Information as Thing

Michael K. Buckland
School of Library and Information Studies, University of California, Berkeley, CA 94720

Three meanings of "information" are distinguished: "information-as-process"; "information-as-knowledge"; and "information-as-thing," the attributive use of "information" to denote things regarded as informative. The nature and characteristics of "information-as-thing" are discussed, using an indirect approach ("What things are informative?"). Varieties of "information-as-thing" include data, text, documents, objects, and events. On this view "information" includes but extends beyond communication. Whatever information storage and retrieval systems store and retrieve is necessarily "information-as-thing."

These three meanings of "information," along with "information processing," offer a basis for classifying disparate information-related activities (e.g., rhetoric, bibliographic retrieval, statistical analysis) and, thereby, suggest a topography for "information science."

Introduction: The Ambiguity of "Information"

An exploration of "information" runs into immediate difficulties. Since information has to do with becoming informed, with the reduction of ignorance and of uncertainty, it is ironic that the term "information" is itself ambiguous and used in different ways. (For a concise and convenient introduction to varieties of meanings of "information" and some related terms see Machlup (1983). See also Braman (1989), NATO (1974, 1975, 1983); Schrader (1983), Wellisch (1972), Wersig and Neveling (1975)). Faced with the variety of meanings of "information," we can, at least, take a pragmatic approach. We can survey the landscape and seeking to identify groupings of uses of the term "information." The definitions may not be fully satisfactory, the boundaries between these uses may be indistinct, and such an approach could not satisfy anyone determined to establish the one correct meaning of "information." But if the principal uses can be identified, sorted, and characterized, then some progress might be made. Using this approach we identify three principal uses of the word "information:"

(1) *Information-as-process*: When someone is informed, what they know is changed. In this sense "information" is "The act of informing...; communication of the knowledge or 'news' of some fact or occurrence; the action of telling or fact of being told of something" (*Oxford English Dictionary,* 1989, vol. 7, p. 944).

(2) *Information-as-knowledge*: "Information" is also used to denote that which is perceived in "information-as-process:" the "knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told; intelligence, news" (*Oxford English Dictionary,* 1989, vol. 7, p. 944). The notion of information as that which reduces uncertainty could be viewed as a special case of "information-as-knowledge." Sometimes information increases uncertainty.

(3) *Information-as-thing*: The term "information" is also used attributively for objects, such as data and documents, that are referred to as "information" because they are regarded as being informative, as "having the quality of imparting knowledge or communicating information; instructive." (*Oxford English Dictionary,* 1989, vol. 7, p. 946).

A key characteristic of "information-as-knowledge" is that it is intangible: one cannot touch it or measure it in any direct way. Knowledge, belief, and opinion are personal, subjective, and conceptual. Therefore, to communicate them, they have to be expressed, described, or represented in some physical way, as a signal, text, or communication. Any such expression, description, or representation would be "information-as-thing." We shall discuss implications of this below.

Some theorists have objected to the attributive use of the term "information" to denote a thing in the third sense above. Wiener asserted that "Information is information, not material nor energy." Machlup (1983, p. 642), who restricted information to the context of communication, was dismissive of this third sense of information: "The noun 'information' has essentially two traditional meanings... Any meanings other than (1) the *telling of something* or (2) *that which is being told* are either analogies and metaphors or concoctions resulting from the condoned appropriation of a word that

Received November 14, 1989; revised March 16, 1990; accepted March 29, 1990.

had not been meant by earlier users." Fairthorne (1954) objected scornfully to information as "stuff": "information is an attribute of the receiver's knowledge and interpretation of the signal, not of the sender's, nor some omniscient observer's nor of the signal itself."

But language is as it is used and we can hardly dismiss "information-as-thing" so long as it is a commonly used meaning of the term "information." Indeed, languages evolve and with the expansion of information technology, the practice of referring to communications, databases, books, and the like, as "information" appears to be becoming commoner and, perhaps, a significant source of confusion as symbols and symbol-bearing objects are easily confused with whatever the symbols denote. Further, "information-as-thing," by whatever name, is of especial interest in relation to information systems because ultimately information systems, including "expert systems" and information retrieval systems, can deal directly with information *only* in this sense. The development of rules for drawing inferences from stored information is an area of theoretical and practical interest. But these rules operate upon and only upon information-as-thing.

The purpose of this examination of the notion of "information-as-thing" is to

(1) Clarify its meaning in relation to other uses of the term "information;"
(2) Affirm the fundamental role of "information-as-thing" in information systems; and
(3) Speculate on possible use of the notion of "information-as-thing" in bringing theoretical order to the heterogeneous, ill-ordered fields associated with "information science."

The distinction between intangibles (knowledge and information-as-knowledge) and tangibles (information-as-thing) is central to what follows. If you can touch it or measure it directly, it is not knowledge, but must be some physical thing, possibly information-as-thing. (This distinction may be overstated. Knowledge may well be represented in the brain in some tangible, physical way. However, for present purposes and for the time being, treating knowledge in the mind as importantly different from artificial stores of information seems reasonable and useful. Academic examinations test individuals' ability to answer questions or to solve problems, which is presumed to provide *indirect* measures of what they know. But that is not the same.) Knowledge, however, can be represented, just as an event can be filmed. However, the representation is no more knowledge than the film is the event. Any such representation is necessarily in tangible form (sign, signal, data, text, film, etc.) and so *representations* of knowledge (and of events) are necessarily "information-as-thing."

