

Data, Information, and Knowledge

a semiotic view of phenomena of organization

Frieder Nake

Informatik, Universität Bremen, D-28334 Bremen, Germany, nake@informatik.uni-bremen.de

Abstract: Phenomena of data, information, and knowledge are important for an organization to function. It is therefore important to agree on concepts for those terms. But the nature of such phenomena may not allow for a broad and lasting agreement, because they possess a historical dimension. In this situation we offer semiotic explications of the three terms. Data is viewed here as the syntactic reduction of a sign, information as its semantic reduction, and knowledge appears tied to the pragmatics of the sign. While we do not suggest that the views expressed here are overly new, we feel they offer a useful perspective on the difficult relation between precise formalism and vague insight.

Key words: Data, information, knowledge, Peircean semiotics.

1. AN INTRODUCTORY REMARK LEADING UP TO SIGNS

The Call for Papers to this Conference on Organizational Semiotics announces as its overarching theme: "evolving a science of information systems". A specific question is: "Can we apply the rigor of formality while including the human aspects?" I want to address this question as a background to the more immediate issue of a differentiation of data, information, and knowledge.

A contradiction of formality and humaneness is implied in the question. It may not be possible to satisfy both requirements simultaneously. On one hand, a *formal* theory appears desirable because formality could support precision and predictability. On the other hand, inclusion of *human* aspects

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35611-2_22](https://doi.org/10.1007/978-0-387-35611-2_22)

appears necessary in a theory of organization, because that could provide validity and motivation on behalf of those who constitute an organization.

Why and how should formality contradict humaneness? A formal theory, like any theory, no matter how formal, is always of human origin. On a very general level, there does not appear to be a contradiction. But the current context is more special. "Human aspects" does not refer here to humanity in the broadest sense but rather to the consideration of human needs when people are exposed to applications of information technology.

The contradiction is between form and content. The degree of formality is a matter of form, the level of human aspects a matter of content. They may not fit together.

Nothing in the world is pure form or pure content. Form and content are merely aspects under which we decide to view the world. It is useful to do so in our attempt of coming to grips with human existence. The more we stress form, the more formal affairs become.

It happens that an author presents a formal definition of concepts and, at the same time, assures us that he was not neglecting human aspects. Experience, however, tells us that we are sometimes forced to drop essential human characteristics when we define formal concepts. Formalism requires independence of context, whereas human aspects are rich of context.

The problem is not formalism by itself. If we gain deeper insight by excluding the context of a situation – why should we not try? The problem is rather our interpretation of what we could possibly achieve by formalism. We create a serious problem only when we *identify* a formal system with a social organization. Human needs and an institutional purpose are the fabric of the organization. They are so heavily influenced by complex networks of contexts that any formalization should be done with care and humility. Formalization pretends that we fully understand a phenomenon, since, in formalizing, we study it out of context. But epistemological skepticism tells us that, in the end, we cannot understand. All we can hope for is the establishment of systems that we study in place of the original situation. In science, we never deal with the "real" thing, but only with models. We try to escape the despair that inevitably emerges from that state of affairs, by the invention of models, aspects, structures. We are justified in claiming that we have managed to understand some phenomenon – but we should not forget that the best we can hope for is a close fit between phenomenon and some semiotic layer that we use to replace the actual thing by.

Our starting question boils down to an interpretation of what we mean by "including human aspects". As long as we are content with a vague inclusion of human aspects in the rigorous treatment of an issue, we may even apply formal methods. There is a great difference between grasping the essence of a process and describing it in a rigorous formal manner.

We may call these two conditions the condition of *essentiality*, and the condition of *formality*, respectively. The condition of essentiality aims at a deep understanding of the phenomenon; it is of a subjective, wholistic, and intuitive kind. The condition of formality aims at a mathematical formulation of the phenomenon; it is objective, reductionist, and explicit. Essentiality leads to religion, formality to science.

The fundamental difference between essentialism and formalism has been widely acknowledged in recent computer science literature. We name, in lieu of many more, only Winograd & Flores (1986), Ehn (1988), Floyd et al. (1992). The crucial point appears to be the treatment of context. In fact, the more we strive for formalism, the more context we must ignore. Alternatively, the closer we attempt to arrive at the essence, the more context we have to take into consideration.

