December 2004

Understanding Data and Information Sharing in Organizations: A Value-Based Approach

Dennis Hart
Australian National University

Sigi Goode
Australian National University

Follow this and additional works at: http://aisel.aisnet.org/acis2004

Recommended Citation
Hart, Dennis and Goode, Sigi, "Understanding Data and Information Sharing in Organizations: A Value-Based Approach" (2004).
http://aisel.aisnet.org/acis2004/109

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Understanding Data and Information Sharing in Organizations: A Value-Based Approach

Dr Dennis Hart
Sigi Goode
Australian National University
School of Business and Information Management
Australian National University
Canberra, Australian Capital Territory
Email: dennis.hart@anu.edu.au
Email: sigi.goode@anu.edu.au

Abstract

The Information Systems literature observes that, while there are ostensibly benefits to sharing of data and information, barriers to organisational data sharing appear significant. Managers may be understandably concerned that the sharing activity is adversely affecting their own organisations. This paper develops a model of data and information sharing based on the traditional system model, and proposes a theory of the sharing activity in organisations. The paper theorises that employees may engage in or oppose sharing based on the assessment of perceived benefits accruing to themselves from the activity. In particular, the paper highlights the contention that data and information sharing are likely to decrease as organisations grow in size, and also offers an explanation for why the sharing activity is so poorly undertaken in modern organisations even though the technological capability to support it may be there.

Keywords

Data integration, information sharing, value, subjective utility, organizational barriers.

INTRODUCTION

An early expectation of information and communication (ICT) technologies was that they would enable greater sharing of data, information and knowledge in organizations. Such sharing would ostensibly lead to improvements in organizational functioning and performance (e.g. Power 1983; Yu et al. 2001; Li 2002; Roth et al. 2002; Sahin and Robinson 2002). However, expectations on this front have become more circumspect over time in light of evidence that actually achieving sharing is rather more difficult than initially expected (e.g. Davenport 1994). Goodhue et al. (1992) note that “the ability to make coordinated, organization-wide responses to today’s business problems is thwarted by the lack of data integration” and, with respect to information sharing, Davenport et al. (1992) argue that “the rhetoric and technology of information management have far outpaced the ability of people to understand and agree on what information they need and then to share it [so] the information-based organization is largely a fantasy”. Even in a scientific research environment, where information sharing is at least prima facie an explicit norm of the community, barriers to sharing appear substantial (Hubbard and Little 1997).

It has been argued that ownership perceptions and organizational norms influence the sharing of data and information (Jarvenpaa and Staples 2001; Hart 2002) and therefore also affect stakeholder attitudes and political activities related to information systems development (Hart 2002). Hirschheim and Newman (1991) observed, in their investigation of information systems development, that:

“it can often be noted that there is a mystical value attached to the ownership of data. The sharing of data is thus to be avoided. There is a strongly held belief that harm will come from others accessing ‘our’ data, often without any basis in experience”.

This paper focuses on data and information sharing behaviour. Because much research interest currently focuses on organizational knowledge, as opposed to data and information, management and sharing, we feel the wider issue of data and information sharing behaviour has been neglected (e.g. Nelson and Cooprider 1996; Hendricks 1999; Augier et al. 2001; Bartol and Srivastava 2002; Bock and Kim 2002; Jones 2002; Tsai 2002; Swart and Kinnie 2003).

For the purposes of explanation and simplicity, we present our analysis from the point of view of data sharing only and consider later what differences, if any, information as opposed to data sharing might entail. Moreover, since our aim is to understand and capture in a formal model at least certain aspects of the process leading to a
decision of whether or not to share data, we need a general model of decision-making on which to base the subsequent argument. Again in the interests of simplicity, we adopt the traditional rational subjective expected utility (SEU) theory model of decision-making (e.g. Schoemaker 1982) even though we acknowledge that this model of actual human decision-making may have significant deficiencies (e.g. Tversky and Kahneman 1986). It is, however, beyond the scope of our purpose in this paper to address the issue of how more sophisticated theories of human decision-making may affect the analysis we present below.

