What Kind of Science Can Information Science Be?

Michael Buckland
School of Information
University of California, Berkeley, CA  94720-4600

ABSTRACT
During the twentieth century there was a strong desire to develop an Information Science from librarianship, bibliography, and documentation and in 1968 the American Documentation Institute changed its name to American Society for Information Science. By the beginning of the twenty-first century, however, departments of (library and) information science had turned instead towards the social sciences. These programs address a variety of important topics, but they have been less successful in providing a coherent explanation of the nature and scope of the field. Progress can be made towards a coherent, unified view of the roles of archives, libraries, museums, online information services, and related organizations if they are treated as information-providing services. But such an approach seems significantly incomplete on ordinary understandings of the providing of information. Instead of asking what Information Science is or what we might wish it to become, we ask instead what kind of field it can be given our assumptions about it. We approach the question by examining some key words: science, information, knowledge and interdisciplinary. We conclude that if information science is concerned with what people know, then it is a form of cultural engagement and, at most, a science of the artificial.

INTRODUCTION
During the twentieth century there was a strong desire for the provision of information services to become scientific, to move from librarianship, bibliography, and documentation to an Information Science. Accordingly, in 1968 the American Documentation Institute changed its name to American Society for Information Science. By the beginning of the twenty-first century, however, departments of (library and) information science had turned instead towards the social sciences. Leading programs have increased their size and visibility with skillful publicity liberally using the words “information,” “society,” and “technology.” “Information school” is currently a name or nickname of choice. These programs address a variety of important topics, but they have been less successful in providing a coherent explanation of the nature and scope of the field. It is wise for organizations to be prospecting for new opportunities, but to be opportunistic without a coherent underlying rationale appears imprudent.

A related problem concerns the analysis of information services. Some progress can be made towards a coherent, unified view of the roles of archives, libraries, museums, online information services, and related organizations if they are treated as information-providing services (e.g. Buckland, 1991a), but such an approach seems significantly incomplete on ordinary understandings of the providing of information. Public libraries, for example, do more than simply provide information. Here again a deeper or wider or different explanation is needed.

APPROACH
Our approach is to consider some key words: “Science,” “Information,” “Knowledge,” and “Interdisciplinary”” and to make distinctions between scientific, scholarly, and critical.

Although the word “science” is sometimes used broadly for any body of knowledge (e.g. domestic science, library science), here we are using it in the normative sense as denoting formal and
Buckland: What Kind of Science can Information Science Be?

physical sciences (e.g. chemistry, mathematics, and physics). Science is a constructive enterprise. Being scientific involves model-building. Hypotheses and theories are developed to explain and to predict observable phenomena. To be scholarly involves more than being knowledgeable. It requires the affirmative search for evidence contrary to one’s theories. This is true for all fields: in the humanities, the social sciences, the sciences, and professional practices. In this context, being critical is not a matter of being hostile or negative, but of asking questions about underlying assumptions and methodological choices. How have conclusions been determined, or at least influenced, by particular assumptions or the choice of method? The ideal is to be scientific and scholarly and critical. The more we can approach that ideal the more robust our ideas will be.

It is also important to remember the distinction between things and their names: Describing some phenomenon is a separate matter from deciding what to call it. Names can be multiple, ambiguous, and unstable. Past discussion of information and related terms has been hindered by failure to recognize this rather obvious distinction and any declarative statement in the form “Information is . . .” should be viewed with suspicion absent some explanation of what is being referred to. Another useful guideline is Ockham’s Razor, the principle that, other things being equal, the simplest explanation is generally to be preferred. These principles provide a basis not only for examining individual notions of information, but also to consider what kind of a field Information Science can be and, thereby, identify plausible terrain for information school programs.

INFORMATION

The word “information” has been used so much that it has come to dominate discourse (Day, 2001). One information school website recently contained two striking statements: “161 exabytes of new information are created each year” (They mean digital bits) and “Information: The power to transform the world” (They don’t mean digital bits). Vagueness and inconsistency are advantageous for slogans and using “chameleon words” that assume differing colors in different contexts allows flexibility for readers to perceive what they wish. However, when clarity is sought more careful definitions are needed.

