Conceptual Approaches for Defining Data, Information, and Knowledge

Chaim Zins

Knowledge Mapping Research, 26 Hahaganah Street, Jerusalem 97852, Israel.
E-mail: chaim.zins@gmail.com

The field of Information Science is constantly changing. Therefore, information scientists are required to regularly review—and if necessary—redefine its fundamental building blocks. This article is one of a group of four articles, which resulted from a Critical Delphi study conducted in 2003–2005. The study, "Knowledge Map of Information Science," was aimed at exploring the foundations of information science. The international panel was composed of 57 leading scholars from 16 countries, who represent (almost) all the major subfields and important aspects of the field. This particular article documents 130 definitions of *data, information*, and *knowledge* formulated by 45 scholars, and maps the major conceptual approaches for defining these three key concepts.

Introduction

Context

The field of Information Science (IS) is constantly changing. Therefore, information scientists are required to regularly review—and if necessary—redefine its fundamental building blocks. This article is one of a group of four articles, which resulted from a Critical Delphi study conducted in 2003–2005. The study, Knowledge Map of Information Science, explores the theoretical foundations of information science. It maps the conceptual approaches for defining *data*, *information*, and *knowledge*, which is presented here, maps the major conceptions of Information Science (Zins, 2007a), portrays the profile of contemporary Information Science by documenting 28 classification schemes compiled by leading scholars during the study (Zins, in press), and culminates in developing a systematic and scientifically based knowledge map of the field (Zins, 2007b).

The three concepts of data, information, and knowledge, which are the foci of this article, are fundamental in the context of information science. They are often regarded as the

basic building blocks of the field. For this very reason, the formulation of systematic conceptions of data, information, and knowledge is crucial for the development of a systematic conception of Information Science, as well as for th construction of a systematic knowledge map of the field.

Data, Information, and Knowledge

The academic and professional IS literature supports diversified meanings for each concept. Evidently, the three key concepts are interrelated, but the nature of the relations among them is debatable, as well as their meanings.

Interrelations. Many scholars claim that data, information, and knowledge are part of a sequential order. Data are the raw material for information, and information is the raw material for knowledge. However, if this is the case, then Information Science should explore data (information's building blocks) and information, but not knowledge, which is an entity of a higher order. Nevertheless, it seems that information science does explore knowledge because it includes the two subfields, knowledge organization, and knowledge management, which can be confusing. Should we refute the sequential order? Should we change the name of the field from *Information Science* to *Knowledge Science*? Or should we go to the extreme of excluding the two subfields of knowledge organization and knowledge management from information science?

Information versus knowledge. Another common view is that knowledge is the product of a synthesis in the mind of the knowing person, and exists only in his or her mind. If this is the case, we might well exclude the subfields of knowledge organization and knowledge management from information science. Besides, is Albert Einstein's famous equation "$E = MC^2$" (which is printed on my computer screen, and is definitely separated from any human mind) information or knowledge? Is "$2 + 2 = 4$" information or knowledge?

Synonyms. The alternative view that information and knowledge are synonyms is problematic too. If information and knowledge are synonyms, could we use the term *Knowledge Science* rather than Information Science?

Such issues are rooted in various subjectivist and empiricist schools of philosophy, and are not addressed here as a philosophical treatise. This article is focused on exploring the meanings of the three fundamental concepts of data, information, and knowledge and the relations among them, as they are perceived by leading scholars in the information science academic community.

