Chapter 1

The value of information in the "e-age"

Alex Verrijn-Stuart & Wolfgang Hesse

1. INTRODUCTION

The observation that fuzzy and ill-defined concepts – and hence, terminology confusion – adversely affect the quality and usefulness of information systems (or: "IS") is valid even more today than it was when the FRISCO Report (Falkenberg et al. 1998) was first conceived in 1987. As in that report, our point of departure is the "babylonian problem" [1] – but viewed in the context of the ever growing interdependence of systems based on "Information and Communication Technology" (or "ICT").

In section 2, it is argued that the problems of conceptual coherence and terminological consistency and hence the necessity of conceptual frameworks are not outdated by the current tendencies to build and use global IS. On the contrary, they are more important and challenging today than ever before.

Subsequently, we address two further, closely related issues of importance, namely: the *value* of ICT-products and services (section 3), and the *persistence* (or permanence) of ICT-provided information (section 4). Both topics will be considered in the light of today's networked environments. It is widely accepted that effective assessment criteria highly

[1] The confusion of languages (Genesis 11.9)

depend on one's point of view. We will demonstrate the power of the FRISCO "semiotic ladder" approach to find the appropriate level of assessment and, generally, the benefits of conceptual frameworks, that is to say, of clear systems of concepts, terms and relationships.

2. CONCEPTUAL FRAMEWORK AND TERMINOLOGY

The FRISCO Task Group was set up in 1988 in response to the feeling that the plethora of vaguely defined terms in the world of 'information systems' causes confusion and harms the very communication they are supposed to facilitate [see FRISCO "Manifesto" (Falkenberg et al. 1998 Preamble)]. It is our view that the globalisation and merging of networks enhances that tendency rather than overcome it. Another cause for this conceptual (and terminological) inflation tendency is the fact that new (or renewed) approaches are better recognised when they are given specific labels.

Typical examples, over the past three decades, were ER (Entity Relationship model), OO (Object Orientation), BPR (Business Process Reengineering), ERP (Entity Resource Planning). Querying the web via Google, the argument "ERP" led to over 500 finds, mostly commercial offerings of ERP-related services, with references to ERM (Enterprise Resource Management), APS (Advanced Planning and Scheduling)[2], MRP (Material Requirements Planning) and CRM (Customer Relationship Management). "Entity Relationship Model" yielded 44600 hits, "Object Oriented" branched into untold directions and "BPR" also brought up a link to a snowmobile supplier.

The acronyms mentioned mostly concern methodological approaches in business modelling, software-engineering and other ICT related fields. As rather specific methodologies, they would not normally attract FRISCO-like analysis, were it not for the hype surrounding them, which incorrectly suggests that aspects are brought into play that had not been considered before. The guru status acquired by their originators causes any questioning to be felt as heresy (and/or jealousy). Yet, for proper appreciation of their innovation, benefits and limitations, it is important to establish some link to subject areas of common understanding. The adoption of a reference framework will help doing just that.

[2] To photographers, however, APS stands for the modern Advanced Photographic System

In this paper, we will not apply the FRISCO concepts to terminological questions, as such, but more specifically consider the area in which every new term has the prefix "e-" (admitting that our title is an example of this). The "e-" (for 'electronic') is associated with activities characterised by a strong element of messaging via an internal or external network, such as "e-mail" and "e-commerce". The requirements for successful entry and positioning in this environment are often perceived as lying on a different level as compared to the "old economy". In fact, they are already implicit in any fundamental view of organisational information systems and communication. The main difference lies in the time scale of the messaging: 'classical' information systems tend to provide support for 'off-line' decision making, whereas the interactive nature of "e-systems" permits on-the-spot commitment, such as enquiry-and-ordering (e-commerce) via the Internet. Likewise, an enterprise-wide network (" intranet") permits a style of seamless organisational communication that is at once less formal and more pervasive.

Anyway, the above and other instances of buzz words either deal with elementary components and their structural aspects in system descriptions, or with ways of pragmatic modelling composite structures in a specific organisational environment. The most general (and elementary) component is a *'thing'* in the real or conceived world and therefore subject to one's world view or 'ontology' [3]. Complexity is involved when combinations of components are taken into account as *'models'* and more specific *'systems'* [4]. The idiosyncratic nomenclature of detailed business analysis hide the fact that their model and system denotations may be expressed in more general terms and hence reduced to common concepts. It is the specificity and terminological complexity of the descriptions resulting from commercially offered software and modelling applications that make this difficult.