Information-as-thing is of special interest in the study of information systems. It is with information in this sense that information systems deal directly. Libraries deal with books; computer-based information systems handle data in the form of physical bits and bytes; museums deal directly with objects. The intention may be that users will become informed (information-as-process) and that there will be an imparting of knowledge (information-as-knowledge). But the means provided, what is handled and operated upon, what is stored and retrieved, is physical information (information-as-thing). On these definitions, there can be no such thing as a "knowledged-based" expert system or a "knowledge access" system, only systems based on physical representations of knowledge.

This introductory discussion can be rounded out by reference to a fourth element: information processing, the handling, manipulating, and deriving of new forms or versions of information-as-thing. (One could regard the process of becoming informed as a sort of information processing, but, to reduce confusion, we prefer to separate and exclude mental information-as-process from the scope of "information processing.")

Our discussion thus far can be summarized in terms of two distinctions (1) between entities and processes; and (2) between intangibles and tangibles. Taken in conjunction, these two distinctions yield four quite different aspects of information and information systems. See Fig. 1.1.

A Reverse Approach: What Is Informative?

Instead of the tedious task of reviewing candidate objects and inquiring whether or not they should be considered to be examples of information-as-thing, we can reverse the process and ask people to identify the things by or on account of which they came to be informed. People will say that they are informed by a very wide variety of things, such as messages, data, documents, objects, events, the view through the window, by any kind of evidence. This point was recognized by Brookes (1979, p. 14): "In the sciences it has long been

	INTANGIBLE	TANGIBLE
ENTITY	2. Information-as-knowledge Knowledge	3. Information-as-thing Data, document
PROCESS	1. Information-as-process Becoming informed	4. Information processing Data processing

FIG. 1. Four aspects of information.

recognized that the *primary* source of information is not the literature of the sciences but observation of the relevant natural phenomena. Scientists (and others) find 'sermons in stones and books in the running brooks'." How might we best sort out these candidates for being regarded as information? (Note we are restricting our attention to physical things and physical events. Some people would say that some of their knowledge comes from paraphysical sources, notably from divine inspiration. Others would deny any such nonphysical source of information, but, to the extent that it may exist, information science would have to be incomplete if it were excluded. Not knowing what to say on the subject we merely note it as a possible area of unusual interest within information science.)

Information as Evidence

One learns from the examination of various sorts of things. In order to learn, texts are read, numbers are tallied, objects and images are inspected, touched, or otherwise perceived. In a significant sense information is used as evidence in learning—as the basis for understanding. One's knowledge and opinions are affected by what one sees, reads, hears, and experiences. Textbooks and encyclopedias provide material for an introduction; literary texts and commentaries provide sources for the study of language and literature; arrays of statistical data provide input for calculations and inference; statutes and law reports indicate the law; photographs show what people, places, and events looked like; citations and sources are verified; and so on. In each case it is reasonable to view information-as-thing as *evidence,* though without implying that what was read, viewed, listened to, or otherwise perceived or observed was necessarily accurate, useful, or even pertinent to the user's purposes. Nor need it be assumed that the user did (or should) believe or agree with what was perceived. "Evidence" is an appropriate term because it denotes something related to understanding, something which, if found and correctly understood, could change one's knowledge, one's beliefs, concerning some matter.

Further, the term "evidence" implies passiveness. Evidence, like information-as-thing, does not *do* anything actively. Human beings do things *with* it or *to* it. They examine it, describe it, and categorize it. They understand, misunderstand, interpret, summarize, or rebut it. They may even try to fake it, alter it, hide it, or destroy it. The essence of evidence is precisely that perception of it can lead to changes in what people believe that they know.

Dictionary definitions of "evidence" include: "An appearance from which inferences can be drawn; an indication, mark, sign, token, trace.... Ground for belief; testimony or facts tending to prove or disprove any conclusion.... Information, whether in the form of personal testimony, the language of documents, or the production of material objects, that is given in a legal investigation." (*Oxford English Dictionary,* 1989, vol. 4, p. 469). If something cannot be viewed as having the characteristics of evidence, then it is difficult to see how it could be regarded as information. If it has value as information concerning something, then it would appear to have value as evidence of something. "Evidence" appears to be close enough to the meaning of information-as-thing to warrant considering its use as a synonym when, for example, describing museum objects as "authentic historic pieces of evidence from nature and society." (Schreiner, 1985, p. 27).

One area in which the term "evidence" is much used is in law. Much of the concern is with what evidence—what information—can properly be considered in a legal process. It is not sufficient that information may be pertinent. It must also have been discovered and made available in socially approved ways. However, if we set aside the issues of the propriety of the gathering and presentation of evidence and ask what, in law, evidence actually *is,* we find that it corresponds closely to the way we are using it here. In English law, evidence can include the performing of experiments and the viewing of places and is defined as: "First, the means, apart from argument and inference, whereby the court is informed as to the issues of fact as ascertained by the pleadings; secondly the subject matter of such means." (Buzzard et al., 1976, p. 6; also Wigmore, 1983).