Our first restriction is to limit our use of “information” to its traditional association with human knowing and learning. This differentiates our scope from other important fields that have also used the name “Information Science.” One is Computer Science, concerned with the theory and application of algorithms. Another, concerned with entropy, probability, Shannon-Weaver information theory, physical patterns (in-form-ing), and related topics, is sometimes referred to as the “physics of information.” Also, the word “information” is, of course, used in “information technology” (IT, also ICT, for Information and Communication Technologies) but largely restricted in practice to the use of electronics for communication and computation. These other areas are not considered here. Instead, we are concerned with those areas generally understood as being within the scope of “library and information science” (LIS) and the interests of the American Society for Information Science and Technology. For a wider and more detailed analysis by Machlup and Mansfield of numerous fields with some interest in information see Study (1983).

Jonathan Furner has wisely reminded us that for each of the multiple meanings of the word “information” there is already another satisfactory more specific word (Furner 2004). Information studies does not require use of the word “information”! Another move is to sort the varied uses of the word “information” into categories, including:

– Information-as-knowledge for knowledge imparted, what was learned as a result of being informed;
– Information-as-process for becoming informed, for learning; and
– Information-as-thing for bits, bytes, books, sounds, images, and anything physical perceived as signifying. The word “document,” which was not, historically, limited to textual media, can be used as a technical term for information-as-thing (Buckland 1991a, b, 1997).
Starting with this last category, information-as-thing, we can ask what documents do or, more, correctly, what people do with information-as-thing, with documents, that is to say with data, records, texts, and media of every kind.

THE USE OF DOCUMENTS

We find, when we look, that documents are widely used for a variety of purposes. Governments use documents to control us, requiring the use of passports, income tax returns, drivers’ licenses, and so on. Schools use textbooks and curriculum standards to guide both students and teachers. Religions use sacred texts to instill beliefs and to influence conduct. Merchants invest heavily in advertisements to influence what we buy. Politicians use slogans and policy statements to win votes and to attract financial and electoral support. Entertainers use varied media to amuse us and, usually, to attract payments from us. Individuals use messages to communicate and social media to attract attention. Museums present interpretations of our heritage through the selective presentation and skilful interpretation of artifacts. Libraries provide access to collections of documents . . . and so on. Anyone can make such a list and the list quickly becomes a long one.

Contemplating this or any similar list reminds us of some important points:

1. Documents are pervasive in society and shape our lives. Dependence on documents has increased over time. Modern economies are based on an ever-increasing division of labor and on the existence of markets, both of which depend on communication and documentation, which in turn have been progressively facilitated by technical innovations (writing, printing, telegraphy, radio, Internet, etc.). As Patrick Wilson put it, we are more and more dependant on “second-hand knowledge” (Wilson 1983).

2. The use of information and “information behavior” are ordinarily understood as referring to the individual who would like to be informed. But, as is clear from the list, that is only a small part of the story. Much of the use of documents is not initiated by the “user” but by a wide and diverse set of very active agents (governments, schools, religions, merchants, etc.) with differing and sometimes competing purposes.

3. The commonest form of information-related behavior is simply noticing things, a minimally active role. It may be unintended (as when we hear thunder), unexpected, or unconscious (when subliminal).

4. The use of documents may include but does not reduce to fact-finding, information-seeking, or problem-solving. As is clear from the list, the agendas and means are varied. Public libraries are not simply information services, at least not in any simple or normal sense. “One of the things that public libraries have done fairly well is to realize that their mission, their job, is about community building,” states Martin Gómez (IMLS 2009, 9).

If we contemplate the list above or any similar list, it is reasonable to ask what term can embrace this range of information-related activities. The common feature is that they are cultural. Here we do not use “culture” in the popular sense of high culture, denoting grand opera and other elitist activities, but in the broader academic sense used in anthropology. The classic definition is by sir Edward Tylor in 1871: “Culture or civilization, taken in its wide ethnographic sense, is that complex whole which includes knowledge, belief, art, morals, law, custom and any other capabilities and habits acquired by man as a member of society”(Tylor 1871, 1). The simplest assumption, then, is that the uses of information, when we speak of information-as-thing, are properly seen as an active engagement in the cultural sphere.