René Magritte, in his painting *La Trahison des images* (Fig. 1), ingeniously expressed the difference we are talking about. By painting the image of a smoking pipe and commenting on it with the line, *Ceci n'est pas une pipe*, Magritte throws the spectator into a semiotic dilemma. The painting has become an icon for the difference of showing and naming. M. Foucault has discussed the problem (Foucault 1983), and Karl-Heinrich Schmidt (1992) has taken it up in his critique of a radiologist's expertise in analysing X-ray images vs. giving a written account of a patient's condition.

Expressed very blatantly, when we show a phenomenon, we preserve the whole but may not have a concept, while when we name it, we lose the phenomenon itself for sake of a concept of it. At closer inspection, the problem of simultaneously dealing with formal expression and essential impression is to the heart of the dialectics of surface appearance and deep essence, or of explicit form and implicit content.



Figure 1. *La Trahison des Images*, René Magritte 1928/29

Rigorous formalism necessitates naming, and it does so in a most radical way. In a formal context, the name occupies the place of the whole by replacing it. But either way, in naming and showing, we enter the realm of *signs*. The name of an entity is obviously a sign standing for the entity; showing the entity becomes possible only through the use of pointers or

other indices, which are also signs. There is no escape from semiotics, and in semiotics we gain a level of study where the impossibility of simultaneous explicit rigor and implicit insight disappears in a play with signs.

2. DISCLAIMER

What does it entail to view an important aspect of this Working Conference as a dialectics? To consider a phenomenon dialectically means to consider it in evolution and under varying aspects. In this note, I offer a semiotic perspective on the concepts of data, information, and knowledge. The perspective is subjective, and I do not claim to present ultimate definitions of the three terms. In fact, the introductory remark is to say that we should rather give up hope to ever find ultimate answers.

Ultimate answers could be turned into definitions that catch the essence. The search for them is a driving force behind rationalism. Rationalism has had a hard time to realize that such definitions cannot be given. We should therefore give up the attempt to find them. We may, however, give explications. An explication is a weaker form of introducing a concept into a (more or less) formal theory. Where the definition should apply to all cases of some kind, the explication only provides a circumscription of an intuitively more or less clear, yet vaguely formulated, concept. It is a statement of temporarily convincing power, good enough for a discourse to continue in good faith that all interested parties agree on some terms until someone objects. The explication is open to new contexts whereas the definition belongs to one particular context.

This short essay will thus give explications of the terms data, information, and knowledge. It will do so by reference to the three aspects of the sign: syntactics, semantics, and pragmatics.

I should caution the reader that, within a short paper like this, ideas are no more than indications. There is a deep and rich discourse on the semiotics of information systems. In the recent collection (Liu et al. 2001), papers by van Heusden & Jorna and by Stamper are particularly relevant to our discussion. It remains to be shown where we agree and disagree.

3. ORGANIZATION

”Organisation is achieved not by doing things but by talking and writing about them”, says Ronald Stamper (1973: 8). Considering that his book is by now almost thirty years old, it carries an astounding message. Information is announced as the fabric of an organization, and organization is the main

topic of the book. Information, on the other hand, is distinctly related to signs. Here we have an early source for the roof of this Working Conference: first the idea that organization is a matter of information and communication – an idea shared by many in organizational theory. Second, more important from today's perspective, the idea that information is a matter of signs.

Semiotics is always a good bet when a phenomenon is so complex that the function and structure of its parts are not conclusive to explain the phenomenon well enough. In our case, an organization is a social structure in pursuit of some aim and scope. Individual members of the organization have to agree to at least a minimal amount with the overall aim and scope of the organization to function. But the organization is abstract. It emerges from the cooperation of its members, and it in turn influences its members' activities. Information and communication appear as decisive on either level because they are relational by nature. If the progression of institutions depends on the division of labor, then communication and cooperation (i.e. organization) are needed to keep things together.