This situation is even exacerbated when *world-wide IS* with all their redundancies, inconsistencies, their mixtures of pearls and garbage wrapped (and obscured) by a lot of internet-related jargon come into play. The FRISCO report defines

Definition E40: Information system[5]

[3] See FRISCO, section 3.1 ('Weltanschauung'); note that the term 'ontology' is used by FRISCO in the traditional sense of 'view of the world'; however, in current knowledge system work, it often stands for 'glossary' (list of canonical descriptions). See also (Braun et al. 2000) for the foundation of "things" and "conceptions" as social constructs.

[4] See FRISCO, sections 3.6-3.7.

[5] In FRISCO, section 3.9, definition E40 is qualified as follows:
Warning: This term is used in various different meanings in the literature!
Synonym: Information System in the Broader sense (Acronym: ISB)

> An **information system** is a sub-system of an organisational system, comprising the conception of how the communication- and information–oriented aspects of an organisation are composed (e.g. of specific communicating, information-providing and/or information-seeking actors, and of specific information-oriented actands) and how these operate, thus describing the (explicit and/or implicit) communication-oriented and information–providing actions and arrangements existing within that organisation.

Hence, the concept of IS is basically reduced to those of *information*, *communication* and *organisation*. The definition is still valid for global IS if we consider the "world community" (or at least its internet-connected parts) as one big "organisation". In fact, this comprehensive view is implied in FRISCO's assumption regarding an *organisation:*

> **Assumption [n]:**
> An *organisation* is a grouping of actors, together with a collection of actands, such that (a) a common goal is pursued or some other characteristic coherence is displayed, and (b) interactions occur that are based on communication. *Information* is used in an organisation in the context of its functioning (actions), both internally and with respect to its environment, the society at large. Because of this social involvement, *norms* are meaningful directives for individuals in the organisation and the organisation as a whole.

We can derive from these definitions what are essential prerequisites for making a global IS a valuable tool for its users: common goals, (social) coherence, functioning communication, individuals driven and directed by social involvement and common norms.

In the next section, we shall apply this line of generalisation of concepts in an area that has not been considered very much to date, namely the economics and persistence of ICT services and usage.

3. THE 'VALUE' OF DATA, INFORMATION AND KNOWLEDGE

In the FRISCO Report (Falkenberg et al. 1998), it was argued that the major moving force for using ICT is the economic advantage ('value addition') to be obtained. In the versions published so far, that basis was not further discussed.[6] The fact that 'information' is intrinsically valuable by enabling rational decision making was implied, but the objective of the report not to advocate any specific methodology precluded a detailed analysis of the elements that make it so. Although we do not ignore the obvious dominance of purely economical considerations and assessments,

[6] However, attention will be paid to this aspect in the forthcoming Revised Report – see (Hesse and Verrijn-Stuart 2000)

we would like to extend the concept of 'value' even further, e.g. by including aspects such as personal independence, education, self-esteem, quality of life, political culture, social harmony etc. It will be seen that that this does not affect the procedure for arriving at sensible results.

As is done more generally, the FRISCO Report distinguishes 'data' and 'information'. However, FRISCO differs from many other approaches in that *data* is defined as a *representation* of knowledge[7], but *information* as an *increment* of knowledge[8]. This follows from FRISCO's view that 'knowledge' is strictly personal, although it may be 'shared' by communication and negotiation[9]. Thus, the elements and concepts involved in a personal setup may be spelled out, even if the full model for relating a specific messaging arrangement and the value of decisions based upon them remain complex. This holds even more so for the value of information within an organisation (or a part of it, such as 'management' or 'a department'). As mentioned above, organisations may possess 'shared knowledge', resulting from internal exchanges and agreement. However, the economics of personal and organisational information cannot be expressed in straightforward relationships.

In the "e-age", these relationships remain as indirect as before - often even more so, and in different ways than in the traditional environment. Again, any misinterpretation may have grave consequences. In this section, we illustrate the options for economic reasoning through a simple example. It may be generalised to cover more complex situations. The case concerns the rendering of some unspecified service, originally sold over the counter of a physical office or shop, but now also offered via the internet. In both setups, customers must be able to make enquiries and place orders.

In FRISCO terminology, the applicable concepts are *actors*, *actions* and *actands*, and their specialisations *sender*/*channel*/*receiver* (actors), *sending-action*/*channelling-action*/*receiving-action* (actions) and *messages* (actands), respectively, as illustrated in Figure 2 [10]. Messages consist of *data*, i.e. representations of knowledge. We shall revert to the semiotic aspect of 'interpretation' by sender and receiver, in a moment.