Types of Information

Pursuing the notion of information as evidence, as things from which one becomes informed, we can examine more specifically what sorts of things this might include.

Data

"Data," as the plural form of the Latin word "datum," means "things that have been given." It is, therefore, an apt term for the sort of information-as-thing that has been processed in some way for use. Commonly "data" denotes whatever records are stored in a computer. (See Machlup (1983, p. 646–649) for a discussion of the use and misuse of the term "data".)

Text and Documents

Archives, libraries, and offices are dominated by texts: papers, letters, forms, books, periodicals, manuscripts, and written records of various kinds, on paper, on microform, and in electronic form. The term "document" is normally used to denote texts or, more exactly, text-bearing objects. There seems no reason not to extend the use of "text" and "document" to include images, and even sounds intended to convey some sort of communication, aesthetic, inspirational, instrumental, whatever. In this sense, a table of numbers can be con-

sidered as text, as a document, or as data. Text that is to be analyzed statistically could also be regarded as data. There is a tendency to use "data" to denote numerical information and to use text to denote natural language in any medium.

Further confusion results from attempting to distinguish two types of retrieval by making and compounding two unwarranted assumptions about "data" and "document": (1) that "data retrieval" should denote the retrieval of records that one wishes to inspect and "document retrieval" should denote *references to* records that one may wish to inspect; and (2) that "data retrieval" would be a "known item" search, but that "document retrieval" would be a "subject search" for an unknown item (van Rijsbergen, 1979, p. 2; Blair, 1984). The former assumption imposes an odd definition on both terms. The second is illogical and contrary to practical experience (Buckland, 1988b, pp. 85–87). It is wise not to assume any firm distinction between data, document, and text.

Objects

The literature on information science has concentrated narrowly on data and documents as information resources. But this is contrary to common sense. Other objects are also potentially informative. How much would we know about dinosaurs if no dinosaur fossils had been found? (*cf.* Orna and Pettit (1980, p. 9), writing about museums: "In the first stage, the objects themselves are the only repository of information.") Why do centers of research assemble many sorts of collections of objects if they do not expect students and researchers to learn something from them? Any established university, for example, is likely to have a collection of rocks, a herbarium of preserved plants, a museum of human artifacts, a variety of bones, fossils, and skeletons, and much else besides. The answer is, of course, that objects that are not documents in the normal sense of being texts can nevertheless be information resources, information-as-thing. Objects are collected, stored, retrieved, and examined as information, as a basis for becoming informed. One would have to question the completeness of any view of information, information science, or information systems that did not extend to objects as well as documents and data. In this we, like Wersig (1979), go further than Machlup (1983, p. 645) who, like Belkin and Robertson (1976), limited information to what is intentionally told: "Information takes at least two persons: one who tells (by speaking, writing, imprinting, signally) and one who listens, reads, watches." Similarly Heilprin (1974, p. 124) stated that "information science is the science of propagation of meaningful human messages." Fox (1983) took an even narrower view, examining information and misinformation exclusively in terms of propositional sentences. Brookes (1974), however, was less restrictive: "I see no reason why what is learned by direct observation of the physical environment should not be regarded as *information* just as that which learned by observing the marks on a document." Wersig (1979) adopted an even broader view of information as being derived from three sources: (1) "Generated internally" by mental effort; (2) "Acquired by sheer perception" of phenomena; and (3) "Acquired by communication." We view "information-as-thing" as corresponding to Wersig's phenomena (2) and communications (3).

Some informative objects, such as people and historic buildings, simply do not lend themselves to being collected, stored, and retrieved. But physical relocation into a collection is not always necessary for continued access. Reference to objects in their existing locations creates, in effect, a "virtual collection." One might also create some description or representation of them: a film, a photograph, some measurements, a directory, or a written description. What one then collects is a document describing or representing the person, building, or other object.

What is a Document?

We started by using a simple classification of information resources: data, document, and object. But difficulties arise if we try to be rigorous. What, for example, is a document? A printed book is a document. A page of hand-writing is a document. A diagram is a document. A map is a document. If a map is a document, why should not a three-dimensional contour map also be a document. Why should not a globe also be considered a document since it is, after all, a physical description of something. Early models of locomotives were made for informational not recreational purposes (Minns, 1973, p. 5). If a globe, a model of the earth, is a document, why should one not also consider a model of a locomotive or of a ship to be a document? The model is an informative representation of the original. The original locomotive or ship, or even a life-size replica, would be even more informative than the model. "The few manuscript remains concerning the three ships that brought the first settlers to Virginia have none of the power to *represent* that experience that the reconstructed ships have." (Washburn, 1964). But by now we are rather a long way from customary notions of what a document is.

