* Computer, 1987. **20**(4): p. 10-19.

5.   Brynjolfsson, E., *An Incomplete Contracts Theory of Information, Technology, and Organization.* Management Science, 1994. (Forthcoming).

6.   Ceri, S. and G. Pelagatti, *Distributed Databases: Principles and Systems*. 1984, .

7.   CISR, *Information Technology: Unleashing the Power*1993, MIT Sloan School of Management, Videotape Series: Cambridge, MA.

8.   Date, C.J., *An Introduction to Database Sysems*. Fifth ed. Vol. I. 1990, Reading, MA: Addison-Wesley. 854.

9.   Davenport, T.H., R.G. Eccles, and L. Prusak, *Information Politics.* Sloan Management Review, 1992. : p. 53-65.

10.   Fenton, N.E., *Measuring Internal Product Attributes,* in *Software Metrics: A Rigorous Approach,* F.H.a. Hall, Editor^Editors. 1991, London.

11.   Goodhue, D.L., *et al., Strategic Data Planning: Lessons from the Field.* MIS Quarterly, 1992. (March): p. 11-33.

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

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

14.   Hart, S., *The Shapley Value,* in *The New Palgrave*. 1987, Stockton Press: New York. p. 318-320.

15.   Kawell, L., *et al. Replicated Document Management in a Group Communication System*. in *Second Conference on Computer-Supported Cooperative Work*. 1988. Portland, Oregon:

16.   Kemerer, C.F. and G.L. Sosa, *Systems Development Risks in Strategic Information Systems.* Information & Software Technology, 1991. **33**(3): p. 212-223.

17.   Kreuger, C.W., *Software Reuse.* ACM Computing Surveys, 1992. **24**(2): p. 131-183.

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

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

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

21.   Maxwell, B., *Beyond "Data Validity:" Improving the Quality of HRIS Data.* Personnel, 1989. **66**(4): p. 48-58.

22.   O'Leary, M., *Dialog and the American Chemical Society Play a High Stakes Game.* Online, 1991. **15**(1): p. 15-20.

23.   Orlikowski, W. *Learning from Notes: Organizational Issues in Groupware Implementation*. in *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*. 1992. Toronto, Canada: ACM.

24.   Osborn, C.S., S.E. Madnick, and Y.R. Wang. *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*. 1989. Honolulu:

25.   Richmond, W.B., A. Seidmann, and A.B. Whinston, *Contract Theory and Information Technology Outsourcing.* Decision Support Systems, 1992. **8**(5): p. 459-477.

26.   Rockart, J.F., C.V. Bullen, and J.S. Leventer, *Centralization vs. Decentralization of Information Systems*1977, Center for Information Systems Research, MIT Sloan School: April, (Unpublished Manuscript)

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

28.   Stallings, W., *Local Networks.* ACM Computing Surveys, 1984. **16**(1): p. 3-41.

29.   Tanenbaum, A., *Network Protocols.* ACM Computing Surveys, 1981. **13**(4): p. 453-489.

30.   Wang, R., V. Storey, and C. Firth, *Data Quality Research: A Framework, Survey, and Analysis*1993, MIT Sloan School: TDQM-93-11