Methodology

The scientific methodology is Critical Delphi. Critical Delphi is a qualitative research methodology aimed at facilitating critical and moderated discussions among experts (the panel). The international and intercultural panel is composed of 57 participants from 16 countries. It is unique and exceptional; it is comprised of leading scholars who represent nearly all the major subfields and important aspects of the field (see Appendix A). The indirect discussions were anonymous and were conducted in three successive rounds of structured questionnaires. The first questionnaire contained 24 detailed and open-ended questions covering 16 pages. The second questionnaire contained 18 questions in 16 pages. The third questionnaire contained 13 questions in 28 pages (see relevant excerpts from the three questionnaires in Appendix B). The return rates were relatively high: 57 scholars (100%) returned the first round, 39 (68.4%) returned the second round, and 39 (68.4%) returned the third round. Forty-three panelists (75.4%) participated in two rounds (i.e., R1 and either R2 or R3), and 35 panelists (61.4%) participated in all three rounds. In addition, each participant received his or her responses that I initially intended to cite in future publications. The responses were sent to the each panel member with relevant critical reflections. Forty-seven (82.4%) participants responded and approved their responses. Twenty three of them, which is 48.9% (23 out of 47), and 40.3% of the entire panel (23 out of 57) revised their original responses. Therefore, one can say that actually the critical process was composed of four rounds.

The Panel's Definitions

Forty-four panel members contributed their definitions and reflections as follows.[1]

> **Data.** In computational systems data are the coded invariances. In human discourse data are that which is stated, for instance, by informants in an empirical study. **Information** is related to meaning or human intention. In computational systems information is the contents of databases, the web, etc. In human discourse systems information is the meaning of statements as they are intended by the speaker/writer and

understood/misunderstood by the listener/reader. **Knowledge** is embodied in humans as the capacity to understand, explain and negotiate concepts, actions and intentions. [1] (Hanne Albrechtsen)

> **Datum** is the representation of concepts or other entities, fixed in or on a medium in a form suitable for communication, interpretation, or processing by human beings or by automated systems (Wellisch, 1996). **Information** is (1) a message used by a sender to represent one or more concepts within a communication process, intended to increase knowledge in recipients. (2) A message recorded in the text of a document. **Knowledge** is knowing, familiarity gained by experience; person's range of information; a theoretical or practical understanding of; the sum of what is known." [2] (Elsa Barber)

> **Data** is a symbol set that is quantified and/or qualified. **Information** is a set of significant sings that has the ability to create knowledge . . . The essence of the information phenomenon has been characterized as the occurrence of a communication process that takes place between the sender and the recipient of the message. Thus, the various concepts of information tend to concentrate on the origin and the end point of this communication process (Wersig & Neveling, 1975). **Knowledge** is information that has been appropriate by the user. When information is adequately assimilated, it produces knowledge, modifies the individual's mental store of information and benefits his development and that of the society in which he lives. Thus, as the mediating agent in the production of knowledge, the information, qualifies itself, in form and substance, as significant structures able to generate knowledge for the individual and his group." [3] (Aldo Barreto)

> **Data** are sensory stimuli that we perceive through our senses. **Information** is data that has been processed into a form that is meaningful to the recipient (Davis & Olson, 1985). **Knowledge** is what has understood and evaluated by the knower." [4] (Shifra Baruchson–Arbib)

> **Datum** is every thing or every unit that could increase the human knowledge or could allow to enlarge our field of scientific, theoretical or practical knowledge, and that can be recorded, on whichever support, or orally handed. Data can arouse information and knowledge in our mind. **Information** is the change determined in the cognitive heritage of an individual. Information always develops inside of a cognitive system, or a knowing subject. Signs that constitute the words by which a document or a book has made are not information. Information starts when signs are in connection with an interpreter (Morris, 1938). **Knowledge** is structured and organized information that has developed inside of a cognitive system or is part of the cognitive heritage of an individual (based on C. S. Peirce; Burks, 1958; Hartshorne & Weiss, 1931). [5] (Maria Teresa Biagetti)

> **Data.** The word "data" is commonly used to refer to records or recordings encoded for use in computer, but is more widely used to refer to statistical observations and other recordings or collections of evidence.

> **Information.** The word "information" is used to refer to a number of different phenomena. These phenomena have been classified into three groupings: (1) Anything perceived as potentially signifying something (e.g. printed books); (2) The process of informing; and (3) That which is learned

---

[1] I added the boldface in this section.

from some evidence or communication. All three are valid uses (in English) of the term "information." I personally am most comfortable with no. 1, then with no. 3, but acknowledge that others have used and may use no 2.