Semiotics may serve as a conceptual framework for such situations. Social processes are contradictory by nature. Contradictions drive developments. Semiotic processes are descriptive means for dealing with contradictory processes. Is it possible to advance the semiotic framework from only descriptive means to predictive instruments? This appears to be the question behind the gap between human needs and formal rigor.

However, human needs develop in a dialectic relation with the means of their satisfaction. If organization is a way to satisfy human needs in a social context, then the successful organization produces its failure: the better the organization leads to the satisfaction of needs, the more those needs change, thereby making necessary a different organization.

Formalism could never cope with this contradiction. The best we may hope for is a descriptive method that carries within itself the power of expressing contradictory processes. Peirce's recursive concept of sign seems to provide a way to exactly this end.

4. SIGNS AND SEMIOTIC DIMENSIONS

A brief indication must suffice to mark my position within the semiotic context. I refer to semiotics as the conceptual framework created by Charles S. Peirce (for a short high-level account cf. Nöth (1990); a readily available source of original texts is Houser & Kloesel (1992, 98)).

A sign is given by a relation. It is not a thing. The sign relation possesses three correlates called representamen, object, and interpretant by Peirce. He was not particularly consistent in his terminology – a fact indicative of his

permanent strive for deeper insight and for explicit expression of what he felt was the essence of the phenomenon of communication. Communication between humans intrigued him, but also communication in a much broader sense. The sign emerged as his concept to describe all phenomena of communication (and thereby of knowledge).

In a sign, according to Peirce, a representamen stands for an object by virtue of an interpretant. The interpretant is an expression of what the sign *means*; the object is an expression of what the sign *signifies*; the representamen is an expression of what the sign is *made of*.

It is important to note that the interpretant, and the sign itself, are connected to a human mind, an intent, a desire. The interpretant is an expression of the human's intent. Therefore, if in a given semiotic situation, we want to make the interpretant explicit, we have to create another sign. The sign is recursive. We have to continue this process indefinitely, if we want to get the whole picture. In real life, however, we are forced to interrupt at some point in time. This amounts to our principal inability of ever totally grasping a sign's meaning. All our understanding is broken and partial. Only the infinite God could totally understand a sign's meaning. He would need only one single sign because all the others would appear somewhere in the indefinite recourse. Our partial understanding is, however, good enough for most practical purposes of communication.

Anything of our choice can be turned into the representamen of some sign. Nothing is a sign unless it is declared to be a sign. The world is not split into signs and non-signs, nor into things and non-things. But when we think about the world, we cannot but think in terms of signs. This is to say we generate signs as our way of understanding the world.

Some thing becomes the representamen of a sign by an act of the mind. Any such act establishes relations (i.e. signs) between things or processes in the world. It needs an individual human mind to start a process of semiosis, but a sign acquires the culturally saturated status of sign only by social processes that are beyond the control of individuals.

The three correlates of a sign cannot be taken away from the sign, and they cannot be understood in isolation. Nevertheless, once we introduce this distinction, we cannot but use the terms separately. The three dimensions of the sign are syntactics, semantics, and pragmatics. They were introduced by Charles W. Morris (cf. Nöth 1990). We give a short explanation of the way we use those words in the present context.

In the *syntactic* dimension, we ask, "How does the sign signify?" The sign is reduced to its corporeal aspects which really says that in syntactics the sign loses its genuine character as a sign, or rather: that the sign is reduced to a special kind of sign, which we call *signal*. The signal is a sign in a state where the object and the interpretant tend to coincide.

In the *semantic* dimension, we ask, "What does the sign signify?" We need the sign as representamen and as object in order to discuss that question. If syntactics is semiotics reduced to material aspects of signs, then semantics is a dyadically reduced semiotics: the field of linguistics.

The semantics of the sign discuss those aspects that are generally agreed upon within a social group, community, or culture. In a dictionary, the bold-faced entry name is the representamen; the explaining text (and picture) attached to the name gives its semantics. An individual may disagree with the particular explanation given to the entry in the dictionary. But the authors of the dictionary would usually take great pains to write texts that do not spark too much of a controversy within the community and context. The signification is the conventional meaning of the sign.