[7] E34: The term *data* denotes any set of representations of knowledge, expressed in a language.

[8] E36: *Information* is the knowledge increment brought about by a receiving action in a message transfer.

[9] The relevant FRISCO definitions imply:
def 34 : Data ~ representation of personal knowledge
def 35 : Message ~ data, transmitted with intent
def 36 : Information ~ increment of personal knowledge
def 37 : Communication ~ exchange of messages
def 38 : Shared knowledge ~ knowledge agreed within group

[10] See FRISCO, sections 3.5 and 3.8.

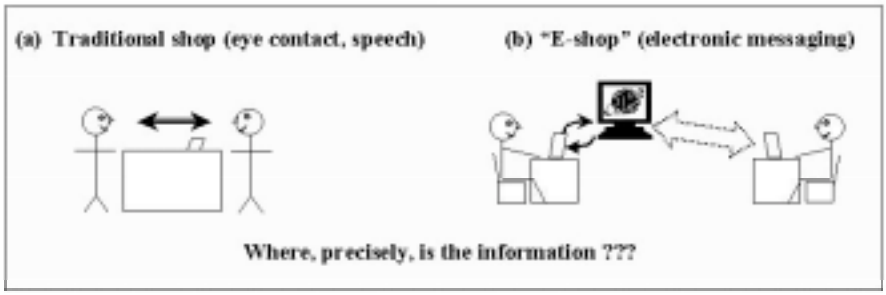

Figure 1 : 'messaging' by means of (a) printed brochures and (b) webposting/e-mailing

The diagrams of Figure 1 show (a) a physical shop and (b) an internet arrangement. In both cases, the customer and salesperson alternate in the roles of sender and receiver of messages, as illustrated in Figure 2. Sending may consist in asking for or providing 'information', filling in or confirming order forms and so on. In each of such actions there is an amount of new data (that may represent knowledge not yet acquired), true redundancy (embellishing data, checking data) and assurance (supporting data, but also data to refresh forgotten knowledge)[11].

The important aspect is that the selling party (service renderer or whatever) must not only be aware of the distinction between data and information, but also understand that the physical (medium, channel) appearance of the information-carrying messages does not necessarily bring about increments (or -refreshment) of knowledge, although often, it will do so. This applies to any kind of 'messaging' in the setup. Let us restrict ourselves for a moment to the activity of a seller advising a customer by means of documentary material.

It is customary but inappropriate to speak about the 'amount of information' packed into the brochures handed across the counter or (e-) mailed to potential customers (who then may be enthusiastic or complain about the overload of the message stream experienced). The FRISCO approach shows us a differentiated way for measuring various aspects of messages. This is where recourse to the so-called 'semiotic levels'[12] is helpful. These permit fairly precise distinctions of concept aspects, which natural language terminology tends to obscure.

[11] The various 'information roles' played by this data is further discussed below (section 3).
[12] See FRISCO, sections 3.4/5, in particular Figure 3.5-2.

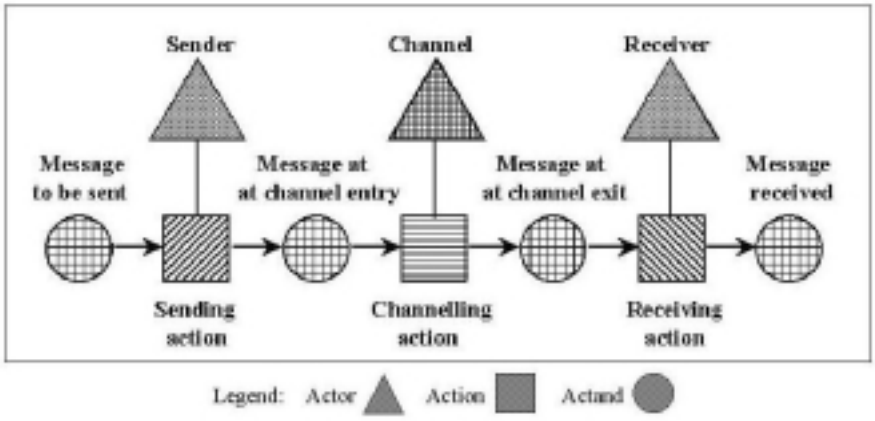

Figure 2 : Message transfer [after FRISCO Figure 3.8-1 (b)]

For instance, it is proper to view the flow of data as a more or less noisy channel, the message stream on which may be quantified in terms of bits, digits, encapsulated forms or whatever. The effectiveness of the channel may be expressed in terms of message elements passing over it in time, at some cost, including a charge for transfer losses. Thus, quantification is feasible at the physical/empirical levels. A well-known way to achieve this follows the Shannon/Weaver approach to "measuring" the contents (or better: representation density) of messages (Shannon and Weaver 1949). However, quantification becomes much more difficult when one tries to establish a link to the effect of the message, at the other end of the ladder, however desirable. The effort (and hence, cost) may be worked out, but the effect (and hence, benefit) is neither predictable nor directly proportional to it. In a way, the ‘amount of data’ is quantifiable, the ‘amount of information’ is not.