Knowledge. The word "knowledge" is best used to refer to what someone knows, which is, in effect, what they believe, including belief that some of the beliefs of others should not be believed. By extension the word "knowledge" is used more loosely for (1) what social groups know collectively; and (2) what is in principle knowable because it has been recorded somehow and could be recovered even though, at any given time, no individual knows (or remembers) it. [6] (Michael Buckland)

Data are the basic individual items of numeric or other information, garnered through observation; but in themselves, without context, they are devoid of information. **Information** is that which is conveyed, and possibly amenable to analysis and interpretation, through data and the context in which the data are assembled. **Knowledge** is the general understanding and awareness garnered from accumulated information, tempered by experience, enabling new contexts to be envisaged. [7] (Quentin L. Burrell)

Data are (or datum is) an abstraction. I mean, the concept of 'data' or 'datum' suggests that there is something there that is purely given and that can be known as such. The last one hundred years of (late) philosophic discussion and, of course, many hundred years before, have shown that there is nothing like 'the given' or 'naked facts' but that every (human) experience/knowledge is biased. This is the 'theory-laden' theorem that is shared today by such different philosophic schools as Popper's critical rationalism (and his followers and critics such as Kuhn or Feyerabend), analytic philosophy (Quine, for instance), hermeneutics (Gadamer), etc. Modern philosophy (Kant) is very acquainted with this question: experience ("Erfahrung") is a product of 'sensory data' within the framework of perception ("Anschauung") and the categories of reason ("Verstand") ("perception without concepts is blind, concepts without perception are void"). Pure sensory data are as unknowable as "things in themselves"."

Information is a multi-layered concept with Latin roots ('informatio' = to give a form) that go back to Greek ontology and epistemology (Plato's concept of 'idea' and Aristotle's concepts of 'morphe' but also to such concepts as 'typos' and 'prolepsis') (See Capurro, 1978; Capurro & Hjøerland, 2003). The use of this concept in information science is at the first sight highly controversial but it basically refers to the everyday meaning (since Modernity): "the act of communicating knowledge" (OED). I would suggest to use this definition as far as it points to the phenomenon of message that I consider the basic one in information science.

Message, information, understanding. Following systems theory and second-order cybernetics, I suggest to distinguish between 'message', 'information' and 'understanding.' All three concepts constitute the concept of communication (See, for instance, Luhmann, 1996, with references to biology (Maturana/Varela), cybernetics etc.). A 'message' is a 'meaning offer' while 'information' refers to the selection within a system and 'understanding' to the possibility that the receiver integrates the selection within his/her preknowledge—constantly open to revision i.e. to new communication—in accordance with the intention(s) of the sender. The receiver mutates each time into a sender.

Knowledge is 'no-thing' (contrary to "information-as-thing" as suggested by Michael Buckland, 1991a), i.e., it is the event of meaning selection of a (psychic/social) system from its 'world' on the basis of communication. The "act of communicating knowledge" (OED's definition of information) is then to be understood as the act of making a meaning offer (=message) leading to understanding (and misunderstanding) on the basis of a selection of meaning (=information). To know is then to understand on the basis of making a difference between 'message' (or meaning offer) and 'information' (or meaning selection). Human knowledge is, as Popper states, basically conjectural. Or, to put it in hermeneutic terms: understanding is always biased, i.e., based on (implicit) pre-understanding. In more classical terms we distinguish following Aristotle between 'empirical knowledge' (or 'know-how' = 'empeiria') and explicit knowledge (or 'know-that', for instance, scientific knowledge or 'episteme').

Data, information, knowledge. Putting the three concepts ("data," "information," and "knowledge") as done here, gives the impression of a logical hierarchy: Information is set together out of data and knowledge comes out from putting together information. This is a fairytale. [8] (Raphael Capurro)

Knowledge is that which is known, and it exists in the mind of the knower in electrical pulses. Alternatively, it can be disembodied into symbolic representations of that knowledge (at this point becoming a particular kind of information, not knowledge). Strictly speaking, represented knowledge is **information. Knowledge**—that which is known—is by definition subjective, even when aggregated to the level of social, or public, knowledge—which is the sum, in a sense, of individual "knowings." Data and information can be studied as perceived by and "embodied" (known) by the person or as found in the world outside the person..." [9] (Thomas A. Childers)