PRELIMINARIES

Consider a dataset $D_0$, which may be partially or wholly non-computerized. In general there may be multiple datasets $D_i$ in an organization, each of which may be considered to consist of other distinct data subsets, but to simplify the analysis we consider only one such set, $D_0$, from now on.

A dataset such as $D_0$ can be used as an input to a process that operates on it in some way to produce an output. This output is commonly called “information” (e.g. O’Brien, 1993: p.21). Diagrammatically:

![Figure 1: Standard Input (Data) – Process – Output (Information) model](image)

Figure 1: Standard Input (Data) – Process – Output (Information) model

A wide variety of different processes can operate on the dataset $D_0$ in this way. As a result, the situation is in practice more like that shown in Figure 2, which shows the dataset as input to multiple processes $P_0, P_1, P_2, \ldots, P_n$ leading to the various information outputs $I_0, I_1, I_2, \ldots, I_n$.

![Figure 2: Multiple processes operating on the same dataset to produce different information outputs](image)

Figure 2: Multiple processes operating on the same dataset to produce different information outputs

While the various processes operating on the dataset $D_0$ may produce different information outputs, it is also possible that they produce outputs that are incompatible with or contradictory to each other. Moreover, the set of processes $P_0, P_1, P_2, \ldots, P_n$ may, in the most general case, be extended to include everything that could conceivably be done with the dataset $D_0$, including things that would be considered legitimate, as well as those that might be considered illegitimate.

The different processes $P_0, P_1, P_2 \ldots P_n$ may be associated with various organizational actors. Process $P_i$ may actually be operated by organizational actor $A_i$. On the other hand the connection between a process and organizational actor may only be potential. For example, process $P_i$ may not exist in reality, but if it did then the perception might be that a certain actor $A_i$ would be the likely one to run it. In general, a single organizational actor may be linked to multiple processes. However, there is no reason why different actors may independently actually undertake the same processing on the same dataset to produce the same information output.

The link between processes and actors is, in general, many-to-many. We illustrate this by further expanding Figure 2 to produce Figure 3, which is re-oriented for clarity and in which $A_0, A_1, A_2 \ldots A_n$ are organizational actors and the dashed connections indicate their associations with the various processes $P$. A dashed link between an actor $A_i$ and a process $P_j$ indicates that $A_i$ either could run $P_j$ on dataset to produce $D_0$ to produce output $I_j$. 

![Figure 3: Illustration of many-to-many relationship between processes and actors](image)
Figure 3: Multiple processes operating on a dataset to produce different information outputs, also showing the
organizational actors associated with those processes.

**Value of process outputs and information**

An organizational actor $A_i$ can be expected to attach some value to actual and potential information outputs,
regardless of whether the process is associated with themselves or another actor $A_j$. We may also expect that the
value placed on a particular information output by one actor will, in general, differ from that placed on the same
output by a different actor. This inherent value may be positive, negative or zero.

**Value of data sharing**

The actual value (the subjective expected utility in SEU theory) an actor $A_i$ attaches to some information output
$I_k$ resulting from process $P_k$ that is associated with actor $A_j$ will depend on their subjectively estimated
probability of that output really being produced. As an example, even though a potential information output $I_i$
might be extremely valuable to some actor, if the probability of its actually being produced is estimated by that
actor to be zero then its actual value to them would be zero. Thus, the value of an output must be weighted by
its probability of production in order to see its actual value.

The perceived probability of production of a certain information output is, however, actor-dependent. That is,
one actor’s estimation of the probability of the production of a certain output may be different from that of
another actor. We may, therefore, represent the probability estimated by actor $A_i$ of an actor $A_j$ operating some
process $P_k$ to produce an information output $I_k$ as $p_{ijk}$. Note that the probabilities $p_{ijk}$ are considered to be
independent of one another. That is, the probability of one actor running a particular process to produce a certain
output is assumed to be unrelated to the probability of occurrence of any other process.

Now consider actor $A_0$ who controls dataset $D_0$. What would be the value to $A_0$ of sharing $D_0$ with all other
actors? This is given by:

$$V_0 = \sum_{j \neq 0} \sum_{k=0..n} p_{ijk} v_{ijk}$$

That is, the value to $A_0$ is the sum, in accordance with SEU theory, across all actors (other than themselves) and
all processes of the actual value to them of the information outputs that those actors and processes will produce
from $D_0$. If this turns out to be positive (but not zero) then a rational value-maximizing actor $A_0$ may be
motivated to share the dataset $D_0$ with all other actors.