In the *pragmatic* dimension, we ask, "Why (and what for) is the sign signifying?" The sign gains its full context only within the purpose of its use. Personal usage of the sign is the pragmatic theme, and quarrels about meanings of words, importance of art works, or the impact of movies are subject matter of sign pragmatics. The study of signs cannot be complete without the pragmatic dimension, and the pragmatic dimension is the genuine dimension of semiotics. Syntactics and semantics are good for a thorough discussion of details, but they have no semiotic relevance by themselves.

We introduce a symbolism to express the triadic relation of a sign and its not-symmetric aspects. The sign can be taken as an R related to an O and these two related to an I in the following directional sense:

$$((R \leq O) \leq I) \text{ or } ((R \Rightarrow O) \Rightarrow I).$$

The left form depicts the *generation* of the sign triad out of a preconceived intent (I) to its perceivable matter (R), whereas the second form shows the act of *interpretation* of the sign from sensual perception (R) to meaning (I). We use the neutral notation $((R - O) - I)$, if we want to leave open which of the two directions we consider. The notation reveals that the interpretant, I, is related to the pair of (R - O). The object is related to R, and R may, within the semiotic context, be taken in isolation.

5. EXPLICATIONS: DATA, INFORMATION, KONWLEDGE

We now link the words "data", "information", and "knowledge" to the three semiotic dimensions. This linkage expresses some insight into the concepts behind the words.

There are reasons to connect "data" to the syntactic dimension, "information" to the semantic, and "knowledge" to the pragmatic dimension. The

reason is the characterization of the dimensions in the preceding section. The syntactic question of "How" leads us to data (the category of firstness, in Peircean terms); in semantics, a phenomenon is questioned by "What" which relates the signifying data to the signified information (secondness); finally, in pragmatics the question of "Why" relates a signifying pair to our subjective knowledge (thirdness).

Knowledge is a person's lived life. It is total, whole, and inseparable. We feel a need to make our knowledge explicit. When we do so, we introduce a new entity: we express and therefore we reduce. At times we mistake explicit knowledge for the total of knowledge. This is most common in computational contexts.

- *Explicit knowledge* is a sign in the full triadic relational meaning of "sign". It is a sign in the pragmatic dimension. Explicit knowledge is what is left of knowledge when it is put on paper (or expressed in some other way). In the formalism it appears as $((R - O) - I)$.
- *Information* is explicit knowledge (a sign) reduced to its semantic dimension, and viewed as a sign. Information is what is left of knowledge when the subjective context and situation is abstracted away. What is left then is the conventional or cultural aspect of knowledge. In our formalism it is $(R - O)$.
- *Data* is information (a sign) reduced to its syntactic dimension, and viewed as a sign. Data is what is left of information when the cultural context is abstracted away. What is left then is the material aspect of information. In the formalism: (R) .

All three – explicit knowledge, information, and data – are signs although they are introduced here as reductions of signs. This is possible because of the intrinsically recursive character of the sign in Peircean semiotics.

The explication starts with knowledge made explicit, and identifies it with the sign in its full meaning. We stress that this starts a process much more than it creates an isolated thing. A sign never exists in isolation. It separates one sign situation from another one.

From explicit knowledge we arrive at information by an operation that neglects pragmatics, but treats the remainder as a sign. The study of information cannot be understood much differently: we proceed from individual meaning to cultural context, thus creating new pragmatics. In science, we reduce the difference between O and I.

Similarly, we reach data from information by first denying semantics but then re-embedding into a full semiotic triad. We take care to concentrate on the reduced representamen R. This step is decisive for the use of computers as information systems. The explication should make clear that the computer is concerned only with data, and that is hard enough. It does not leave room for automatic processing of information in our use of the word.

6. DISCUSSION

Space does not permit a broad discussion of the general concepts applied to special situations. As one such case, human-computer interaction has been treated in (Nake 1994). The conceptual distinction introduced here proved to be helpful to clarify the typical interactive situation.

The necessity to differentiate data from information and knowledge (or the machine view from the collective and individual views) originates in the vast data-bases that today form a second reality layer on top of reality as we usually experience it. Those data get read, interpreted, and operated on in two essentially different ways: by humans and by computers. The human reader of data is free to interpret in what ever way he or she chooses. Through such an act, data get instantaneously embedded and enriched semiotically and are turned into triadic signs by acquiring objects and interpretants of human significance.