Data is the plural of datum, although the singular form is rarely used. Purists who remember their first-year Latin may insist on using a plural verb with data, but they forget that English grammar permits collective nouns. Depending on the context, data can be used in the plural or as a singular word meaning a set or collection of facts. Etymologically, data, as noted, is the plural of datum, a noun formed from the past participle of the Latin verb dare–to give. Originally, data were things that were given (accepted as "true"). A data element, **d**, is the smallest thing which can be recognized as a discrete element of that class of things named by a specific attribute, for a given unit of measure with a given precision of measurement (Rush & Davis, 2007; Landry & Rush, 1970; Yovits & Ernst, 1970).

Information. The verb 'inform' normally is used in the sense to communicate (i.e., to report, relate, or tell) and comes from the Latin verb informare, which meant to shape (form) an idea. Data is persistent while information is transient, depending on context and the interpretation of the recipient. Information is data received through a communication process that proves of value in making decisions.

Knowledge involves both data and the relationships among data elements or their sets. This organization of data based on relationships is what enables one to draw generalizations from the data so organized, and to formulate questions about which one wishes to acquire more data. That is, knowledge begets the quest for knowledge, and it arises from verified or validated ideas (Sowell, 1996). [10] (Charles H. Davis)

Data are symbols organized according to established algorithms.

Information represents a state of awareness (consciousness) and the physical manifestations they form. Information, as a phenomena, represents both a process and a product; a cognitive/affective state, and the physical counterpart (product of) the cognitive/affective state. The counterpart could range from a scratch of a surface, movement (placement)of a rock; a gesture(movement) speech(sound), written document, etc. (requirement). Information answers questions of what, where, when and who and permutations thereof.

Knowledge. Knowledge represents a cognitive/affective state that finds definition in meaning and understanding. Knowledge is reflected in the questions of "how" and "why." Knowledge extends the organism state of awareness (consciousness/ information). Knowledge can be given physical representation (presence) in the material products (technology) thereof (books, film, speech, etc.).

Message. Message is a medium through which data; information and knowledge are transmitted and used. It represents an instrument for moving the state of awareness and meaning with reference to specific events (states, conditions) from one implicit, or explicit source to another. When the physical products of awareness are transferred from one source to another, reference to the collective domain can be realized. [11] (Anthony Debons)

Raw data (sometimes called source data or atomic data) is data that has not been processed for use. [In the spirit of Tom Stonier's definition—Data: a series of disconnected facts and observations] Here "unprocessed" might be understood in a sense that no specific effort has been made to interpret or understand the data. They are the result of some observation or measurement process, which has been recorded as "facts of the world." The word data is the plural of Latin datum, "something given", which one also could call "atomic facts. **Information** is the end product of data processing. **Knowledge** is the end product of information processing. In much the same way as raw data are used as input, and processed in order to get information, the information itself is used as input for a process that results in knowledge.

Theory laden. It is very true that all data are theory laden. That does not mean that you can not produce new data which in the next step will lead to the theory revision, and that new, corrected theory will be the basis for producing new data which after a while will lead to the correction of the existing theory. We use our theory-laden data to refute theories!

Data-Information-Knowledge-Wisdom. According to Stonier (1993, 1997), data is a series of disconnected facts and observations. These may be converted to information by analyzing, cross-referring, selecting, sorting, summarizing, or in some way organizing the data. Patterns of information, in turn, can be worked up into a coherent body of knowledge.

