Data Ownership and Semiotics in Organizations, or Why “They’re Not Getting Their Hands on My Data!”

Dennis Hart
School of Computer Science, UNSW
Australian Defence Force Academy

Abstract
A longstanding aim of information system managers and developers in organizations has been to improve the availability and use of data through its rationalization and sharing across all who need it. This has proven difficult to achieve, not least because of the unwillingness of data “owners” to share what they consider to be theirs. This paper seeks to better understand the nature and origin of these ownership issues by tackling them through the lens of Semiotics (the theory of signs). The discussion is then illustrated through a brief case study drawn from the information systems literature. Moreover, it is argued that, contrary to the usual view, there may in fact be good reasons for the existence of such data ownership perceptions, not only from the “owner’s” viewpoint but perhaps even from that of the organization as a whole.

Keywords: Data, ownership, Semiotics, Organizations

1. Introduction

That information management and information systems development in modern organizations continues to be a significant challenge can hardly be disputed. Moreover, why they are so is more often for human and organizational rather than technological reasons (Avison and Fitzgerald, 1995). Indeed, Keil, Cule, Lytinen and Schmidt (1998) have found that, of the top 11 risk factors identified by three independent panels of experienced software project managers from across the world, only one involved technology.

One non-technological problem arises because of conflict over the ownership of organizational data and/or systems in which it is stored. For example, Wigan (1992) says “The perception that certain types of data assembled or created by one [organizational] group should not be accessed without agreement by the collectors is a major issue”. Moreover, the resulting behaviour – unwillingness to share information – “is probably the single biggest reason for the lack of impact of information resources management on organizations” (Oppenheim, 1997, italics in original).

While it has been argued that “corporate data is data without which an organization cannot function [and which] is ‘owned’ by the organization” (Holloway, 1988), it is still true that one must deal with “diverse business units and work groups… many of whom are uncomfortable with any attempt to dictate the use of “their” data” (O’Brien, 1993). As a result “the rhetoric and technology of information management have far out-paced 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” (Davenport et al, 1992).

---

1 Dr Dennis Hart is a Lecturer at the School of Computer Science and can be contacted via email at: d.hart@adfa.edu.au. The author gratefully acknowledges the detailed comments provided by the reviewers in assisting to improve the quality of this paper.
Where do these perceptions of data ownership and the consequent reticence to share come from? Reflecting the general recognition that possession or control of data and/or information can be a significant source of power (Zand, 1981), they often seem to be ascribed to power-based motives. It is then but a short step to regard them as dysfunctional, for the organization at large if not the perpetrator, and therefore deserving of stamping out or, at the very least, severe discouragement. However, against this it is argued in this paper that power-based motives are not the only source of perceptions of data ownership. Moreover, there can in fact be good reasons for an unwillingness to share data, at least from the point of view of the owner but possibly also from an overall organizational perspective.

2. Semiotics

Concepts from linguistics and sociology have been applied in the field of information systems, particularly in relation to approaches to systems development (e.g. Janson et al, 1993, 1995). However, with some notable exceptions (e.g. Mingers, 1995; Stamper, 1973, 1992), semiotics or the theory of signs seems to have had relatively little impact in the field. Even so, insofar as it has been used, it has been directed primarily towards the area of information systems development. This paper seeks to apply concepts drawn from semiotics to assist in understanding organizational issues related to data ownership but, before doing so, a brief overview of some relevant concepts is needed.

2.1 Basic Concepts

Essentially, semiotics is concerned with understanding how signs and symbols, including spoken and written language, but also other forms of communication such as pictures, gestures, films, etc, are used and come to have the meanings for us that they do (Sless, 1986). It is generally accepted that the scope of semiotics contains several important subfields including syntactics, semantics and pragmatics. Respectively, these are concerned with:

Syntactics: the “relation between a given sign vehicle (i.e. the physical manifestation of the sign) and other sign vehicles” (Nöth, 1990); that is, the rules that specify how the various signs may be combined into correct patterns such as sentences in a written language;

Semantics: the “relations between sign vehicles and their designata (i.e. the thing to which the sign refers)” (Nöth, 1990); that is, what the sign means; and


Stamper (Stamper, 1992) has organized these subfields, plus three additions of his own, into a “semiological ladder” as shown in Figure 1. In this Figure the various characteristics associated with a sign are encapsulated at the different levels. To illustrate the different levels, we consider first a non-linguistic and then a linguistic example.
2.2 A Non Linguistic Example