KNOWLEDGE

The theory of knowledge has been dominated by analytical philosophy with an emphasis on the
truth of propositional sentences and knowledge as justified true belief. This approach is problematic in several ways. We can question the adjective “justified” since nobody is likely to accept that they hold unjustified beliefs. The “true” criterion also does not hold up well on inspection. In ordinary discourse, “true” tends to imply consistency with some objective reality, but the subjective knowing of objective reality is philosophically suspect and in practice, “true” reduces to congruence with some other prior belief or assumption.

Propositional knowledge (justified true belief) is illustrated in this excerpt from the *Stanford Encyclopedia of Philosophy*:

Suppose, for example, that James, who is relaxing on a bench in a park, observes a dog that, about 8 yards away from him, is chewing on a bone. So he believes

5. There is a dog over there.

Suppose further that what he takes to be a dog is actually a robot dog so perfect that, by vision alone, it could not be distinguished from an actual dog. . . . Given these assumptions, (5) is of course false. But suppose further that just a few feet away from the robot dog, there is a real dog. Sitting behind a bush, he is concealed from James's view. Given this further assumption, James's belief is true. So once again, what we have before us is a justified true belief that . . . gives us the wrong result that James knows (5).

Analytical philosophy of this sort has little relevance to the everyday realities of a document-pervaded society, our unavoidable dependence on second-hand knowledge, and the perennial need to decide who and what to trust. A famous seventeen-century textbook on logic summarized the situation rather well:

. . . a wide difference must be made between two kinds of truths: one, which relates simply to the nature of things, and their unchangeable essence, independently of their existence; the others, which relate to things existing, and especially to human accidents and events, . . .

In the first kind of truths, since everything is necessary, nothing is true which is not true universally; and thus we may conclude that a thing is false, if it is false in a single case. But if we think of following the same rules in the belief of human events, we shall always, except by accident, judge falsely, and make a thousand false reasonings about them. For these events being contingent in their nature, it would be ridiculous to seek in them necessary truth: . . . (Arnauld 1662/1850, 345-6).

In 1946 Gilbert Ryle wrote: “Philosophers have not done justice to the distinction which is quite familiar to all of us between knowing that something is the case and knowing how to do things” (Ryle 1946, 4). He argued that knowing how cannot be defined in terms of knowing that and that knowing how was logically prior to knowing that. But this is not enough. The theory of knowledge needs to be extended further to another distinction quite familiar to all of us: knowing about. In our daily lives we operate with necessarily imperfect, incomplete, and uncertain knowledge. We must continually make decisions on whether to depend on this document or that. In real life we have imperfect knowing about and we have to rely more on trust than on truth. Seen this way, the importance of marshalling the most suitable available documents for ourselves or for others, a core concern of LIS, is evident. In this situation a distinction between knowledge and belief seems questionable, and propositional knowledge, preoccupied with the truth of single sentences, becomes an implausible theoretical foundation.

**BECOMING INFORMED**

The remaining category, information-as-process, is concerned with the imparting of knowledge, with learning. So long as we are concerned with understanding rather than mere memorization, learning
depends on what we already know. Learning is incremental, a change in what we knew rather than simple addition, except, it seems, in LIS research where we find a fundamental deficiency. A detailed content analysis of LIS literature by Allan Konrad found that only 5.6% of a selection of 413 canonical texts examined were consonant with the principle that learning is incremental; most (88.8%, including 83% in a subset categorized as cognitive studies) either ignored the principle or made only token mention of it; and 5.6% explicitly or implicitly refuted it (Konrad 2007, 499-569, esp. 508).