In a similar way, the value to some other actor $A_i$ (not $A_0$) of $A_0$ sharing dataset $D_0$ with all other actors is:

$$V_i = \sum_{j \neq 0} \sum_{k=0..n} p_{ijk} v_{ijk}$$

If $V_i$ ($i \neq 0$) is positive then we may expect that the relevant actor ($A_i$) will be motivated to push for $A_0$ to share
$D_0$ with all other actors.
Restricted sharing

The calculations above relate to sharing across all actors. However, there are other possible cases to consider which involve more restricted sharing behaviour. Consider the value to actor \( A_i \) of \( A_0 \) sharing dataset \( D_0 \) with \( A_i \) alone. This is:

\[
V_{ii} = \sum_{k=0..n} p_{ik} V_{ik}
\]

If \( V_{ii} \) is positive but \( V_i \) is negative then actor \( A_i \) will be motivated to get \( A_0 \) to share \( D_0 \) with them, but also for them to block \( A_0 \) sharing \( D_0 \) with any other actor. More generally, the value to \( A_i \) (not \( A_0 \)) of \( A_0 \) sharing dataset \( D_0 \) with a third actor \( A_0 \), is:

\[
V_{ij} = \sum_{k=0..n} p_{jk} V_{jk}
\]

If this is positive, then we may expect that actor \( A_i \) will be motivated to get \( A_0 \) to share \( D_0 \) with the third actor \( A_j \).

It remains to consider the intermediate case of sharing with some proper subset \( A \subset \{ A_0, A_1, \ldots, A_n \} \) of all other actors. Suppose that \( \alpha \) is the set of index values of the members of \( A \). Then, the value to an actor \( A_i \) of \( A_0 \) sharing dataset \( D_0 \) with the members of \( A \) is:

\[
V_{i\alpha} = \sum_{j=\alpha} \sum_{k=0..n} p_{jk} V_{jk} = \sum_{j=\alpha} V_{ij}
\]

and, if this is positive, \( A_i \) will be motivated to get \( A_0 \) to share with the members of \( A \). However, it is important to note that we may expect the value of \( V_{i\alpha} \) to vary if the membership of \( A \) is changed. This being so, if there is a subset \( A \) for which \( V_{i\alpha} \) is positive and maximal, then \( A_i \) will be motivated to share with this specific subset of actors in preference to any other subset (or all) actors. If there is no subset \( A \) for which \( V_{i\alpha} \) is non-negative then the actor \( A_i \) will not be motivated to push for the actor \( A_0 \) to share \( D_0 \) with any other actor and, if it is significantly negative, they can be expected to actively oppose such sharing because they would deem it to be detrimental to their interests.

ORGANIZATIONAL DATA AND INFORMATION SHARING

In the light of the analysis above, we now explore the circumstances under which it can be expected that all organizational actors concur in the sharing of dataset \( D_0 \) with all other actors. On the assumption that the actors are rational value maximizing entities, this would entail that, for every actor \( A_i \), \( V_i \) be not only non-negative, but also maximal. On the less restrictive assumption that the actors are merely satisficing entities (Simon 1957), it would still entail that, for every actor, \( V_i \) be non-negative although not necessarily maximal. Either way, it would seem that as organizations increase in number of actors, the likelihood of such conditions holding would progressively decrease and, consequently, also the likelihood of unproblematic organization-wide data sharing (an expectation that seems consistent with literature evidence).