In the physical shop case, one of the decisions is how many brochures and folders to print, another, whether to mail them to some previously established list of potential customers, yet another, whether just to place this material on the counter and/or on display shelves. In all of these situations, the printed documents serve as messages, which are intended to provide information to the recipient (in the ‘knowledge increment’ sense), but may not necessarily do so, although normally they will not fail to arouse some form of perception on the part of the person coming across it. Although the intention of the ‘message sender’ (in the FRISCO sense) is to ‘inform’, the recipient can only include *data* representing his own knowledge (the verb ‘to inform’ is not a FRISCO term, of course). Economically, the expense (printing, mailing, exhibiting) should be more than offset by the ultimate revenue from whatever is sold (a physical product or service). The selling

party is aware of the fact that the decision to buy will be based on personal knowledge and motivation of the (potential) customer. However, only part of the recipient's knowledge has resulted from viewing the brochure – if at all – even if the triggering effect has been part of what tipped the balance.

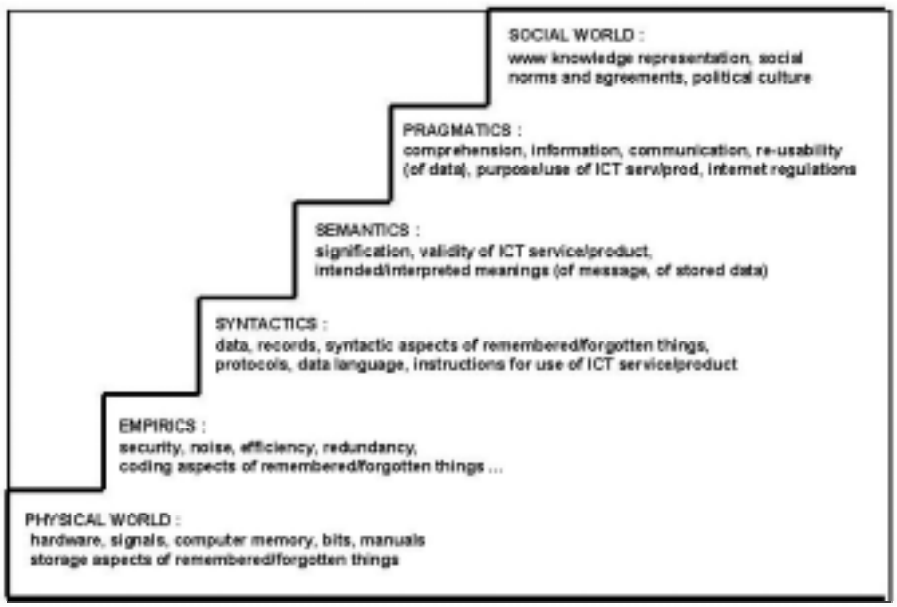

Figure 3 : The 'semiotic ladder' as applicable to 'e-age' concepts

The conclusion of this analysis is straightforward. The brochures and folders, with all their potentially interesting content, remain data and may or may not provide information (in the 'knowledge-increment' sense) to the 'message receiver' - depending on his or her knowledge status, ability or willingness to comprehend etc.. Nevertheless, the awareness aroused by them justifies *informally* calling such 'message-sending' an act of 'informing'.

In the e-business approach, the elements are the same. There are *actors* (the organisation offering the products or services via the internet, the potential customers and the networks as intermediates), there are *actions* (placing attractive pages on a website, e-mailing announcements to a list of potential customers, viewing web sites, reacting to e-mail messages or filling the order boxes associated with an interactive web page) and *actands* (web-pages, e-mail messages, e-enquiries and e-orders). Again, there are costs the seller should be able to work out in fair detail.