The proper meaning of "document" has been of concern to information scientists in the "documentation" movement, seeking to improve information resource management since the beginning of this century. The documentalist's approach was to use "document" as a generic term to denote any physical information resource rather than to limit it to text-bearing objects in specific physical media such as paper, papyrus, vellum, or microform. Otlet and others in the documentation movement affirmed:

(1) That documentation (i.e., information storage and retrieval) should be concerned with any or all potentially informative objects;
(2) that not all potentially informative objects were documents in the traditional sense of texts on paper; and
(3) that other informative objects, such as people, products, events and museum objects generally, should not be excluded (Laisiepen, 1980). Even here, however, except for Wersig's contribution (Wersig, 1980), the emphasis is, in practice, on forms of communication: data, texts, pictures, inscriptions.

Otlet (1934, p. 217), a founder of the documentation movement, stressed the need for the definition of "document" and documentation (i.e., information storage and retrieval) to include natural objects, artefacts, objects bearing traces of human activities, objects such as models designed to represent ideas, and works of art, as well as texts. The term "document" (or "documentary unit") was used as a specialized sense as a generic term to denote informative things. Pollard (1944) observed that "From a scientific or technological point of view the [museum] object itself is of greater value than a written description of it and from the bibliographical point of view it should be regarded therefore as a document." A French documentalist defined "document" as "any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental." ("Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter ou de prouver un phénomène ou physique ou intellectual" (Briet, 1951, p. 7)). On this view objects are not ordinarily documents but become so if they are processed for informational purposes. A wild antelope would not be a document, but a captured specimen of a newly discovered species that was being studied, described, and exhibited in a zoo would not only have become a document, but "the catalogued antelope is a primary document and other documents are secondary and derived. ("L'antilope cataloguée est un document initial et les autres documents sont seconds ou dérivés." (Briet, 1951, p. 8). Perhaps only a dedicated documentalist would view an antelope as a document. But regarding anything informative as a "document" is consistent with the origins and early usage of the word, which derived from the Latin verb *docere,* to teach or to inform, with the suffix "-ment" denoting means. Hence "document" originally denoted a means of teaching or informing, whether a lesson, an experience, or a text. Limitation of "document" to text-bearing objects is a later development (*Oxford English Dictionary,* 1989, vol. 4, p. 916; Sagredo & Izquierdo, 1983, pp. 173–178). Even among documentalists, however, including anything other than text-bearing objects in information retrieval appears to occur only in theoretical discussions and not always then (Rogalla von Bieberstein, 1975, p. 12). Meanwhile the semantic problem remains: What generic term for informative things is wide enough to include, say, museum objects and other scholarly evidence, as well as text-bearing objects? Objecting to the use of "information" or of "document" for this purpose does not remove the need for a term.

Most documents in the conventional usage of the word—letters, books, journals, etc.—are composed of text. One would include diagrams, maps, pictures, and sound recordings in an extended sense of the term "text." Perhaps a better term for texts in the general sense of artifacts *intended* to represent some meaning would be "discourse." We could also characterize these texts as "representations" of something or other. However, we could hardly regard an antelope or a ship as being "discourse." Nor are they representations in any ordinary sense. Their value as information or evidence derives from what they signify about themselves individually or, perhaps, about the class or classes of which they are members. In this sense they represent something and, if not a representation, they could be viewed as represent*ative.* If an object is not representative of something, then it is not clear how far it can signify anything, i.e., be informative.

One might divide objects into artifacts intended to constitute discourse (such as books), artifacts that were not so intended (such as ships), and objects that are not artifacts at all (such as antelopes). None of this prevents any of these from being evidence, from being informative concerning something or other. Nor does it prevent people from making uses different from that which may have been intended. A book may be treated as a doorstop. Illuminated initial letters on medieval manuscripts were intended to be decorative, but have become a major source of information concerning medieval dress and implements.

"Natural sign" is the long-established technical term in philosophy and semiotics for things that are informative but without communicative intent (Clarke, 1987; Eco, 1976).

Events

We also learn from events, but events lend themselves even less than objects do to being collected and stored in information systems for future edification. How different the study of history would be if they could! Events are (or can be) informative phenomena and so should be included in any complete approach to information science. In practice we find the evidence of events is used in three different ways:

(1) Objects, which can be collected or represented, may exist as *evidence associated with* events: bloodstains on the carpet, perhaps, or a footprint in the sand;

(2) There may well be *representations* of the event itself: photos, newspaper reports, memoirs. Such documents can be stored and retrieved; and, also,
(3) Events can, to some extent, be *created or recreated*. In experimental sciences, it is regarded as being of great importance that an experiment—an event—be designed and described in such a way that it can be replicated subsequently by others. Since an event cannot be stored and since accounts of the results are no more than hearsay evidence, the feasibility of reenacting the experiment so that the validity of the evidence, of the information, can be verified is highly desirable.

Regarding events as informative and noting that, although events themselves cannot be retrieved, there is some scope for recreating them, adds another element to the full range of information resource management. If the recreated event is a source of evidence, of information, then it is not unreasonable to regard the laboratory (or other) equipment used to reenact the event as being somehow analogous to the objects and documents that are usually regarded as information sources. In what senses does it matter whether the answer to an inquiry derives from records stored in a data base or from reenacting an experiment? What significant difference is there for the user of logarithms between a logarithmic value read from a table of logarithms and a logarithmic value newly calculated as and when needed? The inquirer might be wise to compare the two, but would surely regard both as being equally information. Indeed it would be a logical development of current trends in the use of computers to expect a blurring of the distinction between the retrieval of the results of old analyses and the presentation of the results of a fresh analysis.