This model can also explain an organizational actor’s willingness to share processed outputs, or what we have here termed “information”, but at the same time for them to be reluctant to share the input data that was used to produce that information. As an example, in her classic study of the Financial Information System (FIS) implementation, Markus (1983) says:

“Prior to FIS, divisional accountants summarized raw data on the transactions in their divisions and sent the summaries to the corporate accountants for consolidation. Divisions retained control of their own data and exercised substantial discretion in summarizing it...After FIS, however, all transactions were collected into a single database under the control of corporate accountants [so] FIS automatically performed the divisional summaries [and also] corporate accounts had the ability to “look into” the database and analyse divisional performance”

The explanation is as follows. If the actor \( A_0 \) provides a certain information output \( I_0 \) directly to another actor \( A_1 \), rather than the dataset \( D_0 \) from which it was drawn, then \( A_0 \) has successfully constrained the information received by \( A_1 \) to \( I_0 \) only. On the other hand, if \( A_0 \) were to release the dataset \( D_0 \) to \( A_1 \), then \( A_1 \) could at least potentially generate not only \( I_0 \) but also any number of other potential outputs \( I_0 \) that will be of varying value to the original actor \( A_0 \) in line with our analysis above. Moreover, the receiving actor \( A_1 \) may signal that they will only generate output \( I_0 \) for themselves but in fact generate a different output \( I_0 \) that is significantly different from the original \( I_0 \) that \( A_0 \) would have provided. In any case, it may well be that the actor \( A_0 \), in assessing the value to them of \( A_1 \) having \( I_0 \) is greater than the value to them of \( A_1 \) generating \( I_0 \) for themselves together with the risk that they will also generate some or all of the other possible outputs \( I_0 \) in addition.
In terms of the discussion above, actor \( A_0 \) here represents the divisional accountants, the corporate accountants are represented by actor \( A_1 \), the divisionally provided transactions summaries are \( I_0 \), the corporately generated divisional transaction summaries are \( I_1 \), and the analyses of divisional performance based on the transactional data are all the other potential outputs \( I_i \). Markus’ description of the FIS implementation and more particularly the effects it elicited regarding data and information sharing are evidently readily understandable and explainable in terms of our value-based analysis.

LIMITATIONS AND SIMPLIFYING ASSUMPTIONS

A number of simplifying assumptions have been made in the foregoing analysis. Among these are:

Non-informational results of sharing and other pressures

Non-information based outputs and other pressures, have not been considered. For example, an organization may implement various incentives in order to encourage information sharing. These may affect the attitudes and value calculations of the various organizational actors, and consequently their decisions about whether and how to share. In order to incorporate these aspects our analysis could be extended along the lines of Ajzen and Fishbein’s “Theory of Reasoned Action” or TRA (Fishbein and Ajzen 1975, Ajzen and Fishbein 1980) in which a behaviour is deemed the combined result of a behavioural intention, attitudes and social norms as follows (after Upmeyer 1989):

\[
B = w_1 BI + (A_B) w_2 + (SN) w_3
\]

where

- \( B \) = overt behaviour
- \( BI \) = behaviour intention
- \( w_1 \) = empirical weight attached to \( BI \)
- \( A_B \) = attitude toward behaviour \( B \)
- \( w_2 \) = empirical weight attached to \( A_B \)
- \( SN \) = subjective norms
- \( w_3 \) = empirical weight attached to \( SN \)

and

\[
A_B = \sum B_i E_i
\]

where

- \( B_i \) = belief that behaviour will lead to outcome \( i \)
- \( E_i \) = evaluation of expected outcome \( i \)

and a similar sum is made over the various sources of social norms to whom the actor is subject, and the actor’s motivation to act in accordance with those norms. Our analysis is limited to the attitudinal component in TRA, omitting the distinction between a behavioural intent, the behaviour itself and the effect of social norms or other pressures that may exist.

Definition of the dataset to be shared

It has been assumed that all actors share a common understanding of what comprises the dataset \( D_0 \) and that they base their subjective value calculations on that understanding. However, even though the different actors may believe they have such a common understanding about what \( D_0 \) is, they may not have it in fact. Thus, even if all the actors concur that sharing \( D_0 \) is desirable as a result of their various value calculations, it may then emerge that the “\( D_0 \)”, on which their considerations were based, was different, possibly invalidating their conclusions regarding sharing it. See, for example, our discussion of the FIS described by Markus (1983) above in relation to the outputs \( I_0 \) and \( I_1 \).