Knowledge consists of an organized body of information, such information patterns forming the basis of the kinds of insights and judgments which we call wisdom. The above conceptualization may be made concrete by a physical analogy (Stonier, 1993): consider spinning fleece into yarn, and then weaving yarn into cloth. The fleece can be considered analogous to data, the yarn to information and the cloth to knowledge. Cutting and sewing the cloth into a useful garment is analogous to creating insight and judgment (wisdom). This analogy emphasizes two important points: (1) going from fleece to garment involves, at each step, an input of work, and (2) at each step, this input of work leads to an increase in organization, thereby producing a hierarchy of organization." [12] (Gordana.Dodig-Crnkovi)

Datum is a unique piece of content related to an entity. **Information** is the sum of the data related to an entity. [13] (Henri Jean-Marie Dou)

Data are a set of symbols representing a perception of raw facts (i.e., following Debons, Horne, & Cronenweth (1988), events from which inferences or conclusions can be drawn). **Information** is organized data (answering the following basic questions: What? Who? When? Where?). **Knowledge** is understood information (answering following basic questions: why?, how?, for which purpose?)." [14] (Nicolae Dragulanescu)

Data. Here, data typically means the "raw" material obtained from observation (broadly understood, but not necessarily, as "sense impressions," which is a key notion of empiricist philosophy). Such data is typically quantitative, presented in numbers and figures. [15] (Hamid Ekbia)

Prolog. These definitions are offered as an elaboration on physicist Heinz Pagels' (1988) observation:

"Information is just signs and numbers, while knowledge has semantic value. What we want is knowledge, but what we often get is information. It is a sign of the times that many people cannot tell the difference between information and knowledge, not to mention wisdom, which even knowledge tends to drive out." (p. 49, cited in O'Leary and Brasher, 1996, p. 262).

These distinctions in turn trace back at least as far as T. S. Eliot's lament:
Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
Choruses from the Rock

Data can be defined as a class of information objects, made up of units of binary code that are intended to be stored, processed, and transmitted by digital computers. As such, data consists of information in a narrow sense—i.e., as inscribed in binary code, units of data are not likely to be immediately meaningful to a human being. But units of data, as "informational building blocks," when collected and processed properly, can form information in the broader sense (see below), i.e., that is more likely to be meaningful to a human being (as sense-making beings).

Information. Collocations of data (information in the narrow sense—see above) that thereby become meaningful to human beings—e.g., as otherwise opaque units of binary code are collected and processed into numbers, artificial and

natural languages, graphic objects that convey significance and meaning, etc. Such collocations of data can be made meaningful by human beings (as sense-making beings) especially as such data collocations/information connect with, illuminate, and are illuminated by still larger cognitive frameworks—most broadly, worldviews that further incorporate knowledge and wisdom (see below). On this definition, information can include but is not restricted to data. On the contrary, especially as Borgmann (1999) argues, there are other forms of information (natural, cultural) that are not fully reducible to data as can be transmitted, processed, and/or produced by computers and affiliated technologies.

Knowledge is one step above information, and one step below wisdom. Knowledge in the broadest sense approaches a reasonably comprehensive worldview, i.e., a cognitive framework that establishes the major parameters and ten thousand details of human social and ethical realities, including basic values, beliefs, habits, notions of identity, relationships among human beings (including gender identity and issues) and relationships between humanity and larger realities (political, environmental, religious). Knowledge, however, can remain detached, objective, and thereby useless. Transforming cognitive forms of knowledge into ethical judgment and action is a primary task and goal of **wisdom** (see Dreyfus, 2001; Ess, 2003, 2004). [16] (Charles Ess)

Data are a string of symbols. **Information** is data that is communicated, has meaning, has an effect, has a goal. **Knowledge** is a personal/cognitive framework that makes it possible for humans to use information. [17] (Raya Fidel)

Data. It depends on your framework. If you are a Kantian, it is the foundation for the a priori categories of the understanding. If you are a computer programmer it is preprocessed information (data collected according to some algorithm for some purpose) or post-processed information (e.g., tables of such information). In this latter case data cannot be defined apart from information, because it is dependent on it. If you are a biologist, it might be stimuli, but these scientific approaches are built on a faulty understanding of perception (e.g., perception is sensations (i.e., stimuli) glued together—which is false).

Information is resources useful or relevant or functional for information seekers.

Knowledge. For some philosophers, validated, true information is that which coheres with other truths (coherence theory of truth). For others, what corresponds to reality (correspondence theory of truth). For others, it is what works or is functional (pragmatic theory of truth). At any event it is always contextual.

