# A definition of information

*A.D. Madden*
*JFS, Learning Resources Centre, 175 Camden Road, London NW1 9HD*
*admadden@hotmail.com*

One difficulty faced by students on many information management courses is the lack of any attempt to teach concepts of information. Therefore, if a core module does not fit in with a student's existing concept of information, it can make it hard for the student to recognise the relevance of that module. This paper addresses that problem by summarising concepts of information, and by presenting a simple model that attempts to unite the various concepts listed. The model is based on the idea that the meaning in a message depends on the context in which the message originated (the authorial context), and the context in which it is interpreted (the readership context). Characteristics of authors, readers and messages are discussed. The impact of the 'knowledge' of 'information' users, and of their community, is considered. Implications of the model are discussed. A definition of information is suggested, which attempts to encapsulate the nature of information implied by the model.

## Introduction

Following a return to full-time education after several years in scientific research, the author recently completed a MSc in Information Management. As student representative on the teaching committee the author was able to gain insights into some of the difficulties experienced by the course organisers. A disproportionate number of complaints were received about one module in particular. The module in question, though no more intellectually demanding than any other, was technically more demanding. It was a core module, so students had to take it, and many resented the fact. One common criticism was that the relationship of the module to the course as a whole was unclear. It became apparent that the issue of whether or not the module was appropriate was dependent on a student's concept of information. It was also clear that concepts of information were not being taught on the course. Discussion with students on courses elsewhere in the UK suggested that this is not unusual.

The paucity of theory in information science has often been commented on [1, 2], and has led to a pragmatic approach to both the teaching and practice of information science. The resulting piecemeal view of the subject has, not surprisingly, led to problems of the kind described above. The purpose of this short paper is to explore concepts of information that have been suggested, and to propose a definition that encapsulates them. It is not intended to present new ideas, but rather to consolidate existing ideas in a way that makes them easy to put across to students. The complexity of ideas concerning information is such that a paper this length can scarcely do them justice, but it is intended that this paper should provide a starting point to which students can relate modules as diverse as document retrieval and systems modelling.

## What is information?

Attempts to answer the question 'What is information?' have, not surprisingly, occupied the thoughts of information scientists for a long time: almost certainly since before the term 'information science' was coined in 1955 [3]. The lay person, asked to define information, is most likely to regard it as:

> An item of information or intelligence; a fact or circumstance of which one is told. (OED)

This is just one of the many dictionary definitions of the word. Indeed, information scientists appear to have been reluctant to propose definitions of information, preferring

rather to discuss concepts: the difference being, according to Belkin [4, p. 58], a definition 'says what the phenomenon defined is, whereas a concept is a way of looking at or interpreting the phenomenon'.

In their recent paper, McCreadie and Rice [5] review concepts of information proposed over the last fifty years. A summary of the concepts they consider is given below.

- **Information as a representation of knowledge**
  Information is stored knowledge. Traditionally the storage medium has been books, but increasingly electronic media are becoming important.
- **Information as data in the environment**
  Information can be obtained from a range of environmental stimuli and phenomena; not all of which are intended to 'convey' a message, but which can be informative when appropriately interpreted.
- **Information as part of the communication process**
  Meanings are in people rather than in words or data. Timing and social factors play a significant role in the processing and interpretation of information.
- **Information as a resource or commodity**
  Information is transmitted in a message from sender to receiver. The receiver interprets the message as intended by the sender. There may be added value as the information is disseminated or exchanged.

## Information in context

The model presented below rests on the assumption that information cannot be evaluated without an awareness of the context in which it is being interpreted. This assumption leads to a model comprising three components.

***1. Readership context***
The context in which a message is received and interpreted. The reader is any system which derives (or attempts to derive) information from a message. A system may be a mechanism, an organism, a community, or an organisation.

***2. Authorial context***
The context in which the message originates. The author is any system that transmits (intentionally or otherwise) a message from which a reader can derive information.

***3. Message***
The means by which information is transmitted. It may be written, spoken, facial expression, pheromonal, etc.

These components are described more fully below.