For the non-linguistic example we will use a well-known sign – the swastika. At the physical semiotic level the swastika may be carved in stone, scratched into the sand on a beach, drawn on a piece of paper, daubed on a wall, engraved on a medal, and so on. At the empiric level, the question of how recognizable it is arises. For example, is it partially obscured but still visibly a swastika? Or is it distorted, though not enough to prevent its recognition? At the syntactic level, the concern is with usage of the symbol. Would it make sense to insert it into the middle of a sentence such as this one? No, not really. Would it be sensible to paint it onto a flag, or find it engraved in stone at an archaeological dig, or adorning the cover of a book? Yes, very likely. Now, at the semantic level, what does the swastika mean? According to the Encyclopædia Britannica, before the mid-twentieth century its meanings have variously been:

- Prosperity, abundance and good fortune;
- The god Thor’s hammer;
- Religious (as an icon of the Hindus, Jainas and Buddhists and as a version of the Christian cross);
- The Sun;
- Night, magic and the Indian goddess Kali.

Now, however, the meaning it virtually universally evokes is as the Nazi emblem. To be fair, however, these various meanings are spread across several variations of the symbol (mirror images and rotations) each of which may be argued to be a separate symbol in its own right. But nevertheless, even if this is admitted, differences in meaning still remain.

At the pragmatic level, the concern is with what use of the symbol communicates about the intentions of its user. For example, there can be little doubt that an author who uses the swastika on the cover of a novel he or she is writing is likely to have different intentions through its use than someone who paints it on a flag to wave at a street march.

Lastly, at the social level, the interest is in what the symbol communicates about the social beliefs, values, commitments and so on of its user. For example, nowadays a person waving a flag on which the swastika appears immediately identifies him/herself with the Nazi belief system and may reasonably be assumed to support the kinds of ideas and actions that that
belief system entails. Such a conclusion cannot, of course, necessarily be drawn for the author who uses the swastika on the cover of a novel that he or she has written.

2.3 A Linguistic Example

Now we consider a linguistic version of the transition from one level to another in Stamper’s semiological ladder, but this time restricting consideration to the syntactic and semantic. Why restrict attention to these two levels? Inspection of Figure 1 reveals that, in addition to showing the various semiotic levels, Stamper has separated the diagram between the syntactic and semantic with a dividing line labelled “The IT Platform”. This indicates that our technological information systems are (arguably) capable only of capturing and storing the characteristics of signs at the syntactic and lower levels. The higher levels of semiotic content are, at least for the present and perhaps foreseeable future, the domain of human additions and interpretations.

It may perhaps be contended that the division in Figure 1 should be drawn somewhat higher in the diagram. For example, so-called knowledge-based and expert systems exist and these might be argued to contain higher-level semiotic content than the syntactic. Nevertheless, it is undeniable that such a line exists somewhere near the middle of the diagram. Therefore, for the sake of argument, it will be assumed that the division is appropriate where Stamper has placed it.

Now to the linguistic example: consider the following two pieces of data:

Transfer amount = $1000
Transfer date = 23 October 1999.

Someone inspecting these directly may combine them, using the syntactic rules of language (i.e. grammar), into a sentence such as:

“The transfer amount was $1000 and the date of the transfer was 23 October 1999”.

On the other hand, someone else, given the same data and grammar might construct:

“The transfer amount was $1000 but the date of the transfer was 23 October 1999”.

These two very similar sentences, both syntactically correct and based upon the same data, transmit very different meanings to a listener. The second says there is something special, suspicious or otherwise out of the ordinary about what is being reported whereas this aspect is totally absent from the first. This semantic difference occurs despite the similarity of the syntactic structure of the two sentences and the fact that their truth-values, in the logical sense, are necessarily the same. So, what is the point? It is that even small, subtle changes in the way the same data is combined using syntactical rules to produce correct outputs can have a major impact on the meaning or semantics conveyed to the recipient. This effect is not limited to language-based communication either. As an illustration, there are well known techniques such as baseline and unit choice, or method of graphical presentation, for conveying numerical data in a way designed to carry the desired semantics. For example, the jump from $1 to $3 may be either an increase of $2 or 200% depending on whether one wants to paint it as small or large.
3. Semiotics versus the Traditional View