The computer as reader of data, however, is bound to no more than exactly one interpretation. Its act of interpretation reduces to the *determination* of a machine sequence of operations. There is no *meaning* for a computer, only a signification. In a forthcoming book, Andersen and the author introduce the dual concept of an intentional and a causal interpretant to describe that discrepancy. The semiotic view provides a rational approach to the issues of human and computer behavior that otherwise could stir up heated debates. We briefly take up one particular case.

A computer-based information system in an organization is a socio-technical system whose technical subsystem resides on a computer. We consider it as nothing but a data storage. Inside the computer memory, and during all kinds of computer processing, the entities of that system must be classified as data and nothing else. As such they are signals, i.e. signs of the most reduced form. Only when data get output to the periphery, the human user embeds the data into contexts and thus turns signals into signs. Within those contexts, knowledge and information emerge. We should be very clear about the actors here: signs, and thus information and knowledge, appear only by constructive activities on behalf of humans. The fabric of the organization remains entirely within the social subsystem, outside of the reach of the technical subsystem. Only the way the organization is organized changes when an information system is introduced.

Our explication of the words data, information, and knowledge has lead us to a sober view of an information system within an organization. We should, however, keep in mind the dialectics inherent to any change of infrastructure. Once the initial enthusiasm has gone, the technical system develops its specific dynamics. It is likely that people create new needs because of their acquaintance with the technological power. The semiotic

analysis offers a way to describe such processes as semioses between parts of an organization. Once these semioses emerge, they may generate an impression as if the technical system itself created them. Perhaps this note pulls away the veil from such superficial belief.

In closing, I would like to crossreference the work of the FRISCO group. Since 1988, their contributions to terminology have centered around the same three concepts. When scientific terms can be defined in a quantitative way, they can be subjected to hard experiments. It is not possible, in any meaningful way, to quantitatively define “knowledge”. In the case of “information”, Shannon and Weaver have tried just that. But organizational semiotics does not get anywhere with that measure of data channel capacity.

We must concede that knowledge and information (data to a lesser extent) are qualitative concepts that emerge from semioses. It appears unlikely that they can be *defined* precisely: the circumstances of their appearance can be *described*.

Stable concepts – definitions! – are what scientists like so much to cling to in their continuous search for knowledge. They are looking for rocks in a river. There they meet both, the hard and the fluid, intricately in exchange.

REFERENCES

- Andersen, P. B. & Nake, F. (2002) Semiotics and informatics. In preparation
- Ehn, P. (1988) Work-oriented design of computer artifacts. Stockholm: Arbetslivscenter
- Falkenberg, F.D. et al. (eds.) (1998) A framework of information systems concepts. IFIP WG 8.1 Task Group Report
- Floyd, C., Züllighoven, H., Budde, R. & Keil-Slawik, R. (eds.) (1992) Software development and reality construction. Berlin, Heidelberg, New York: Springer Verlag
- Foucault, M. (1983) Dies ist keine Pfeife. Frankfurt, Berlin, Wien: edition passagen
- Houser, N. & Kloesel, Ch. (eds.) (1992, 98) The essential Peirce. Selected philosophical writings. 2 vols. Bloomington, IN: Indiana University Press
- Liu, K., Clarke, R.J., Andersen, P.B. & Stamper, R.K. (eds.) (2001) Information, organisation and technology: studies in organisational semiotics. Boston, Dordrecht, London: Kluwer
- Nake, F. (1994) Human-computer interaction: signs and signals interfacing, Languages of Design 2, 193-205
- Nöth, W. (1990) Handbook of semiotics. Bloomington, IN: Indiana University Press
- Schmidt, K.-H. (1992) Texte und Bilder in maschinellen Modellbildungen. Tübingen: Stauffenburg
- Stamper, R. (1973) Information in business and administrative systems. London: B.T. Batsford
- Winograd, T. & Flores, F. (1986) Understanding computers and cognition. A new foundation for design. Norwood, NJ: Ablex