### *Information as data in the environment: reading the signs*

Many information scientists accept that information is a property of all living organisms [6, 7]. It is not unreasonable, therefore, to illustrate the prime importance of context with an example from biology. It has long been recognised by zoologists that there is an association between brightly coloured markings on an animal and unpalatability or toxicity [8]. Predators learn to associate such markings with unpleasant experiences and so are less inclined to attack similarly marked creatures in future. So a potential predator, seeing the markings of a brightly coloured male butterfly, will derive information about the insect's suitability as a food source. A female butterfly of the same species will derive no such information. She might, however, find the markings useful in assessing the male's quality as a mate. Clearly, therefore, both the predator and the female butterfly derive information from the markings, but the message of the markings depends on the context in which it is read.

The importance of context in the interpretation of information has long been recognised. The above example, however, indicates that, unless it is assumed that butterflies and birds have knowledge, knowledge is not necessary for a signal to be informative.

### *Information as part of the communication process: the authorial and readership contexts*

Messages exchanged between humans frequently differ from those discussed in the

two examples above because there is often an intention that they should carry information. There is no reason to assume that a male butterfly means to inform either a predator or a potential mate when he flaps his wings. By contrast, a message designed to communicate has two informing contexts: that of the author and that of the reader. These correspond to the two points described by Shannon and Weaver [9, p. 31] when they stated that 'The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.' However, Shannon and Weaver were merely talking about the difficulties involved in transmitting a signal from a sender to a receiver. If that signal is to be a message, it is necessary for the sender to be an author, or the recipient to be a reader, or both. Characteristics of these two contexts, and the message itself, are described below.

***The readership context***

As Meadow and Yuan noted, 'Most views of the difference between data and information … depend on the recipient.' [10, p. 701] The information derived from a message by a reader depends on a wide range of factors, all of which affect the reader's understanding of that message. Some of these are listed below:

- Geographical – nation, culture, language, physical community.
- Social – interests, pastimes.
- Educational – level of education, subjects studied.
- Professional – area of professionalism, career history.

The different contexts overlap. A mathematical treatise will be understood in the same way by both Russian and American mathematicians. A Birmingham newspaper will be more informative to Jamaican and Punjabi immigrants living in Solihull than it would to a tenth generation cockney in Lambeth. To understand what makes sense to a reader, therefore, it is necessary to understand the structure of the society of which he or she is a part [11].

***The authorial context***

As well as sharing the characteristics of the readership context, the authorial context has an additional property: that of intention. Two possible states of intention are assumed:

*Message intended to convey information*

The author produces the text with the intention of informing the reader. This is the usual authorial context, in which a text 'is a collection of signs purposefully structured by a sender with the intention of changing the image-structure of a recipient' [12, p. 20]. The closer an author's context is to that of a reader, the greater is the chance that the author's work will be informative. In exceptional circumstances, an author may choose to convey more than one message in a given text (see Appendix).

*Message not intended to convey information*

The author ascribes no meaning to the message of the text: any meaning is derived within the readership context. Examples include the predictions of fortune-tellers and output from artificial intelligence programs such as Eliza [13].

## Information as a resource or commodity: getting the message

Messages are classified according to how focused they are. Traditionally, information scientists have dealt largely with focused messages. Highly focused messages are ones in which the context for interpretation is very specific, making ambiguities difficult or impossible. The most obvious example would be a mathematical document, but other examples include command line computer interfaces and technical publications.

Looser (but still focused) messages would include descriptive works and histories, which will be interpreted according to the reader's culture and experience. A less positive example would be poorly written documents [4], which may be confusing, ambiguous, or misleading: a typical contemporary example of this is email. Totally loose messages would have no obvious interpretation in any context. An example is surrealist literature. In many ways this model is similar

to the communication model proposed by Jakobson [14]. This too comprises three components (addresser, addressee, and message). Because it deals with the intentional transfer of information, however, it places greater emphasis on the means by which messages are transmitted, and excludes a great deal of information sources.

## The importance of recognising context

Wilson [3] stresses the importance of context in dictating information needs, but makes no mention of the impact of context on the interpretation and effectiveness of information materials. Hjørland considers 'subject analysis of documents as one of the most fundamental activities of library and information professionals' [2, p. 610].