Commonality of the process and output sets across actors

We have assumed that that each actor has perfect and identical knowledge of what comprises the set of processes \( P = \{P_0, P_1, \ldots P_n\} \) and their corresponding information outputs \( I = \{I_0, I_1, \ldots I_n\} \). We have also
assumed that each actor perceives the same set of organizational actors and their connection to the processes P in their environment as every other actor. However, none of these assumptions are likely to be correct, at least in general. Different actors may be expected to know about and visualize different sets of processes P and their information outputs I. Therefore their subjective value calculations will vary not only because of differences of perception regarding the value of the various information outputs, but also because of differences of perception regarding what those outputs will in fact be in the first place. Moreover, because different actors may operate with different conceptions of who the actors other than themselves are, as well as the linkage of these actors to the information producing processes P, their value based motivations for data sharing (if any) may then be distributed across a different set of actor entities than is the case for others doing similar data sharing value assessments. While this would complicate the analysis in a particular case, it does not affect the analysis already presented.

**On-sharing**

It is possible that, once a dataset has been shared with an actor $A_i$ (by, say, $A_0$) then the receiving actor’s value calculations may differ so significantly from $A_0$’s that it results in them passing on the dataset that originated from $A_0$ to further actors that $A_0$ would not have shared with on the basis of $A_0$’s own value calculations. It may be, however, that if $A_0$ knows about and assesses this risk, it will have its effect on their original subjective probability and value estimations regarding sharing with $A_i$. If this is indeed the case, then the model can be regarded as taking this possibility into account without need of modification.

**Context**

The number of actual and possible information producing processes P applicable to any particular dataset of significant size or complexity will most likely be large. Moreover, it is probable that even if different actors think they are talking about the same information producing process and output, they are not. The reason lies in the presence of contextual factors and interaction effects with other datasets, as illustrated in Figure 4.

![Figure 4: Complicating effect of contextual factors and interaction effects.](image)

Even if the input dataset ($D_0$) is the same and the actors think that they are running the same process $P_i$ on it, the output information is unlikely to be the same because of differences in other inputs and context, even if the actors themselves do not recognise them as being present, relevant or different. Shared data and information are subject to interpretation and cannot be assumed to hold the same meaning for those who have it (Miranda and Saunders 2003). It is therefore likely that the majority of processes shown in Figure 3 are actually unique to a single actor. In sharing a dataset with other actors, the originating actor should therefore expect that the receiving actor may well derive different information outcomes from the shared dataset than what they (the originating actor) might expect, even if they know in considerable detail what the receiving actor intends doing with it.

**Network, feedback and learning effects**

The analysis has assumed that data resident in some dataset is input to information producing processes P, and that the output is produced from that input in a single stage process. However, the output of one process $P_i$ may itself form the input for a different process $P_j$, and so on. The reality, then, is rather a complex network of interlinked information producing processes, many of which operate upon or are affected by, in a contextual or interactive sense, the outputs of other information producing processes.

Feedback effects may also exist to complicate matters even further. In subsequent time periods, actor $A_0$ may have an improved ability to assess the value to themselves of sharing particular data sets. Similarly, other actors in the organisation may have a different understanding of $A_0$’s motivation to share data. As these actors learn, the sharing dynamic may change commensurately.
FACTORS AFFECTING SHARING

In order for sharing to occur, the actor that is deciding whether or not to share a dataset D needs to have a set of possible processes \( P_1, P_2, \ldots, P_n \), their corresponding information outputs \( I_1, I_2, \ldots, I_m \), and a set of estimated probabilities \( \{ p_{ijk} \} \) in mind on which to base their deliberations. Factors affecting this are as follows.

Awareness of Other Organizational Actors and the Effect of Ignorance

While it may be possible in principle to develop a list of all possible processes one could carry out on a given dataset, this would be infeasible in reality. Instead, the set of processes \( P_1, P_2, \ldots, P_n \) would more realistically be constructed from the sharing actor’s awareness of the other organizational actors and what they might conceivably do with the data if they were given the opportunity. In a similar way, the subjective estimation of the probability of each actor actually running a particular process on the dataset would depend on the knowledge and understanding the sharing actor had of the other party with whom they were considering sharing.