Given the trend of the field towards the social sciences and explicit talk of a “cognitive turn” (e.g. Ingwersen & Järvelin 2005), these findings are striking. We can speculate on the reasons. One consideration is that it is very difficult in practice to take into account what individuals already know. Another factor is that formal and algorithmic techniques by their nature resist the inclusion of culture (Ekbia 2008). Newtonian physics allowed no place for heaven or hell, and information science has focused heavily on information storage and retrieval systems, in effect on document-supplying systems, rather than systems that inform (Buckland 1991a, 78-80). Third, with some commendable exceptions (such as Allen Bryce and Carol Kuhltau) the so-called cognitive turn has tended instead to be a rather narrowly-based cognitive science turn. (Witness the frequent reference to artificial intelligence papers rather than the wider realms of educational psychology). The frequent reference to “states of knowledge” implies a dubious simplification, since each time we remember something we create a slightly different recollection. Søren Brier characterized the situation as follows: The current dominant paradigm is heavily influenced by cognitive science which is a logical and algorithmic research program that investigates information processing in humans, animals, and machines. This approach is based on Wiener's Cybernetics, Shannon-Weaver information theory, logic, set theory, and computation. It is inadequate because it fails to accommodate the cultural realities of knowing and communicating, the phenomenological complexity of perception and understanding, or the interaction of the social and the personal. The result is a general confusion among many alternative meanings of the word "information" and an approach to information behavior that is inhospitable to both communication and learning (Brier 2008).

LANGUAGE AND FACTS

Information retrieval, widely regarded (along with bibliometrics) as being the most scientific part of information studies, depends heavily on algorithmic operations on text, especially the co-occurrence of specified words (actually character strings) in both query and searched documents. These methods are enormously useful despite some weaknesses that arise from words having multiple meanings and variant forms, different words having the same spelling, and meanings being unstable. Human communication, in contrast, depends on cultural codes and meaning. Robert Fairthorne had good insights into these issues with his careful distinction between mention and meaning and his explanation of the irresistible obsolescence of subject indexing. Language evolves in dialog and discourse. The indexer is necessarily backward-looking because index terms need to be based on usage already established in past discourse. But the indexer also needs to be forward-looking because indexing is intended for future use. Word meanings continue to evolve with time, but an index term inscribed at some fixed point in time recedes into the past as discourse, language, and the indexer flow forward (Fairthorne, 1974; Buckland, 2007, in-press).

If culture and language resist algorithms and formal techniques, more progress might be made if we could reduce the literary to the factual. Paul Otlet thought so. He considered books and articles to be inefficient, opinionated, and duplicative. His idea was to extracts facts from texts, like peas from pods, and to organize the facts into an authoritative semantic web using concise unitary factual statements (“monographs”) described, positioned, and collectively associated using the Universal Decimal Classification (Frohmann, 2008). The result, he declared, could be shared as a communal extension of the brain. (Otlet’s understanding of a “world brain,” shared with Wilhelm Ostwald and H. G. Wells, was a community resource more like a rigorously edited Wikipedia than an autonomous entity like the computer
But, at the very same time that Otlet was summarizing his ideas in his encyclopedic *Traité de documentation*, published in Brussels in 1934, Ludwik Fleck in Poland was arguing a very different view in his *Genesis and development of a scientific fact* published the next year in 1935 (Fleck, 1935/1979). Fleck argued that facts found in popular encyclopedias were oversimplified when reduced to simple statements out of context and isolated from explanatory narratives. Further, facts arose only in a triadic relationship of a concept, the individual, and the prevailing cultural mindset (*Denkkollektiv*) that both enabled and constrained. Even scientific facts, argued Fleck, long anticipating the paradigms and scientific revolutions of Thomas Kuhn and the archaeology of knowledge of Michel Foucault, are culturally-situated constructs. Even a little attention to intellectual history illustrates his case. Paracelsus, the renaissance physician who struggled towards modern science by pioneering the medicinal use of chemicals and acute attention to the size of dose, was so immersed in medieval alchemy that he lacked adequate concepts and terminology (Ball, 2006). He would not have understood our modern medical texts and we cannot comprehend his.

**THE SUITABLE ARRANGEMENT OF DOCUMENTS**

Vesa Suominen’s answer to the question “What constitutes a good librarian?” was that a good librarian is one who achieves a suitable arrangement of documents for the reader (Suominen, 1997). That there are many different readers, that each has multiple interests, that there are very many documents, and that readers, interests, and documents are all quite unstable greatly complicates the task, but the notion is, in principle, attractive. Certainly, library and information science is much concerned with the suitable arrangement of documents in one way and another.

*Bibliography* and *bibliographical description* are concerned with establishing suitable arrangements in two ways: Relationships between documents are established through descriptions, descriptive lists, and indexes and these descriptions, descriptive lists, and indexes are used to identify suitable documentary means for some purpose (Wilson, 1968).