One implication of the model presented here is that such an analysis requires an appreciation of the context. As Hjørland notes:

> The subject of a book (or any other document, or message) is closely related to what kind of answers people can find from reading the book... Any document thus has an infinite number of subjects [2, p. 610].

So to predict the effect of information on a particular user, it is necessary first to envisage the potential user. In classifying information, therefore, the information scientist is implicitly classifying the user.

## Information and knowledge

Earlier, the importance of context in causing a stimulus to become informative was discussed, but arguably the most fundamental context was omitted: that of the knowledge of the recipient of the information. The examples of contexts listed above will all shape that knowledge, hence their significance, but in addition the knowledge will be affected by an incalculable array of experiences and aptitudes.

The association between knowledge and information seeking is well established: '...the idea of using cognitive models as the basis for information retrieval system design has aroused considerable interest...' [15, p. 63].

It has been argued that what motivates someone to seek information is a recognition by the seeker of 'an anomaly in his/her state of knowledge' [4, p. 81]. Moser [16, p. 350], questions how 'normal' and 'anomalous' states of knowledge are to be identified or measured, and makes the point that 'information, to be generated, need not actively be instigated on the 'recipient's' side'.

This observation is highly relevant to the context-reliant model of information reception described in this paper. Checkland argues that 'consciousness makes man, via his W(orld View)s, a meaning-endowing animal' [17, p. 219]. If this is accepted, then it is the meaning endowed within the World View that will determine whether information is sought, what information is sought, and how it is interpreted. Anomalies may be a motivating factor. People who, unlike Lewis Carroll's 'Humpty Dumpty', are not happy to 'believe six impossible things before breakfast' may seek information in an effort to resolve some of the discrepancies in their World View.

Alternatively, however, the meaning conferred by a World View may provide someone with a paradigm of 'normality', which can be used in the generation of hypotheses. Here, information may be sought to test the hypotheses in order to establish or to extend 'normality'.

## Personal paradigms as context

The word 'paradigm' above is used in its dictionary sense: 'pattern, example, to exhibit beside, show side by side' (OED) rather than in the ways in which Kuhn [18] used the term in his philosophy of science. While Kuhn used 'paradigm' to describe systems by which meaning could be shared in a research community, the word is used above to describe the ways in which an individual organises information within his or her World View.

The difference is significant when considering the evaluation of information. Information scientists can only assess information insofar as their World Views match those of the people for whom they are evaluating it. It is because the match is inexact that browsing and serendipity are important

factors in information seeking. Data and texts that appear irrelevant to an evaluator may provide the missing piece of a puzzle to a researcher; but the data and texts will only be informative if the puzzle is known.

Probably the best known example in science of a serendipitous discovery arose because of the problem of King Hieron's crown. The king, wishing to know whether the crown was pure gold as claimed by the goldsmith, or whether a gold/silver alloy had been used, asked Archimedes to investigate. Archimedes is reputed to have arrived at the solution when, as he climbed into his bathtub, he observed water overflowing from it. Within the context of his knowledge, his thoughts, and his ideas, the stimulus of overflowing water was informative. Archimedes deduced that the quantity of water displaced was equivalent to the volume of his body, and so had a means of determining the density of the crown [19].

The history of science is full of such tales: from the apple that gave rise to Newton's thoughts on gravity, to the dream of snakes from which Kekule derived the structure of benzene. Such examples, however, are of little practical relevance to the information scientist, since information of this kind is impossible to organise. A more constructive and more recent example involves the work of Heisenberg in quantum mechanics. According to C.P. Snow, in the early 1920s, Heisenberg was seeking to find mathematical tools which would enable him to relate the set of rules associated with any given atom to that atom's set of properties.

> The trouble was, he didn't know enough of the curiosities of nineteenth- century mathematics, when all kinds of mathematical arts had been developed. Not for use, but for the sheer beauty of the game.
> Fortunately... Max Born ... [knew of the] old subject of matrix algebra, half forgotten but completely available [20, p. 67].