It is interesting, in this context, to consider the effect of ignorance of the originating actor regarding the other party with whom they may be contemplating sharing. As Tversky and Fox (1995) note, “People typically do not know the exact probabilities associated with the relevant outcomes, but they have some vague notion about their likelihood”. Tversky and Fox find that the decision-making impact of a possible outcome moving from possibility to impossibility, or vice versa, is significantly greater than if it just becomes more or less probable.

Now, with respect to data or information sharing, ignorance of the other actor amounts to the admission of possibility to impossibility, or vice versa, is significantly greater than if it just becomes more or less probable. Moreover, under conditions of uncertainty, it is known that lower probabilities are overestimated compared to intermediate to high ones (Fox and Tversky 1998): if something is deemed to be highly likely to be negative. More importantly, under conditions of ignorance of the other actor, overestimated value, this may have a disproportionate impact on the decision to share compared to a simple change from one non-zero probability to another.

Trust and Social Interaction

Trust, or the willingness to allow oneself to be vulnerable to the actions of another over whose behaviour one has no control, has a positive effect on information sharing (e.g. Zand 1972, Butler 1999). Shapiro et al. (1992) identify three sources of trust: deterrence based, knowledge based, and identification based. Knowledge-based trust is founded in the confidence the trusting actor has that they know what the other party will do, and that the trusted party will act benevolently with respect to their interests. In terms of our analysis, this amounts to the sharing actor knowing (or thinking they know) what the receiving actor will do with the dataset they are contemplating sharing (i.e. knows the processes they probably will or will not run), and that their value calculation based on that knowledge for sharing is positive. Accordingly, the sharing actor will be inclined to trust and therefore share the dataset with the other party, as we have argued above. Alternatively, if the sharing actor does not know what the receiving actor will do but nevertheless identifies with them as being similar or empathic to themselves, then they can safely make assumptions about what the receiving actor may do with the shared dataset.

Social interaction is known to have a “significant positive effect” on knowledge sharing (Tsai 2002) and, presumably, on data and information sharing also (Phillips et al. 2004). Moreover, explicit reward systems appear to be less effective at encouraging sharing behaviour (e.g. Bartol and Srivastava 2002, Bock and Kim 2002). However, organizational actors that engage in social interaction learn, through that interaction, more about each other as it proceeds. With respect to potential data, information and knowledge sharing, this learning enables each actor to assess more and more accurately how the other is likely to behave in the event of sharing taking place. This suggests that knowledge of the other party through social interaction can encourage sharing, but we would argue that instead of there being a direct linkage between such interaction and sharing, this rather occurs through an intervening value calculation step based on increased knowledge of, or identification based assumptions regarding, what the other party is likely to do with the shared dataset.
CONCLUSIONS

A primary implication of the analysis presented here is that in organizational situations where there are multiple datasets, many actual or potential information producing processes that do or could act on those datasets, and many organizational actors who may operate those processes, the unhindered sharing and integration of data across the organization is very likely to be difficult if not impossible to achieve. The process of trying to achieve such sharing or integration is likely to be fraught with difficulty and associated organizational strife. While this conclusion is no doubt unsurprising since this has been the experience of many organizations over many years, despite the undoubted existence of information technology capabilities capable of delivering the data and information sharing aimed at, our intention in this paper has been to try to establish a firmer and more formal theoretical understanding of at least some of the reasons why this might be so.

Most business applications of information technology now entail data and information sharing not only internally but increasingly externally to the organization. It is, therefore, more important than ever to better understand the motivations that encourage as well as hinder such sharing. The motivations for data and information sharing are well discussed and documented in the relevant literature. There are any number of articles touting the benefits waiting to be reaped from increased intra-organizational and inter-organizational sharing of data and information. Nevertheless, despite these undoubted benefits the sharing involved is often very difficult to achieve.

This paper has provided a value-based analysis of data and information sharing that throws some light on why organizational actors may view information systems and other technologies that are targeted at achieving increased data and information sharing with misgivings and why the data and information sharing that they can undoubtedly provide in a technological sense is often not achieved in practice.

REFERENCES


COPYRIGHT

Dennis Hart and Sigi Goode © 2004. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.