Several of the different levels in Stamper’s semiological ladder correspond, approximately, to terms in more common use. For example, Ramaprasad and Rai (1996) compare syntactics with data, semantics with information, and pragmatics with knowledge but this linkage is not exact. Liang (1996) also speaks of an ascending hierarchy of “basic entities”, namely Data, Information, Knowledge and Wisdom, in his theoretical model of information processing, decision making and information systems. These may be roughly identified with the Syntactic, Semantic, Pragmatic and Social World levels of Stamper’s semiological ladder. There is, however, an important distinction in the way these are regarded. For Liang, the transformation from one entity to another (e.g. from data to information) brings “better orderliness to the entity [which] reduces the entropy of the entity so that further analysis can be executed more systematically” (Liang, 1996). This is very much in line with the conventional textbook view of the relationship between data and information where, for example, “information is data that have been organized so that they have meaning and value to the recipient” (Turban et al, 1999). Implicit in this view is the assumption that nothing is contained in the higher level entity that is not already present in the lower level one from which it is derived. However, this is not so in the semiotic version. In this case, the higher the level of a sign (entity) in the semiological ladder, the more meaning it has accreted through its embedding in and interaction with an ever richer surrounding environment.

4. Semiotics and Data Ownership

We now propose that having the ability to control the syntactic construction (e.g. a sentence) in which data is embedded and communicated to a recipient is a powerful motivator towards control and ownership of organizational data.

The reasoning behind this statement is as follows. If an organizational actor, say A, has direct access to data then they, by choosing the syntactic construction in which it is embedded when communicated, heavily influence the semantics or meaning that the message carries. However, if another actor B lacks direct access to the data then B is subject to the semantic content of the message communicated by A. While it is important to note that B may actually receive a different meaning than that intended by A, it is nevertheless A’s context, agenda, prejudices, knowledge of exceptions, unusual circumstances and so on that determine the syntactic construction and its attendant semantic content, and these are very likely different from those of B. From A’s point of view, if there is significant risk of B creating a syntactical construct from the data that carries the wrong semantics, then there is a strong case for them to retain or obtain control of access to, or ownership of that data. Of course B may think the reverse on similar grounds, but this will probably be of little concern to A.

An example might be a situation where A is a sales manager and B a senior manager. When asked by B how sales progress is going for the month, A replies that they are within the expected range. In making his reply, A knows that the raw sales figures are down compared to previous months but this is because a significant number of staff have been away for a week attending a training seminar. If, on the other hand, B had direct access to the raw sales data then she might well have deduced that something was amiss because the figures were noticeably down by comparison with previous months.
4.1 Semiotic Networks

The situation described in the previous paragraphs may be shown as in Figure 2, which we term a Semiotic Network (SN).

![Figure 2](image)

The message of Figure 2 is not that actor B completely lacks all access to the data holding. Rather, the access they have is mediated by the interpretations of A through their syntactic constructions and the semantic content carried through them. As in the sales manager example, B may need information based on the data but, in order to obtain it, must ask A. Actor A may then be both able and willing to provide a considered answer in the form of a simple statement, a short summary, a lengthy report or whatever, even though still being unwilling to grant B direct access to the data itself.

![Figure 3](image)

Now consider the situation shown in Figure 3, where both actors A and B have direct access to the data holding. In this Figure, since both A and B have direct access to the data holding each can apply their own interpretation to generate the appropriate (as they separately see it) syntactical constructions and associated semantics. However, these syntactical outputs may differ significantly in terms of their semantic content, which may perhaps be in direct conflict as shown by the crossed swords symbol in the diagram. Such would, at least potentially, be the situation if the senior manager in the sales example had direct access to the raw sales data rather than having to ask for it through the sales manager.

In connection with Figure 3 the sales example is simple, relatively trivial and easily corrected or overcome by the provision of extra data. However, this is not necessarily always true. As a non-information systems related example, consider the case where A and B have substantially different worldviews, or Weltanschauung as Checkland and Scholes (Checkland and Scholes, 1990) would say in the context of SSM. These different worldviews might be opposing political stances for example. In such a case, the same data given to A and B is perhaps more than likely to be expressed by quite different syntactic outputs, and invested with entirely different meaning by each of them. Moreover, these meanings may remain entirely irreconcilable even in the face of whatever further data on the subject may be supplied to either or both of them.
5. Data Definition

The Semiotic Networks of Figures 2 and 3 show only data being accessed (as well as interpreted). However, data is not only accessed but also needs to be defined. In an information systems sense, the syntactic aspect of data definition entails its formal structure; firstly whether or not it exists and, if so, its name, data type, size, relationship to other data items and so on. At the next (semantic) level there is the matter of what it means, as already described and illustrated above.

5.1 Data Definition and Syntactics/Semantics

The way data is defined and structured in an information system, an aspect of syntactics (as shown by Stamper’s semiological ladder in Figure 1), has important implications for the possible semantics that can be supported. In fact there is in general a set of possible data definitions/structures that can support a particular semantic view and, in the reverse direction, several distinct semantic views that can be supported by a particular data definitional structure. Nevertheless, some definitional structures and semantic views will be incompatible. The definitional structure will at least to some extent constrain the possible semantics that can be captured, albeit only partially, by it.