A lot of our so-called truths, knowledge, or known 'facts' are really orthodoxy—what we collectively believe at a certain point in time.[2] Today when someone would observe an unsupported object falling, when pushed for an explanation, they would utter the phrase/explanation: "the law of gravity." Unfortunately, it is an explanation that fails to explain—we still do not know what the "weak force" is, what gravity is, but we are taught in our so-called scientific approach, to utter a phrase that is supposed to—in the

---

[2]Plato differentiated between "doxa" (opinion) and "orthodoxa" (right or true opinion).

naming of it—to explain it. Four centuries back, it was attributed to the "will of God." Is this a worse explanation? Possibly. In both cases, we are living in images and metaphors and the orthodox frameworks of the time. Most reference collections in libraries are expressions of orthodoxies of various subject domains. [18] (Thomas J. Froehlich)

Data are representations of facts about the world. **Information** is data organized according to an ontology that defines the relationships between some set of topics. Information can be communicated. **Knowledge** is a set of conceptual structures held in human brains and only imperfectly represented by information that can be communicated. Knowledge cannot be communicated by speech or any form of writing, but can only be hinted at. [19] (H.M. Gladney)

Data is one or more kinds of energy waves or particles (light, heat, sound, force, electromagnetic) selected by a conscious organism or intelligent agent on the basis of a pre-existing frame or inferential mechanism in the organism or agent. **Information** is an organism's or an agent's active or latent inferential frame that guides the selection of data for its own further development or construction. **Knowledge** is one or more sets of relatively stable information. A **Message** is one or more inferred data sets gleaned from external or internal energetic reactions. [20] (Glynn Harmon)

Data are facts and statistics that can be quantified, measured, counted, and stored. **Information** is data that has been categorized, counted, and thus given meaning, relevance, or purpose. **Knowledge** is information that has been given meaning and taken to a higher level. Knowledge emerges from analysis, reflection upon, and synthesis of information. It is used to make a difference in an enterprise, learn a lesson, or solve a problem. [21] (Donald Hawkins)

Datum is smallest collectable unit associated with a phenomenon. Normally, data occur in collections that are collected in order to monitor a process, assess a situation, and/or otherwise gain a referent on a phenomenon. This does not mean that data are always defined, collected or used appropriately for the question in hand, but that that is the intention when doing so. They are building blocks, even if the building is engineered incorrectly.

Information. I would usually expect information to be an assessment or interpretation of data. Often information is not far removed from the 'smallest collectable unit' as I have defined "datum." But I expect it to be some abstraction from data.. Information does not inherently mean empirical or first hand analysis of data. It also does not guarantee correct interpretation of data although that is expected.

Knowledge is more subject, and intangible compared to information or data. It is what an individual takes from information and data, and what they incorporate into their beliefs, values, procedures, actions, etc. It is heavily internally oriented, understood completely only to the person possessing it. Much work around knowledge implies how to get the knowledge "out of" one head and in to another. Such transfer entails encoding knowledge into transferable information and decoding again into knowledge. Knowledge and information are not the same, but they feed from and support each other.

A message is the encoded information or codified/explicit knowledge that is disseminated to others. Very much a Shannon and Weaver transmission model, but I also consider

that encoding and decoding have a heavy personal, contextual and historical influence. [22] (Caroline Haythornthwaite)

Data are dynamic objects of cultural experience having the aspect of being meaning-neutral and a dual nature of description and instruction. **Information** is dynamic objects of cultural experience having the aspect of being belief-neutral and a dual nature of content and medium. **Knowledge** is dynamic objects of cultural experience having the aspect of being action-neutral and a dual nature of abstracting to and from the world. [23] (Ken Herold)

Data are the raw observations about the world collected by scientists and others, with a minimum of contextual interpretation. **Information** is the aggregation of data to make coherent observations about the world. **Knowledge** is the rules and organizing principles gleaned from data to aggregate it into information. [24] (William Hersh)

Data are observations and measurements you make on objects (artifacts, sites, seeds, bones) and on their contexts. Data are theory-laden.