SEMIOTIC LEVEL	ELEMENTS OF STORAGE	ELEMENTS OF COMMUNICATION	VALUE-DETERMINING ASPECTS
Social level	Knowledge bases, libraries, etc.	Web communications	Knowledge privileges, economical & political power, social & democratic culture
Pragmatics	Enterprise information systems, repositories, data warehouses	News groups, chatrooms	Data privacy, decision support, market leadership, educational quality
Semantics	Data bases, information systems, web sites	Annotated data, graphics & images, e-mails	Data consistency & reliability, informedness
Syntactics	Text, programs, graphic elements, pixels	Messages, applets	Data storage costs & revenue
Empirics	Bit streams, recods, files	Data packets	Data security & integrity
Physical level	Bits & bytes	Signals, bits & bytes	Channel capacity, data transfer costs & revenue

Table 1: Static, dynamic and value-related aspects of knowledge representation.

Similarly, there may be revenues that hopefully make the e-business profitable. But, as in the physical shop case, the economic model is many-faceted, without any evident relationships between expenses and income. Certainly, the number of pixels and digits emitted is no direct measure of the information flow intended.

While the term *information* is part of ‘IT’ and ‘ICT’, of ‘information system’ and ‘information age’, it continues confusing the parties involved. Building some system, the designer, the software engineer and the future system user have different interests and ‘agendas’ [13]. Likewise, someone offering services or products via the internet must not confuse the ‘informing’ intent of his message stream the with information sought and appreciated by the potential customer. On the other hand, by considering the semiotic levels, the various relevant concepts may be identified more specifically and the economic value analysis narrowed more appropriately.

Whereas in Figure 3, a number of characteristic aspects relevant to the "e-age" problem area were just listed by semiotic level, in Table 1, we go one step further. Here, we separately consider the static (storage), dynamic (communication) and value-related (economic and non-economic) aspects of

[13] See FRISCO Report, section 1.2

knowledge representation and exchange. By doing this, we can isolate areas which are then more or less open to full valuation. Even so, the items mentioned are primarily suggestive, at this stage.

Now, let us attempt modelling the economics of the specific cases of the traditional and e-commerce arrangements, in particular the valuation of making available 'information' by means of over-the-counter brochures or web-posting (or e-mailing) them. From the semiotic point of view, this means that we have recorded some 'knowledge' as data and make available a selection of this data in some form.

At the physical and empirical levels, the size and cost of the message stream may be quantified by considering the amount of data stored or transmitted, together with the expense involved. There is no benefit in the form of directly related revenue, as such. However, having made the hardware and software investments needed, this may be used for other applications without incurring further costs.

The syntactic level is certainly involved regarding message formats, internet protocols and so on. Being able to cope with these requires training of the persons in charge of the storage and transmission tasks. Again, there is no directly related income but that training needs not be fully repeated for further tasks.

At the semantic and pragmatic levels, the provider is normally aware that the messages are intended to convey 'information' (but not necessarily thus perceived by the recipients). The owner of some data not only must must take care of its storage, but also of ways of composing messages out of it (cost). On the other hand, the recipient of a message must perform the task of interpreting and comprehending it (which also involves effort and hence cost). At these levels, there appear potential and actual benefits, both to the sender and the receiver. For the sender (whose intention was to trigger a hopefully positive reaction), the revenue remains potential, so in an economic analysis, this would be subject to some probability distribution. For the recipient, the benefits are tangible, namely, the ability to comprehend and subsequently to take action (having received the new information or a reminder of already existing knowledge). For both, especially for the sender, there also is the henceforth free availability of the representation. One may even take into account the social level, in connection with the internet etiquette and any transmission/publication regulations issued by the responsible authorities. Certain cost and revenue features may be associated with these, as well.

In Table 2, these considerations have been combined, in particular by itemising and further detailing the cost and revenue factors. Table 1 only hinted at these in a qualitative fashion. In this way, a basis for actual

quantification is obtained. Naturally, completeness requires more intensive study than provided by the examples shown here.

SEMIOTIC LEVEL	ASPECT	ECONOMICS (COST)	ECONOMICS (REVENUE)
Social level	Recording: information[a] E-message: regulations	Sender: Internet infrastructure Receiver: confusion, distraction by "over-information"	None to society, as such Potential information, to others
Pragmatics	Recording: (re)usability Printout: information[a] E-message: information[a]	Sender: free (future use) Receiver: study, comprehension	Sender: potential reaction Receiver: ability to react
Semantics	Printout: contents E-message: contents	Sender: composition Receiver: interpretation Sender: composition Receiver: interpretation	Receiver: ability to comprehend Receiver: ability to comprehend
Syntactics	Extract: data language E-message: protocol	Sender staff: education Sender staff: education	Directly, none[b] Directly, none[b]
Empirics	Recording: security, etc. E-message: efficiency, security	Database software Communication software	Directly, none[c] Directly, none[c]
Physical level	Recording: computer memory E-message: bit stream	Hardware Network connection	Directly, none[c] None