As a simple example, consider the meaning (i.e. semantics) of the term unemployed and how this may be captured in a data definitional (syntactic) sense. Most people, hearing the term would immediately understand it and, even if only intuitively and implicitly, have some sort of internal semantic definition of it. Probably that definition would be expressed by a syntactic construction along the lines of “A person is unemployed if they do not have a job”. This definition could be captured by, say, a relational database containing two tables: Person(name, ...., job) and Job(id, title, ...., salary).

Then the unemployed are simply those persons whose database records are not linked to any job record. However, a different definition of “unemployed” might be “A person is unemployed if they do not have a paid job”. Though quite semantically distinct from the first definition this is equally well supported by the relational database design. That is, it is possible for the same definitional structure – the database in this case - to support different semantics. That is, in answering the question of who is unemployed the same structure can support either meaning of “unemployed” although, of course, the outputs in each case would be very different.

On the other hand, it is easy to think of definitions of “unemployed” that the basic database design above cannot capture at all. For example: “A person is unemployed if they do not have a paid job or they work at a paid job for less than 10 hours per week”. This simple example therefore establishes the point that a given data definitional structure may support multiple semantic interpretations but, at the same time, also constrains those that are representable within it.

5.2 Control of Data Definition

Data definitions are not static. They evolve over time as understandings and needs change. From the point of view of an organizational actor it is essential that their data definitions and structures stay in step with their (continually evolving and changing) semantic understanding of their organizational world.
From this perspective, an organizational actor that is dependent upon a particular data resource (e.g. a database) risks the loss of this compatibility if control of the data definitions resides elsewhere than with themselves. For example, they might have to go “cap in hand” to the Information Systems Department to get a data item definition changed, only to be told that there are many others in the queue ahead of them, or perhaps that their requested change cannot be implemented because it conflicts with someone else’s requirements. Because of this, the situation shown in the SN of Figure 4a is likely to be much more acceptable to them than that of Figure 4b.

![Figure 4](image)

In Figure 5a, actor A is in full control of the data resource both in terms of accessibility to it, construction of syntactic/semantic outputs based upon it, and its definitional and structural evolution. Inasmuch as A considers this possession of control over access to and definition of the syntactic resource to be justified and right, they are likely to consider it belongs to them – that they own it.

Furthermore, attempts to broaden access to the data holding owned by A in Figure 4a are likely to be met with resistance by them, or at the very least reluctance and concern, for the reasons described above. Specifically, there is in such broadened access always the risk of “incorrect” interpretation by others.

6. An Example Case from the Literature

Loeb, Rai, Ramaprasad and Sharma (Loeb et al, 1998) describe the implementation of a “Global Information Warehouse” or GIW within an IBM division. The aim was to “integrate data spread all over the world [to] solve the data accessibility problems” that were being caused by the existence of different legacy databases containing data in incompatible forms and formats. The approach was, however, not to construct a single enterprise-wide system to replace the existing databases but rather to build a data/information warehouse that would extract data from these and other external databases.

6.1 Data Ownership

Discussing the challenges to implementation of the GIW, Loeb et al (Loeb et al, 1998) note that it was the political and cultural challenges that were the most difficult. In fact “the most difficult barrier to overcome was the ownership of data.” Their words in this area are worth quoting at some length [emphases added] with added commentary where appropriate:

“It was the political structure and organizational culture that made people feel that information was something to be carefully analysed and evaluated, not necessarily an
**asset that needed to be shared.** From their perspective, the GIW system was taking what they always owned and giving it to ‘everyone’. … If information is power then the understanding (analysis) is even more power.”

This reflects the discussion above under Figure 2 where the point was made that an actor may be prepared to provide appropriate syntactic (and semantic) outputs that are “carefully analysed and evaluated” but without allowing the underlying data to be accessed.

“There was also a sense of insecurity among ‘owners’ of data. … **Owners did not have a chance to check its meaningfulness and accuracy.** This induced fear in owners – fear of not being able to correct their data before it was too late. Driving away this fear and insecurity proved to be a major step in securing customers’ cooperation.”

This is again supportive of the point about willingness to provide interpreted syntactic/semantic output but not direct data access to external actors.

