Abstract

The literature observes that, while there are ostensibly benefits to sharing of data, 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 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 sharing decreases as organisations grow, and also offers an explanation for why the sharing activity is so poorly undertaken in modern organisations.

Keywords: Data Sharing, Value Analysis

1. 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 actual sharing is rather more difficult than initially expected (e.g. Davenport 1994). Moreover, 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, as another example, 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”.

We present a value-based analysis of data sharing with the aim of shedding increased theoretical light on why organizational data sharing is as difficult as it evidently is. In order to do so, we adopt the traditional rational subjective expected utility (SEU) theory model of decision-making (e.g. Schoemaker 1982), while acknowledging that this model may have significant deficiencies (e.g. Tversky and Kahneman 1986).

2. A Value-Based Analysis of Data Sharing

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) and we will adopt this terminology. Diagrammatically:

![Diagram 1](image1)

Figure 1

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 $P_0$, $P_1$, $P_2$, …, $P_n$ leading to the various information outputs $I_0$, $I_1$, $I_2$, …, $I_n$.

![Diagram 2](image2)

Figure 2

The various processes operating on the dataset $D_0$ may produce different information outputs, but it is also possible that they may produce outputs that are incompatible with or contradictory to each other. Moreover, the set of processes $P_0$, $P_1$, $P_2$, …, $P_n$ 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 illegitimate or wrong, as well as those that might be considered legitimate and correct.

The different processes $P_0$, $P_1$, $P_2$ … $P_n$ may be associated with various organizational actors. Process $P_i$ may actually be operated by organizational actor $A_j$. On the other hand the connection between a process and organizational actor may only be potential. In general, a single organizational actor may be linked to multiple processes. However, there is also no reason why different actors may independently actually undertake the same processing on the same dataset to produce the same information output.

![Diagram 3](image3)

Figure 3
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 $A_0$, $A_1$, $A_2$ … $A_m$ 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 does or could run $P_j$ on dataset to produce $D_0$ to produce output $I_j$.

2.1 The 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 value may be positive, negative or zero.

2.2 The 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 $I_k$ 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:

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

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 $A_0$ can be expected to 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.
2.3 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:

$$V_{ii} = \sum_{k=0..n} p_{iik} V_{ik}$$

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 $A_j$, is:

$$V_{ij} = \sum_{k=0..n} p_{ijk} V_{ijk}$$

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 $A \subset \{A_0, A_1, \ldots, A_m\}$ 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:

$$V_{i\alpha} = \sum_{j \in \alpha} \sum_{k=0..n} p_{ijk} V_{ijk} = \sum_{j \in \alpha} V_{ij}$$

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.

3. 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, coincidentally, also the likelihood of unproblematic organization-wide data sharing (an expectation that seems consistent with literature evidence).

4. Limitations and Simplifying Assumptions

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

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. This may not, however, be the case.

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. But while this would complicate the analysis in a particular case, it does not affect the analysis already presented.

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.

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 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.

5. 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
conflict. This conclusion is 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 sharing aimed at.

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 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 sharing that throws some light on why organizational actors may view information systems and other technologies that are targeted at achieving increased data sharing and, by extension, also 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.

6. References


Simon, H.A. Models of Man, 1957, John Wiley & Sons