To include objects and events, as well as data and documents, as species of information is to adopt a broader concept than is common. However, if we are to define information in terms of the potential for the process of informing, i.e., as evidence, there would seem no adequate ground for restricting what is included to processed data and documents as some would prefer, e.g., by defining information as "Data processed and assembled into a meaningful form." (Meadows, 1984, p. 105). There are two difficulties with such a restricted definition: Firstly, it leaves unanswered the question of what to call other informative things, such as fossils, footprints, and screams of terror. Secondly, it adds the additional question of how much processing and/or assembling is needed for data to be called information. In addition to these two specific difficulties there is the more general criterion that, all things being equal, a simpler solution is to be preferred to a more complicated one. Therefore we retain our simpler view of "information-as-thing" as being tantamount to physical evidence: Whatever thing one might learn from (Orna & Pettit, 1980, p. 3). Fortunately there are moves in the English-language literature of information retrieval toward a more ecumenical approach to information and information systems (Bearman, 1989).

When is Information not Information?

Even if we dismiss the argument that untrue information is not information, we could still ask what *could not* be information? Since being evidence, being information, is a quality *attributed to* things, we may well ask what limits there might be to what could or could not be information. The question has to be rephrased as "What things could not be regarded as informative?" We have already noted that a great variety of things can be regarded as informative so the range is clearly very large.

We might say that objects of which nobody is aware cannot be information, while hastening to add that they might well become so when someone does become aware of them. It is not uncommon to infer that some sort of evidence, of which we are not aware, *ought to* or *might* exist and, if found, would be of particular importance as evidence, as when detectives search, more or less systematically, for clues.

Determining what might be informative is a difficult task. Trees, for example, provide wood, as lumber for building and as firewood for heating. One does not normally think of trees as information, but trees are informative in at least two ways. Obviously, as representative trees they are informative about trees. Less obviously, differences in the thickness of tree rings are caused by, and so are evidence of, variations in the weather. Patterns reflecting a specific cycle of years constitute valuable information for archaeologists seeking to date old beams (Ottaway, 1983). But if lumber and firewood can be information, one hesitates to state categorically of any object that it could not, in any circumstances, be information or evidence. We conclude that *we are unable to say confidently of anything that it could not be information.*

This leads us to an unhelpful conclusion: If anything is, or might be, informative, then *everything* is, or might well be, information. In which case calling something "information" does little or nothing to define it. If everything is information, then being information is nothing special.

Being Information is Situational

Information-as-process is situational. Therefore, evidence involved in information-as-process is so situationally also. Hence, whether any particular object, document, data, or event is going to be informative depends on the circumstances, just as the "relevance" of a document or a fact is *situational* depending on the inquiry and on the expertise of the inquirer (Wilson, 1973). It follows from this that the capability of "being informative," the essential characteristic of informa-

tion-as-thing, must also be situational. We may say of some object or document that in such-and-such a combination of circumstances, in such-and-such a situation, it would be informative, it would be information, i.e., information-as-thing.

But, as noted above, we could in principle say that of *any* object or document: One just has to be imaginative enough in surmising the situation in which it could be informative. And if one can describe anything this way, we are making little progress in distinguishing what information-as-thing is. Further, it is a matter of individual judgement, of opinion

(1) whether some particular thing would be pertinent; and, if so,
(2) whether the probability of it being used as evidence would be significant; and, if so,
(3) whether its use as evidence would be important. (The issue might be trivial or, even if important, this particular evidence might be redundant, unreliable, or otherwise problematic.) And, if so,
(4) whether the importance of the issue, the importance of the evidence, and the probability of its being used—in combination—warrant the preservation of this particular evidence. If all of these are viewed positively, then one would regard the thing—event, object, text, or document—as likely to be useful information and, presumably, take steps to preserve it or, at least, a representation of it.

Information by Consensus

We have shown that (1) the virtue of being information-as-thing is situational and that (2) determining that any thing is likely to be useful information depends on a compounding of subjective judgements. Progress beyond an anarchy of individual opinions concerning what is or is not reasonably treated as information depends on agreement, or on at least some consensus. We can use an historical example to illustrate this point. It used to be considered important to know whether a woman was a witch or not. One source of evidence was trial by water. The unfortunate woman would be put in a pond. If she floated she was a witch. If she sank she was not. This event, the outcome of the experiment, was, by consensus, the information-as-thing needed for the identification of a witch. Nowadays it would be denied, by consensus, that the exact same event constituted the information that it had previously been accepted, by consensus, as being.

Where there is a consensus of judgement, the consensus is sometimes so strong that the status of objects, especially documents, being information is unquestioned, e.g., telephone directories, airline timetables, and textbooks. In these cases arguments are only over niceties such as accuracy, currency, completeness, and cost. As a practical matter some consensus is needed to agree on what to collect and store in retrieval-based information systems, in archives, data bases, libraries, museums, and office files. But because these decisions are based on a compounding of different judgements, as noted above, it is not surprising that there should be disagreement. Nevertheless, it is on this basis that data are collected and fed into databases, librarians select books, museums collect objects, and publishers issue books. It is a very reasonable prediction that copies of the San Francisco telephone directory will be informative, though there is no guarantee that each and every copy will necessarily be used.