**6.2 The GIW Semiotic Network**

The type of situation, in Semiotic Network terms, of the IBM division GIW is shown in Figure 5 where “Def” stands for “Definition”. Syntactic/semantic links are not shown for reasons of diagrammatic simplicity. Nevertheless, it is evident that the GIW provides the possibility of independently generated syntactic/semantic outputs by different actors from identical commonly accessible data, and that these may conflict (as in Figure 3). This was clearly a significant cause of resistance faced by the GIW developers and one that was very difficult to overcome.

![Figure 5](image_url)

In Figure 5 it is of interest to note that the definitional links remain under the control of the original owners. This would not have been the case had the approach been to try to develop an overall enterprise-wide corporate database as the target solution to the integration/data accessibility problem. We may surmise, based on the argument presented above relating to data definition, that such a “solution” (the SN for which is shown in Figure 8) would have been politically and culturally even more difficult than the data warehouse/GIW approach that was actually taken. And, as Loeb et al report, that in itself was obviously difficult enough.
6.3 Discussion

Interestingly, Loeb et al make use of semiotic concepts in their description of the GIW case. However, they argue that the system itself transforms the original data from the syntactic level through to semantics by the provision of GIW charting, querying and information delivery tools, and finally to pragmatics when the top level of the senior managers and other GIW users is reached. This is rather different from the semiotic view propounded here, however, and moreover they do not attempt to provide a semiotic explanation of the data ownership issue as is the aim here.

The SNs of Figures 5 and 6, while derived from the discussion of the IBM division’s GIW are nevertheless more generally applicable than that. In particular, they give an important reason why data warehouses are generally politically and culturally easier to construct than all-inclusive corporate-wide integrated systems. The former, whilst allowing broader direct access to data – which is politically sensitive enough by itself because of the possibility of syntactic/semantic conflicts – still allow for the original owners of the data to retain control of their data definitional/semantic alignment. All-inclusive corporate-wide integrated systems, however, inherently imply that this control must be surrendered as well, at least by most of the involved organizational actors (see Figure 6).

Similar issues clearly face those who would implement Business Process Re-engineering in an organization, at least in its original wipe the slate clean and start anew form proposed by Hammer and Champy (1993). Such wiping clean of the slate and starting again obviously eliminates, or at the very least threatens to eliminate, established syntactic/semantic and data definitional links between actors and data holdings, with all its attendant consequences for possible mis-interpretations and definitional misalignment described above.

7. Implications

Data ownership perceptions in organizations and the consequent unwillingness to share are often considered an entirely and unarguably bad thing. For example, according to Martin, DeHayes, Hoffer and Perkins (1991) “A major source of resistance to managing data as a shared corporate resource is the history of data and system ownership”. Moreover, “consolidated and coordinated plans for data are needed to rid organizations of miscommunication and open new opportunities not possible with blinders on data” (Martin et al, 1991, emphasis added). However, the semiotic viewpoint proposed in this paper would indicate that the issue is not quite this simple.

Free access to data by anyone in the organization who needs it, or thinks they need it, leaves the way open to them to interpret it within their own context and according to their own worldview. If they are somewhat removed from the “owners” of the data there is, then,
perhaps a significant risk of them mis-interpreting it. This would most probably not occur if they received instead the pre-digested syntactic/semantic version from the data owner. Balancing this, of course, is the more widely known and understood risk of data owners deliberately hiding or misrepresenting data that they perceive may be damaging to them if released or made accessible “in the raw”. This too is a valid concern. Nevertheless, as argued above, the issue turns out to be less simple than at first sight and in fact turns out to be a two-way street. Data ownership perceptions and unwillingness to share are understandable, even perhaps justifiable, and possibly are not always the unambiguously bad thing for organizations that they have hitherto largely been assumed to be.

8. Conclusion

The open sharing of data within organizations, with accessibility to anyone who needs it whenever they need it, has for a long time now been something of a Holy Grail. CIOs and Data Administrators pursue it, IT professionals and system developers try to provide systems that deliver it, and academics bemoan its difficulty of achievement. By contrast, on the basis of a semiotic approach this paper has argued that there are powerful and perhaps even justifiable reasons why data sharing does not readily occur. There are potential risks as well as benefits involved, and those actors in organizations who consider that they “own” their data may have good and defensible reasons for feeling that way, not only on their own behalf but also with respect to the organization to which they belong.

Moreover, the semiotic approach seems to offer a way of pulling together, in a theoretical sense, at least some of the experiences relating to a variety of areas such as BPR, information management, enterprise-wide integrated system development versus end-user computing, data warehousing and the like.

References


Sless D. *In Search of Semiotics*, Barnes and Noble, New Jersey, 1986.