*Information retrieval systems*, however complex, have the same underlying properties as bibliography. All selection machinery (both for retrieval and for filtering) are composed of chains of just two primitive types of operation: the modification of documents (including the derivation of indexes) and their (re)arrangement (sorting, ranking, clustering, and the like) (Buckland & Plaunt, 1994; Plaunt, 1997). Information retrieval is algorithmic, quantitative, and hugely useful, but is it scientific? (cf. Neill, 1992). Selection system operations depend on set theory and relevance. Set theory is a convenient simplification since documents are not really discretely different in content or meaning. A larger problem is that there is no such thing as relevance, at least nothing tangible, as becomes clearer if we substitute the word “suitable” for “relevant,” which we can do without changing the meaning. *Relevance* may seem more scientific because it has a formal meaning of entailment in logic and because of fifty years of relevance measures in information retrieval evaluation. In practice documents are ranked using some arbitrary surrogate for relevance. These substitutes range from the co-occurrence of character strings through the use of third party relevance judgments, sometimes modified, after the initial retrieval by the subjective perception of those for whom the retrieval has been performed. In general, the fall-back position is whether the topic (or word usage) is similar (Buckland, 1983). It is little wonder that the definitions and literature on relevance have remained stubbornly problematic for fifty years despite the sustained efforts of so many talented and motivated researchers. A natural science (like chemistry or physics) would require a measurable physical property. A formal science (like logic or mathematics) would require a clear and rigorous definition. It is characteristic of softer social sciences that neither is available and one must do the best one can with the least unsatisfactory surrogates.

*Bibliometrics*, mainly based on citation analysis, is the other epicenter of quantification in Information Science. Here, both the motives and the significance of individual acts of citation tend to remain unclear, except in vague general terms, so bibliometrics resembles information retrieval in that
virtuoso calculation is not based on firm foundations. Both bibliometrics and information retrieval bring methods developed in and for formal (logical, well-defined) environments and use them on objects and in environments that are not formal, logical or well-defined. This yields useful results but also necessarily compromised, incongruous processes.

INTERDISCIPLINARY

Being “interdisciplinary” is widely considered to be a good thing and sometimes it is. A good practical example would be when a desired academic program has not been approved, presentation of an alternative program framed as being interdisciplinary might well succeed. Nevertheless, words beginning with “inter” commonly imply a position of weakness (e.g. interval, intermission, interregnum, and interim) and indicate something positioned inbetween other more substantial entities.

A personal view is that in a university environment claims to being interdisciplinary have an attraction among planners but that in times of economic crisis political power tends to reside in well-established disciplines. So, arguing a claim to resources based on being interdisciplinary or on being an emerging discipline is, in general, to choose to occupy a weak position.

Fortunately for information studies, there is a strong alternative: societal need. Who wants to have to deal with an ignorant mechanic, a physician with outdated medical knowledge, an ill-informed manager, obsolete manuals, or a secretive government? With some exceptions, notably relating to privacy and security, we all have a substantial vested interest in a knowledgeable society. We need people who are well-informed, who know about what they are doing. Major social needs are typically complex. Whoever undertakes to try to solve them needs to be methodologically versatile in a way that is inadequately captured by “interdisciplinary.” There is an irony in this since the most respectable academic departments (e.g. history, chemistry, and languages) originated in nineteenth century perceptions of the societal needs of the nation-state.

Søren Brier’s book, Cybersemiotics: Why information is not enough! is an unusually erudite, complete, and cohesive theorizing of the nature of information studies. Is this book interdisciplinary? It draws widely on several fields, including biology, cybernetics, psychology, semiotics, and more, so it clearly is. But to say so misses the more important point that it is a coherent unifying theory for an existing field.

Each academic specialty develops its own culture of knowledge, language, values, and social structures. In consequence they are necessarily more or less different from each other in scope and potentially incompatible or, at least, dissonant. No specialty is likely to prefer a unified culture (epistemology, terminology) to its own evolving native culture, so a tension is to be expected between a desire for the benefits of compatibility with other specialties and the discomfort of dealing with the more or less alien cultures of other specialties. The late Ylva Lindholm-Romantschuk, who studied the flow of ideas within and among disciplines, was of the opinion that the most productive position was to be firmly grounded in one’s own field and to then go prospecting at or over the frontiers with other fields (Lindholm-Romantschuk, private communication 1994).