"Information-as-thing", then, is meaningful in two senses: (1) At quite specific situations and points in time an object or event may actually be informative, i.e., constitute evidence that is used in a way that affects someone's beliefs; and (2) Since the use of evidence is predictable, albeit imperfectly, the term "information" is commonly and reasonably used to denote some population of objects to which some significant probability of being usefully informative in the future has been attributed. It is in this sense that collection development is concerned with collections of information.

Copies of Information and Representations

Copies: Type and Token

In the provision of access to information by means of formal information systems, the question of whether or not two pieces of information are the same (or, at least, equivalent) is important. When copies are identical one would speak formally of types and tokens. Examples that are not the same as each other are referred to as *different types*; identical copies are referred to as *tokens*. If only one example exists, then one would say that there is only one "token" of that "type."

The creation of identical, equally authentic copies is the result of particular technologies of mass production, such as printing. If you want to reread a particular title (type), you would want to read some copy (token) of it, but you would not insist on rereading the exact same copy as before. Similarly, if you had read a book on some subject and wanted to know more, you would ordinarily move on to reading a copy of another *different* title in preference to reading a different copy of the same title.

This feature of equally acceptable copies can be found in other examples of information systems. Some sorts of museum objects are mass-produced, such as telephones. With telephones as with printed books, one example is as acceptable as any other from the same production run. There is, however, a major qualification. In archival practice, as in museums, two physically identical documents are regarded as different if they occur in different places in the original order of the files. The rationale is that their unique positioning

in relation to other documents makes them unique by association and, thereby, different.

In electronic data bases the situation is a little less clear. One can have copies of two sorts: There can be temporary, virtual copies displayed on a screen; or one can make copies of a longer lasting form on paper or other storage medium. These copies might not, from some engineering error, be quite the same as the original. However, it is ordinarily *assumed* that either the copy is authentic or that errors will be so marked as to be self-evident. There may be difficulty in knowing whether the copy is a copy of the latest, official version of the database, but that is a different issue. With handwritten, manuscript texts, one should *expect* each example to be at least slightly different, even if it purports to be a copy. The person making a copy is likely to omit, add, and change parts of the text. A significant feature of medieval studies is the necessity of examining closely all copies of related manuscripts not only to identify the differences, but also to infer which might be the more correct versions where they do differ.

In general, then, the existence of identical, equally informative, equally authoritative copies is unusual. Printed materials in libraries are a notable exception. More general is the case where copies are not altogether identical, though they may be equally acceptable for most purposes.

Interpretations and Summaries of Evidence

Progress in information technology increases the scope for creating and using information-as-thing. Much of the information in information systems has been processed by being coded, interpreted, summarized, or otherwise transformed. Books are a good example. Virtually all of the books in the collections are based, at least in part, on earlier evidence, both texts and other forms of information. Scholarship is permeated with descriptions and summaries, or, as we prefer to call them, representations.

Representations have important characteristics:

(1) Every representation can be expected to be more or less incomplete in some regard. A photograph does not indicate movement and may not depict the color. Even a color photograph will generally show colors imperfectly—and fade with time. A written narrative will reflect the viewpoint of the writer and the limitations of the language. Films and photographs usually show only one perspective. Something of the original is always lost. There is always some distortion, even if only through incompleteness.

(2) Representations are made for convenience, which in this context tends to mean easier to store, to understand, and/or to search.

(3) Because of the quest for convenience, representations are normally a shift from event or object to text, from one text to another text, or from objects and texts to data. Exceptions to this, such as from object to object or from document back to object (physical replicas and models) can also be found (Schlebecker, 1977).

(4) Additional details related to the object but not evident from it might be added to the representation, either to inform or to misinform.

(5) Representation can continue indefinitely. There can be representations of representations of representations.

(6) For practical reasons representations are commonly (but not necessarily) briefer or smaller than whatever is being represented, concentrating on the features expected to be most significant. A summary, almost by definition, is an incomplete description.

Progress in information technology continually permits improvements in our ability to make physical descriptions, examples of information-as-thing. Photographs improve on drawings; digital images improve on photographs. The voice of the nineteenth century singer, Jenny Lind, was described by Queen Victoria as "a most exquisite, powerful and really quite peculiar voice, so round, so soft and flexible..." (Sadie, 1980, v. 10, p. 865). Although this description is better than none, we could learn much more from a phonograph recording.

Reproductions of works of art and of museum artifacts may suffice for some purposes and have the advantages that they can provide much increased physical access without wear and tear on the originals. Yet they will always be deficient in some ways as representations of the original, even though, as in the case of works of art and museum objects, even experts cannot always identify which is an original and which is a copy (Mills & Mansfield, 1979).