Regarding the theory of knowledge organization we may say that knowledge is not organized by elements called data combined or processed according to some algorithmic procedure. What data are is domain specific and theory-laden. At the most general level what is seen as data is depending of the epistemological view that one subscribes to.

Information. The most fruitful theoretical view is here based on Karpatschof's interpretation of information and activity theory, AT (2000, p. 128). In order to define information, Karpatschof introduces the concept of release mechanisms, being systems having at their disposal a store of potential energy, the systems being "designed" to let this energy out in specific ways, whenever trigged by a signal fulfilling the specifications of the release mechanism. The signal that triggers a certain release mechanism is a low energy phenomenon fulfilling some release specifications. The signal is thus the indirect cause, and the process of the release mechanism the direct cause of the resulting reaction, which is a high-energy reaction compared to the energy in the signal. Information is thus defined as a quality by a given signal relative to a certain mechanism.

The release mechanism has a double function: (1) it reinforces the weak signal and (2) it directs the reaction by defining the functional value of a signal in the pre-designed system of the release mechanism. There has been a tendency to consider information to be an obscure category in addition to the classical categories of physics. Information is indeed a new category, but it cannot be placed, eclectically, beside the prior physical categories. Information is a category, not beside, but indeed above the classical categories of physics. Therefore, information is neither directly reducible to these classical categories, nor is it a radically different category of another nature than mass and energy.

Information is, in fact, the causal result of existing physical components and processes. Moreover, it is an emergent result of such physical entities. This is revealed in the systemic definition of information. It is a relational concept that includes the source, the signal, the release mechanism and the reaction as its reactants. The release mechanism is a signal processing system and an information processing system.

Information is thus defined in physical terms of signals, mechanisms and energy, but probably first arose with the biological world. Hjørland (2002) outlines the development of information processing mechanisms in the biological, the cultural and the social world.

Many professionals can claim to work with "the generation, collection, organization, interpretation, storage, retrieval, dissemination, transformation and use of information". This is not specific to information professionals. (Their specific work is discussed in Capurro & Hjørland (2003) and elsewhere). Hjørland (2000) investigates when and why the word "information" became associated with library schools (and thus knowledge organization) and what the theoretical implications in the shift from documents to information imply."

Knowledge. Different epistemologies (theories of knowledge) have different views on the nature of knowledge. I subscribe to the pragmatic theory of knowledge. The most important influence from pragmatic philosophy has been skepticism towards any claim of knowledge. A claim of knowledge should never be regarded as finally verified. It should just be regarded as just a claim. However, claims may be supported by empirical and logical arguments. Knowledge claims are parts of more comprehensive theories. Knowledge claims are not purely arbitrary. Instead of regarding science as a collection of true statements, it should be regarded as a collection of supported knowledge claims. In ordinary speech, knowledge then means that part of our background assumptions, that we do not find it fruitful to put questions to. [25] (Birger Hjorland)

Data are atomic facts, basic elements of "truth," without interpretation or greater context. It is related to things we sense. **Information** is a set of facts with processing capability added, such as context, relationships to other facts about the same or related objects, implying an increased usefulness. Information provides meaning to data. **Knowledge** is information with more context and understanding, perhaps with the addition of rules to extend definitions and allow inference. [26] (Donald Kraft)

Datum (in our sector mainly electronic) is the conventional representation, after coding (using ASCII, for example), of information. **Information** is knowledge recorded on a spatio-temporal support. **Knowledge** is the result of forming in mind an idea of something (Le Coadic, 2004). [27] (Yves François Le Coadic)

Data are commonly seen as simple, isolated facts, though products of intellectual activity in their rough shape. **Knowledge** is the appropriation of information in the process of learning, acting, interpreting. Knowledge is in the head of people, yet knowledge can be shared. Knowledge refers to the way information is used during the intellectual process. [28] (Jo Link-Pezet)

Data are formalized parts (i.e., digitalized contents) of sociocultural information potentionally proccessable by technical facilities which disregard the cognitive process and that is why it is necessary to provide them with meanings from outside (i.e., they are objective). **Information** is a relationship between an inner arrangement (i.e., a priori set structure (Šmajs & Krob, 2003), implicate order[3]) of a system and its