This half-forgotten branch of mathematics proved to be 'precisely what they needed'. If Snow's analysis is correct, however, and matrix algebra had indeed been developed for 'the beauty of the game' rather than for use, it would presumably have been held to have little informative value. Despite this, the text was 'completely available' and clearly retrievable; and within the context of Heisenberg's research it became invaluable information.

## Information in the community

Although the knowledge of the information user may be the ultimate informing context, from the point of view of an information manager wishing to provide relevant information it is probably an impractical starting point. It is for this reason that information specialists tend to deal with information as a representation of knowledge, or 'information-as-thing' [21].

As was stated above, however, community is also an important context. For the purposes of the information manager, this is usually the context by which information is defined. What is stored in collections of informative things, whether those collections be archives, libraries, or digitised records, is stored with a view to the needs of the target community.

The idea that information is embedded in socio-cultural contexts is not new of course [22]. What has changed for the information professional, however, is the extent to which the nature of the community being served must be considered. The traditional librarian was usually a professional, catering for other professionals educated to a similar level, and therefore well able to anticipate their requirements. In the case of public libraries, the user community defined itself by its desire to use the library. This is still true, but given the ever-expanding range of alternative sources of information and entertainment, this community is declining.

## Memes and the transfer of information

As was stated at the start of this paper, information is widely regarded as being a property of living organisms. Dennett in particular, stresses the connection between information and awareness in many forms of life [23], and discusses the impact of information on

consciousness. In so doing, he draws heavily on some of the ideas proposed by Dawkins [24] in *The Selfish Gene*. Here, Dawkins argues that a lot of ideas are reproduced in human society in a manner analogous to genetic replication. He refers to such ideas as memes, and the resulting study (mimetics) has begun to gain acceptability. Mimetics draws heavily on comparisons with biological evolution, and the succession of overlapping contexts described above bears a similarity to Hutchinson's classic definition of an ecological niche [25] as an n-dimensional hypervolume: a mathematically defined space in which each of the factors affecting the viability of an organism occupying that niche is seen as a separate dimension.

It has been commonly observed that 'Almost always the men who achieve... fundamental inventions of a new paradigm are either very young or very new to the field whose paradigm they change.' [18, p. 90]

If the ideas presented above on community as a informing context are 'linked' to those on World View, the possibility is raised that what is learned in one community will, in the context of a different community, be informative in ways that were not previously recognised. To extend the evolutionary analogy use in mimetics, this perhaps represents a cross-fertilisation of ideas.

## Conclusion

As has been argued, the materials with which the information scientist routinely works represent just a small proportion of potential information. This paper therefore proposes that, to capture the 'breadth' of possible information sources, information should be defined as:

> a stimulus originating in one system that affects the interpretation by another system of either the second system's relationship to the first or of the relationship the two systems share with a given environment
> (where a system is as defined above, in Readership Context).

Brown [1, p. 185] suggests that, in attempting to define information, information scientists have tended to restrict the term to only a part of the whole and that 'no satisfactory concept of information for information science will ever be formulated in the sense of supplanting all others'.

The model described above represents an attempt, not to supplant concepts of information, but to unite them. The idea that information is only information in certain 'informing contexts' incorporates nearly all widely held concepts of information, and would provide a useful point at which to introduce students to the range of directions in which information science could take them.

## Appendix

Deliberate ambiguities are common in puzzles, codes, and as literary devices. Simple codes may carry both overt and cryptic messages. For example, the message:

> Coming in tomorrow evening. Meet at
> Dinnington depot, eleven nineteen.

would convey to most readers the information that a liaison is being requested. Those able to apply an appropriate geographical context will derive additional information concerning the location. This instruction may be relevant to the intended recipients of the message, but they would also receive further information by being aware that they should read the initial letters of the words. Punning headlines are commonly used to summarise two aspects of a newspaper story. A fictional example would be that of John Smith, an aspiring rock star arrested for assault. A report on his chart success and his appearance in a police identity parade may be headlined: 'SMITH IN HIT PARADE'. Such exercises need not be limited to one language. A French-speaking reader of the book *Mots d'Heures: Gousses, Rames* (Verse 11) [26] will struggle to understand esoteric and surreal verse such as:

> *Chacun Gille*
> *Houer ne taupe de hile*
> *Tôt-fait, j'appelle au boiteur*
> *Chaque fêle dans un broc, est-ce crosne?*
> *Un Gille qu'aime tant berline à fêtard.*
> (Verse 11)
> *(Every bumpkin*
> *While hoeing uncovers a mole and*

*part of a seed.*
*Quickly finished, I call to the limping man that*
*Every pitcher has a crack in. It is it a Chinese cabbage?*
*A bumpkin loves a life of pleasure and a carriage.)*

An English listener, however, hearing the verse read aloud in French, will find it 'hauntingly familiar'.

## References


1. Brown, A.D. *Towards a theoretical information science: information science and the concept of a paradigm.* (Department of Information Studies, Occasional Publication No. 5) Sheffield: University of Sheffield, 1987.

2. Hjørland, B. Theory and metatheory of information science, *Journal of Documentation,* 54 (5), 1998, 606-621.

3. Wilson, T.D. Models in information behaviour research. *Journal of Documentation,* 55(3), 1999, 249-270.

4. Belkin, N.J. Information concepts for information science. *Journal of Documentation,* 34(1), 1978, 55-85.

5. McCreadie, M. and Rice, R.E. Trends in analyzing access to information. Part I: cross-disciplinary conceptualizations of access. *Information Processing and Management,* 35(1), 1999, 45-76.

6. Debons, A. and Horne, E. NATO advanced study institutes of information science and foundations of information science. *Journal of the American Society for Information Science*, 48 (9), 1997, 794-803.

7. Vickery, B. Metatheory and information science. *Journal of Documentation*, 53(4), 1997, 457–476.

8. Ruxton, G. Sheep in wolves' clothing. *Nature*, 394, 1998, 833–834.

9. Shannon, C.E. and Weaver, W. *The mathematical theory of communication*. Chicago: University of Illinois Press, 1949.

10. Meadow, C.T. and Yuan, W. Measuring the impact of information: defining the concepts. *Information Processing and Management*, 33(6), 1997, 69-714.

11. Dervin, B. On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing and Management*, 35(6), 1999, 727–50.

12. Belkin, N.J. and Robertson, S.E. Information science and the phenomena of information. *Journal of the American Society of Information Scientists*, 27(4), 1976, 197–204.

13. Allen, J. *Natural language understanding*. 2nd edition Redwood City, CA: Benjamin/ Cummings, 1995.

14. Jakobson, R. Closing statement: linguistics and poetics. In: Sebeok, T., ed. *Style and Language*. Cambridge, Mass.: MIT, 1960, 350–377.

15. Ellis, D. *Progress and problems in information retrieval.* 2nd edition London: L.A. Publishing, 1998.

16. Moser, A. Information concepts. *Journal of Documentation*, 34(3), 1978, 350–1.

17. Checkland, P.B. *Systems thinking, systems practice*. 2nd edition Chichester: Wiley, 1984.

18. Kuhn, T.S. *The structure of scientific revolutions*. 3rd edition Chicago: University of Chicago Press, 1996.

19. Ronan, C.A. *The Cambridge illustrated history of the world's science.* Cambridge: CUP, 1983.

20. Snow, C.P. *The physicists: a generation that changed the world.* London: Macmillan, 1981.

21. Buckland, M.K. Knowledge as thing. *Journal of the American Society for Information Science*, 42(5), 1991, 351–360.

22. McGinn, C. *Wittgenstein on meaning: an interpretation and evaluation.* Oxford: Blackwell, 1984.

23. Dennett, D.C. *Consciousness explained*. London: Penguin, 1993.

24. Dawkins, R. *The selfish gene*. Oxford: Oxford University Press, 1989.

25. Hutchinson, R.E. Concluding remarks. Cold Spring Harbor. *Symp. Quant. Biol*. 22. 1957, 415-427.

26. Van Rooten, C.H.K., *Mots d'Heures: Gousses, Rames.* UK: Angus & Robertson Ltd, 1968.