Information, Information Systems, Information Science

We started with two academically respectable usages of the term "information" ("information-as-knowledge" and "information-as-process") and we noted that information systems can deal *directly* only with "information-as-thing." Stating this paradox differently, information systems handle information only in a sense of information dismissed by leading theorists of information. We also concluded that anything *might* be information-as-thing. Small wonder that progress in the development of paradigms for describing and explaining phenomena in the shapeless, ill-defined reaches of "information science" has been slow. But, perhaps, "information-as-thing" could be used to provide some order or arrangement with respect to information-related activities, along with the two more respectable definitions.

First, although all information systems deal directly with "information-as-thing," we might create some order within this area if we could identify a subset of

information-handling activities that are concerned with information *only* in this sense. As examples we might choose information theory (in the sense of the mathematical theory of signal transmission associated with Shannon and Weaver and that has nothing to do with semantic content (Bar-Hillel, 1964); historical bibliography (the study of books as physical objects); and statistical analysis (identifying and defining patterns in populations of objects and/or events). Each of these fields has refined techniques for developing and formalized ways of describing concise and effective representations of their particular kind of information-as-thing. The findings of these useful arts may well be of great significance, but their concern is primarily with the evidence itself. An analysis of a channel, a book, or a population would cease to be valid if the physical characteristics of the channel, book, or population were changed.

Second, information storage and retrieval systems can deal *directly* only with "information-as-thing," but the things that can be stored for retrieval in actual or virtual collections vary in significant ways. Historic buildings, films, printed books, and coded data impose different constraints on the tasks associated with information retrieval systems: selection, collection, storage, representation, identification, location, and physical access. Put simply, a museum, an archive, library of printed books, an online bibliographic database, and a corporate management information system of numeric data can all validly be regarded as species of information retrieval system. But differences in their physical attributes affect how the stored items can be handled (Buckland, 1988a). These differences provide *one* basis for the comparative analysis of information storage and retrieval systems.

Third, representations of knowledge form a distinguishable subset of information-as-thing and so could, in principle, be used to identify and define another class of information systems in which the primary concern is based on the knowledge represented. This is the conventional area of information storage and retrieval, subject bibliography, and "knowledge bases" for expert systems. In these cases the information-as-thing is unavoidably of concern, but only a *means* for dealing with information-as-knowledge and, being merely a means, considerable latitude is imaginable. In providing an information service different physical forms of information and different text-bearing media (texts on paper, on microform, or displayed on a terminal) may be equally acceptable. Further, a wide variety of variant texts could be more or less substitutable—in English or in French, lengthy or concise, recent or old—if they represented the same knowledge to an acceptable degree.

Fourth, information-as-process could also be the basis for defining a class of information-related studies. Here again, information-as-thing cannot be ignored, but is, again, of secondary interest as a means. Cognitive psychology, rhetoric, and other studies of interpersonal communication and persuasion would be examples. Alternative means, i.e., alternative physical media, might be equally acceptable. Indeed, inasmuch as the primary interest is on cognition and persuasion, the actual information-as-knowledge, also a necessary ingredient, may also be of little direct interest. The focus could well be more on *how* beliefs change than on *which* beliefs are changed or *which* knowledge is represented.

It is not asserted that sorting areas of information science with respect to their relationship to information-as-thing would produce clearly distinct populations. Nor is any hierarchy of scholarly respectability intended. The point is rather that examination of "information-as-thing" might be useful in bringing shape to this amorphous field and in avoiding simplistic, exclusive boundaries based on past academic traditions.

Summary

Numerous definitions have been proposed for "information." One important use of "information" is to denote knowledge imparted; another is to denote the process of informing. Some leading theorists have dismissed the attributive use of "information" to refer to *things* that are informative. However, "information-as-thing" deserves careful examination, partly because it is the only form of information with which information systems can deal directly. People are informed not only by intentional communications, but by a wide variety of objects and events. Being "informative" is situational and it would be rash to state of *any* thing that it might not be informative, hence information, in some conceivable situation. Varieties of "information-as-thing" vary in their physical characteristics and so are not equally suited for storage and retrieval. There is, however, considerable scope for using representations instead.

Acknowledgment

This work was partially supported by a Fulbright Research Scholarship at Graz University of Technology, Austria, during sabbatical leave from the University of California at Berkeley. The helpful comments of William S. Cooper, Brian Peaslee, W. Boyd Rayward, and Patrick Wilson are gratefully acknowledged.

References

Bar-Hillel, Y. (1964). *Language and information.* London: Addison-Wesley. Chap. 16: "An examination of Information Theory," 275–297.

Bearman, D. (1989). Archives and manuscript control with bibliographic utilities: Challenges and opportunities. *American Archivist 52,* 26–39.

Belkin, N; Robertson, S.E. (1976). Information science and the phenomena of information. *Journal of the American Society for Information Science, 27,* 197–204.

Blair, D.C. (1984). The data-document distinction in information retrieval. *Communications of the association for Computing Machinery, 27,* 369–374.

Braman, S. (1989). Defining information. *Telecommunications Policy, 13,* 233–242.

Briet, S. (1951). *Qu'est que la documentation?* Paris: Editions Documentaires Industrielles et Techniques.

Brookes, B.C. (1974). Robert Fairthorne and the scope of information science. *Journal of Documentation 30,* 139–152.