Table 2 : Aspects of the economics associated with having available a recording of some 'knowledge', extracts of which are disseminated in printed form (say, a brochure), or e-mailed as promotional message

[a] Potential information (intended by sender)
[b] Indirectly, staff may undertake similar tasks without further training
[c] Indirectly, usable in other applications without further expense

This kind of analysis, based on the semiotic ladder approach, may easily be extended to other kinds of ICT-related services and products. The different levels permit separating out concepts such that the personal, organisational and social relevance of each becomes clear. This contrasts with the partial (almost one-dimensional) modelling approaches encountered in the older literature [notably the classical 'value-of-information' studies of Stigler (Stigler 1961), the theoretical 'removal-of-uncertainty' analysis of Marshak (Marshak 1963) and the pragmatic 'statistical-outcome' analysis of Emery (Emery 1969)]. More recently, the attention has focussed on overall assessments of costs and benefits of ICT investment (Strassman). In our view, all may benefit from the decomposition permitted by considering aspects at separate levels, while maintaining a coherent framework.

In the above examples, there is no fundamental qualitative difference between the traditional and electronic ways of doing business - but a

quantitative one is rather likely. Obviously, this may not always be so in other cases, but the likelihood is that, using a framework such as FRISCO, similarities will outweigh differences. In any event, the conceptual clarity will assist establishing essential aspects.

4. INFORMATION PERSISTENCE AND KNOWLEDGE EROSION

It is stating the obvious when one remarks that "*more information is not necessarily better information***"**. This must be true under any definition of the concept 'information'. Quantification in a naïve way is just not on. More specially selected data does not guarantee per se a better basis for decision making or for the essential answer to some specific question, since the selection criteria may not be straightforward or be affected by unexpected situations. The accumulation of more data, as such, does not automatically result in the desired increase of knowledge, especially if it is not clear a priori what it is one aspires to. We found that a fundamental semi-quantitative analysis of this problem area can benefit from a 'semiotic ladder' similar to that of part 3. However, it is more difficult, in that the 'values' in question are even harder to associate with specific levels. Before addressing the most serious problem – that of the degradation (or 'erosion') of recorded knowledge – we will attempt relating the quality and quantity of information in a general sense.

The *quality* of information offered by a certain amount of data has no correlation to its *quantity*. Take, for example, the wish to buy a 'good' book. Book stores are full of titles, among which one may be struck by something stimulating (the result of 'serendipity'[14]). It is possible, but unlikely, that one would hit on an ideal book just by browsing, over a given amount of time, in a megashop. The same applies when one accesses internet booksellers. On the other hand, a smaller, specialised shop may be precisely the environment in which that inspiring title stands out. Admittedly, this is what the internet shop tries to provide by signalling a few new books of probable interest to a regular 'visitor' on the basis of earlier purchases or enquiries, but that kind of offering cannot compete with the atmosphere and extent of 'information' available at the specialised place.

Generalising, we may conclude that it is difficult to formulate a query for a complex or vaguely specified situation and that bulk offerings ('more information') do not by themselves contribute to the solution and, in extreme

[14] The often quoted "faculty of making happy and unexpected discoveries by accident" (after *The three Princes of Serendip* by Horace Walpole (1717-1797)

cases, may even obscure it. Further assistance by filtering the information provided is needed. We shall revert to this point.

"*Not being able to see the wood for the trees*" is a way of saying that too much data may be a burden. It applies especially to large knowledge bases, which do not necessarily become more valuable the more data and facts are recorded. The advanced techniques of 'data warehouses' and 'data mining' (another buzz word, referring to an extension of the toolset formerly covered by 'information retrieval') try to deal with this problem, often with remarkable success (Chaudiri and Dayal 1997), (Han and Kamber 2000). They suggest that the structuring and accessibility of computerised storage require creative, often 'lateral' thinking [cf. e.g. (Van de Riet 1998)]. This still leaves open the question whether one is best served by large data collections (the megastore) or by specialised support sets (the intimate bookstore). Much depends on the nature of the questions a person or an organisation wishes to answer over time. We shall restrict ourselves to the following three areas:

1. *Decision support*

Decisions are time-dependent and often time-critical. At some point in time a decision has to be made (an investment after consideration of all options including an assessment of the estimated economic outcome). The nearer the moment of action the 'better' the information produced by be the support system might be, but there may be a cut-off moment (closing of a bid, need to buy raw material before it deteriorates, departure of one's train). Sometimes there is no sharp cut-off, but arriving later in a market may adversely affect or even ruin one's position. Clearly, there is a trade-off between timeliness and amount of relevant data.