CONCLUSIONS

“Information” and other vague and/or polysemic words can be very valuable in slogans and in rhetoric.

“Information science” has been used to denote different fields which we can distinguish by using different names: library and information science; computer science; the physics of information, entropy, etc.; and information technology, meaning electronic technology applied to communication and computation. Of these, only the first is directly concerned with knowing and learning.

Enabling people to become better informed (learning, becoming more knowledgeable) is, or should be the central concern of information studies and information services are, in practice, more directly concerned with knowing about than with knowing how or knowing that. Knowledge in everyday
life is belief, is cultural, and is not necessarily well-justified or true in any strong sense. One consequence is that the niceties of analytical philosophy provide an unsuitable basis for theorizing Information Science.

In everyday life we depend heavily and more and more on second-hand knowledge. We can determine little of what we need to know by ourselves, at first hand, from direct experience. We have to depend on others, largely through documents. Correspondingly there is a multiplicity of agencies eager to influence our lives and using documents as their means to achieve their varied and sometimes controversial ends. In this flood of information, we have to select and we have to decide what to trust. What we believe about a document influences our use of it and also, more importantly, our use of documents influences what we believe. Suzanne Briet was recognizing these issues when she wrote in 1951 of documentation not only as “a necessity of our time” but also as “a new cultural technique” (Briet, 1951/2006, emphasis added; Day, 2006).

Information retrieval and bibliometrics, both very useful, are quantitative and technical but not scientific in the normative sense because they are based on ill-defined foundations. If Information Science is a science it is a science of the artificial (Simon, 1996) rather than a natural science (like physics) or a formal science (like mathematics). Patrick Wilson was right: Information studies involves a broad range of the social sciences (and humanities) and some highly specialized engineering (Wilson, 1996, also 1983).

These conclusions would have displeased many who in the twentieth century were determined to create a serious science of information science. The response has to be that if a problem is important the character of the problem should determine the methodology, not the other way around. It should be some consolation that any (re)frameing of the field illuminates opportunities as well as limitations. Some that spring to mind are that if the techniques of analytical philosophy in propositional knowledge appear sterile for our needs, then an emphasis on knowledge as belief and as cultural should be fertile; if we take a functional view and note that relationships between documents have long been studied in the humanities, then bibliography can be enriched by regarding it as a form of paratext and vice versa; if the use of algorithms depends on a useful simplification, a deeper investigation of the consequences of this compromise is indicated; the Internet has amplified the question of deciding what documents to trust, but study of how belief affects document use needs to be complemented by study of how document use affects belief; and if the incremental nature of becoming informed has been neglected, there is much to be done to shift the emphasis from information-supplying services to systems that inform.

For the reasons set out in this paper Information Science is concerned with cultural engagement. Formal and quantitative approaches are extremely valuable, but the field itself is incorrigibly cultural. Formal and quantitative methods, however useful, can never be more than in highly-valued auxiliary roles. Characterizing information retrieval and bibliometrics as sciences of the artificial is a description not a criticism.

These conclusions are not directed at other, different kinds of information study, notably computer science, the physics of information, or information technology, which are not directly concerned with what people believe.

In the end we can see that our arguments have a somewhat circular form. Once we choose to recognize the core notion of information as having to do with knowing and learning there are consequences. First there is a separation from the essentially knowledge-free zones occupied by computer science, the physics of information, and information technology. Second, any notion of information studies involving what and how we know can only be a cultural inquiry. Third, useful formal and quantitative tools depend on significant simplifying compromises needed to diminish the subjective and cultural qualities of the field. Finally, accepting the cultural context of Information Science should lead to a more realistic and more effective contribution to our document-pervaded society.

ACKNOWLEDGMENT Earlier versions of this paper were presented at that Document Academy
conference hosted by the College of Information, University of North Texas, Denton, TX, on March 19, 2010 and at the School of Library and Information Science, University of South Carolina, April 7, 2011.

REFERENCES