Brookes, B.C. (1979). Theoretical informatics in stage II mechanization of IR processes. In International Federation for Documentation, *Theoretical problems of informatics: New trends in informatics and its terminology.* FID 568. (pp. 10–18). Moscow: VINITI.

Buckland, M.K. (1988a). Library materials: Paper, microform, database. *College and Research Libraries 49,* 117–122.

Buckland, M.K. (1988b). *Library services in theory and context.* 2nd ed. New York: Pergamon.

Clarke, D.S. (1987). *Principles of semiotic.* London: Routledge & Kegan Paul.

Buzzard, J.H. et al. (1976). *Phipson on Evidence.* 12th ed. (The Common Law Library, 10). London: Sweet & Maxwell.

Eco, U. (1976). *A theory of semiotics.* Bloomington, Ind.: Indiana University Press.

Fairthorne, R.A. (1954). The theory of communication. *Aslib Proceedings 6,* 255–267. (Reprinted in Fairthorne, R.A. (1961) *Towards Information Retrieval.* London: Butterworths. pp. 64–79.

Fox, C.J. (1983). *Information and misinformation.* Westport, CT: Greenwood.

Heilprin, L. (1974). Operational definitions. In: A. Debons (Ed.), NATO Advanced Study Institute in Information Science, Champion, 1972. *Information science: Search for identity,* (pp. 115–38). New York: Marcel Dekker.

Laisiepen, K. et al. (1980). *Grundlagen der praktischen Information und Dokumentation: Eine Einführung.* 2. Aufl. München: Saur.

Machlup, F. (1983). Semantic quirks in studies of information. In F. Machlup, & U. Mansfield, (Eds), *The study of information: Interdisciplinary messages.* (pp. 641–671). New York: Wiley.

Meadows, A.J. et al. (1984). *Dictionary of computing and new information technology.* London: Kegan Paul.

Mills, J.F., & Mansfield, J.M. (1979). *The genuine article.* London: British Broadcasting Corporation.

Minns, J.E. (1973). *Model railway engines.* London: Octopus Books.

NATO Advanced Study Institute in Information Science, Champion, 1972. (1974). *Information science: Search for identity.* (pp. 115–38). New York: Marcel Dekker.

NATO Advanced Study Institute in Information Science, Aberystwyth, 1974. (1975) *Perspectives in information science.* Leyden: Noordhoff.

NATO Advanced Study Institute in Information Science, Crete, 1978. (1983). *Information science in action: Systems design.* Boston: Martinus Nijhoff.

Orna, E., & Pettit, C. (1980). *Information handing systems in museums.* New York: Saur.

Otlet, P. (1934). *Traité de documentation.* Brussels: Editiones Mundaneum.

Ottaway, B.S. (Ed.). (1983). *Archaeology, dendrochronology, and the radiocarbon calibrating curve.* Edinburgh: Edinburgh University, Dept of Archaeology.

Oxford English Dictionary (1989) 2nd ed. Oxford: Clarendon Press.

Pollard, A.F.C. (1944). *British Society for International Bibliography Proceedings 6,* 54.

Rogalla von Bieberstein, J. (1975). *Archiv, Bibliothek und Museum als Dokumentationsbereiche: Einheit und gegenseitige Abgrenzung.* Pullach bei München: Verlag Dokumentation.

Sadie, S. (Ed.). (1980). *The new Grove dictionary of music and musicians.* Vol. 10, p. 865. London: Macmillan.

Sagredo, F., & Izquierdo, J.M. (1983). *Concepción lógico-lingüístico de la Documentación.* Madrid: Ibercom—Red Comnet de la Unesco.

Schlebecker, J.T. (1977). The use of objects in historical research. *Agricultural History 51,* 200–208.

Schrader, A. (1984). In search of a name: Information science and its conceptual antecedents. *Library and Information Science Research 6,* 227–271.

Schreiner, K. (1985). *Fundamentals of museology: On the theory and methodology of collecting, preserving, decoding, and utilizing musealia* Waren, GDR: n.p.

van Rijsbergen, C.R. (1979). *Information retrieval.* 2nd ed. London: Butterworths.

Washburn, W.E. (1964). Manuscripts and manufacts. *American Archivist 27,* 245–250.

Wellisch, H. (1972). From information science to informatics. *Journal of Librarianship 4,* 157–187.

Wersig, G.; Neveling, U. (1975). The phenomena of interest to information science. *Information Scientist 9,* 127–140.

Wersig, G. (1979). The Problematic Situation as a Basic Concept of Information Science in the Framework of the Social Sciences: A Reply to Belkin. In International Federation for Documentation. *Theoretical Problems of Informatics: New Trends in Informatics and its Terminology.* FID 568. (pp. 48–57). Moscow: VINITI.

Wersig, G. (1980). Informationstätigkeit. In K. Laisiepen, et al. (Eds.), *Grundlagen der praktischen Information und Dokumentation: Eine Einführung.* 2. Aufl. (pp. 161–192). München: Saur.

Wigmore, J.H. (1983). *Evidence in trials at common law.* Vol. 1, revised by P. Tillers. (2–12). Boston: Little Brown, 1983.

Wilson, P.G. (1973). Situational relevance. *Information Storage and Retrieval 9,* 457–471.