2. *Administrative control*

Administrations are often originally required to record certain financial data for the purpose of accountability (financial and trading accounts). These records (representing past history) usually are seen as a source of potential information. Thus, they may be used as input to forecasting and planning models. While the past may influence the future in many ways, many new options may arise that do not have the same background as the situations in which the earlier activity took place. Specifically, cost control may well look back to past experience, but revenue forecasts should be based on external data. More internally collected data only clutters up one's records.

3. *Reassurance*

There is a special category of knowledge, which may be available all along, but needs to be made available at critical (decision) moments. A typical example is that of a road map (representing travelling knowledge), which not merely provides the information needed to decide when to turn off

a road, but also reassures the traveller that no special action is required (just carry on) until the decision point arrives. More generally, a data collection may represent a large amount of useful knowledge of which the decision maker, in principle, only wishes to be warned at a critical time ('need-to-know'). A more relaxed working atmosphere will result when some redundant recall from time to time reassures one that no action is needed yet.

In all of these, there are at least four problems:

(a) There is too much – often ill-structured - data or one is tempted to add too much ('over-information').

(b) There may be a lot more 'out there', but one does not know where to look, how to find it.

(c) The data held is becoming obsolete (no longer relevant and in need of refreshment, replenishment).

(d) There is a lot of valuable data available, but the store can only deal with the more recent additions.

Case (a) may be solved by appropriate restructuring (specialised stores, 'data-warehouses') and/or modern data-mining techniques (such as the insertion of hyperlinks or the application of special input-filters).

Case (b) is a call for 'serendipity', which is by definition insolvable; however, broad scanning, using appropriate input-filters may be helpful.

Case (c) often applies and may be treated in two ways, (1) just keeping the contents up to date and ignoring past data ('current-data' files), or (2) always applying a time stamp to new additions ('historical' files). The first solution suits decision making and control in a stable environment, but is open to the danger of unintended loss of important data. The second one is essential wherever a dynamic view is to be taken (however, the danger remains that mere extrapolation of data leads to incorrect conclusions).

Case (d) is very serious from the point of view of any historian. Society is always in danger of loosing track of its past because archives are too full or poorly accessible. This is the weakness of a developing storage technology. What has been punched into cards is only accessible if one possesses properly functioning punched card equipment, just like early magnetic recordings (wire, first-generation tape recordings) and optically stored images (microfiches, 19th century photographs and films) are hardly accessible today, let alone in the future. However, the problem is not restricted by technology. It also applies to a person (personal knowledge) and an organisation or even a whole society (shared knowledge), where much may be 'forgotten', literally or metaphorically. We may categorise these problems under the single heading 'data and knowledge erosion'.

4.1 Data and knowledge erosion

Once again, the semiotic ladder may help analysing this new problem area. Loss and/or inaccessibility of data, hence 'erosion' of knowledge, manifest themselves in a variety of ways. We shall not repeat the full economic analysis, but merely point out the most important conceptual aspects. These are listed in table 3 along with the concepts discussed in the previous section. Once again, the entries are meant to be illustrative rather than definitive.

SEMIOTIC LEVEL	ASPECTS SUBJECT TO EROSION	EROSION CHARACTERISTICS
Social level	Human culture, tradition, standards, survival techniques	Lack of education, staff/finance problems, lack of referencing to sources, loss of "common sense"
Pragmatics	Data collections Human memory	Different or incompatible versions of data & software Forgotten human skills and abilities, (sub)conscious suppression of knowledge
Semantics	Data bases, programming systems	Incompatible, unexplained data, non-executable or incomprehensible programs
Syntactics	Programs, data, historical recordings (e.g. hieroglyphs, early films)	Lost compilers, unavailable platforms, cryptic messages
Empirics	Code, recorded data	Code corruption, decoding failures (e.g. lost keys)
Physical level	Storage, carrying media, memory (machine and human	Degradation and/or noise, illness and age

Table 3 : Data and knowledge erosion analysis by semiotic level

Physical level: Erosion of the carrying media (by degradation or 'noise') is an evident danger. This applies to any storage device, including the 'inaccessible' human mind.

Empirical level: Here, encoding patterns are discussed. These may be addressed both in the context of code corruption (e.g. an impossibility of detection in spite of parsing capability) and of recording overload (e.g. impossibility of recognition for lack of sufficient parsing power). Decoding, of course, becomes very difficult or even impossible when the key to an otherwise uncorrupted code is lost.

Syntactic level: This arises e.g. when the necessary 'reading' tools are no longer available (i.e. resulting in a loss of parsing capability). Examples of

such tools are ‘forgotten’ programming languages, that is to say, situations where the required compilers or even the appropriate platforms no longer exist. Classical examples are the scripts of ancient languages, which can only be deciphered by extensive comparative research and speculation (hieroglyphs, linear-B).

Semantic level: Typical examples of semantic incompatibility are data collections recorded on different storage systems (which make combined usage mostly meaningless). Similar cases are: diagrams with missing legend, databases without domain names (or references thereto), ICT products (devices or programs) without user instructions or "help" functions.

Pragmatic level: A recurrent problem arises with different versions of data collections (analogous to the semantic incompatibility problem mentioned above). Likewise, there are the cases of truly forgetting remembered things, subconsciously (or consciously) suppressing knowledge, and so on. A similar problem occurs with human skills (e.g. loosing the ability to repair and maintain products of everyday use where it is "cheaper" to buy new things)

Social level: There are many instances of erosion of knowledge in a societal context, for instance due to insufficient education or a loss of tradition, to changes of standards, shortage of staff and finance for archiving services, etc. In the same category, elementary survival techniques may be lost as a result of missing joint memory refreshing exercises or failure to refer to historical sources.

In conclusion, it may be said that the erosion problem is pervasive, but may have many different causes. Each may apply individually in some practical situation, but often, a combination of effects applies. These may be best understood (and hopefully combated) with reference to a conceptual framework which brings out and illuminates the relevant aspects.

5. CONCLUSION

There is no direct relationship between the frequency of receiving (or being repeatedly bombarded by) a message and its importance (or its "value") for the receiver. In the age of e-commerce (and implied: the age of ubiquitous promotion and advertising), this even turns into the opposite: good and valuable information becomes rarer and rarer! On the other hand, the rarity of a piece of "information" does not guarantee its quality! Thus we should carefully examine myths such as "the more information - the better!", but also "the more exotic the information - the better!". Furthermore, we should be aware of the many levels at which recorded and transmitted knowledge is in danger of erosion.

In the e-age, we will not *a priori* be better informed. We certainly should look for even better filter mechanisms (technical and intellectual ones!) and improved skills for "finding needles in haystacks".

REFERENCES

Braun, H. von, Hesse, W., Andelfinger, U., Kittlaus, H.B. & Scheschonk, G. (2000) Conceptions are social constructs - Towards a solid foundation of the FRISCO approach, in Falkenberg, E.D., Lyytinen, K. & Verrijn-Stuart, A.A. (2000) (eds.): *Information System Concepts - An Integrated Discipline Emerging*. Kluwer Academic Publishers, Boston, pp. 61-73.

Chaudhuri, S. & Dayal, U. (1997) An Overview of Data Warehousing and OLAP Technology, *SIGMOD Record* **26** (1), 65-74.

Emery, J.C. (1969)*Organizational Planning and Control Systems*, Macmillan.

Falkenberg, E., Hesse, W., Lindgreen, P., Nilsson, B.E., Oei, J.L.H., Rolland, C., Stamper, R.K., Van Assche, F.J.M., Verrijn-Stuart, A.A. & Voss, K. (1998) *FRISCO - A Framework of Information System Concepts - The FRISCO Report*. IFIP WG 8.1 Task Group FRISCO. Web version: ftp://ftp.leidenuniv.nl/pub/rul/fri-full.zip.

Han, J. & Kamber, M. (2000) *Data Mining: Concepts and Techniques*. Morgan Kaufmann .

Hesse, W. & Verrijn-Stuart, A.A. (2000) Towards a Theory of Information Systems: The FRISCO Approach, in Kangassalo H., Jaakkola H. & Kawaguchi E. (2001) (eds.), *Information Modelling and Knowledge Bases XII*. IOS Press, Amsterdam, pp. 81-91.

Marshak, J. (1963) The payof-relevant description of states and acts, *Econometrica* **31** 719-725.

Shannon, C.E. & Weaver, W. (1949) *The mathematical theory of communication*, University of Illinois Press, Urbana (reprinted 1976).

Stigler, G.J. (1969) The economics of information, *Journal of Political Economy* **69**, 213-285.

Strassman, P.A. (1990) *The business value of computers*, The Information Economics Press, New Canaan, Connecticut.

Van de Riet, R.P. (1998) Complex objects: Theory and practice from a data- and knowledge engineering perspective, as seen from Yellowstone Park, *Data & Knowledge Engineering* **25**, 217-238.