---

[3]The concepts of implicate and explicate orders are explained in Bohm (1980).

present embodiment in reality (explicate order) including mediating memory processes (i.e., historically dependent processes) releasing the meaning. **Knowledge** is tacitly or consciously grasped and interiorized content of information related and meaningfully integrated into a unifying frame of experience among other information contents interiorized in the same way, the complex of which reflects subjective understanding of environment. Mistakes arise from integration of misinformation or from integration of contradictory information into a unifying frame of experience (the second leads to cognitive dissonance and motivates to seek another information). [29] (Michal Lorenz)

Data are perceptible or perceived—if and when the signal can be interpreted by the 'user'—attributes of physical, biological, social or conceptual entities. **Information** is recorded and organized data that can be communicated (Porat & Rubin, 1977). However, it is advisable to distinguish between the various states or conditions of information (e.g. information-as an object (Buckland, 1991b), or semantic, syntactic and paradigmatic states (Menou, 1995). **Knowledge** is information that is understood, further to its utilization, stored, retrievable and reusable under appropriate circumstances or conditions. [30] (Michel Menou)

Data are sets of characters, symbols, numbers, and audio/visual bits that are represented and/or encountered in raw forms. Inherently, knowledge is needed to decipher data and turn them into information. **Information** is facts, figures, and other forms of meaningful representations that when encountered by or presented to a human being are used to enhance his/her understanding of a subject or related topics. *Knowledge* is a reservoir of information that is stored in the human mind. It essentially constitutes the information that can be "retrieved" from the human mind without the need to consult external information sources. [31] (Haidar Moukdad)

Data are raw material of information, typically numeric. **Information** is data which is collected together with commentary, context and analysis so as to be meaningful to others. **Knowledge** is a combination of information and a person's experience, intuition and expertise. [32] (Charles Oppenheim)

Datum is an object or crude fact perceived by the subject, non-constructed nor elaborated in the consciousness, without passing through neither analysis processes nor evaluation for its transfer as information. **Information** is a phenomenon generated from knowledge and integrated therein, analyzed and interpreted to achieve the transfer process of message (i.e., meaningful content) and the cognitive transformations of people and communities, in a historical, cultural and social context. **Knowledge** is a social and cognitive process formed by the passing or assimilated information to thought and to action. **Message** is the meaningful content of information. [33] (Lena Vania Pinheiro)

Data are primitive symbolic entities, whose meaning depend on it integration within a context that allow their understanding by an interpreter. **Information** is the intentional composition of data by a sender with the goal of modifying the knowledge state of an interpreter or receiver. **Knowledge** is the intelligent information processing by the receiver and it consequent incorporation to the individual or social memory (Belkin & Robertson, 1976; Blair, 2002) [34] (Maria Pinto)

Signs. The distinctive feature of signs is that they denote something, regardless of whether that something exists or does not exist, is concrete or abstract, possible or impossible, a thing or an event, a substance or a determination, an individual or a collective. Analysis even of one single sign leads to a multiplicity of signs and their denoted items. For this reason, we may say that the sign contains a reference to both the denoted item considered per se, in isolation, and the contexts or situations in which the denoted item appears. And of these of especial importance are those that, for lack of better terminology, we can call the proximal context and the distal context. The proximal context is the net of relations that hold among the items denoted by signs. On the other hand, the distal context is the outcome of a categorization procedure. Its most usual form is that constituted by the reply to questions like 'what is this?', where acceptable replies are of the type 'this is an animate being', 'this is an artifact', 'this is a property', etc. This codification of the two types of context enables me to propose the following distinction between data and information.

Datum. Def. 1. x is a datum = x is a sign that denotes entities or attributes in a proximal context. In the light of this definition one understands why conventional analyses of consistency and integrity, or procedures of normalization, are effective techniques for the organization and rationalization of data. From a technological point of view, relational databases are the currently most advanced products available for the efficient handling of data.

Information. Def. 2. x is an item of information = x is a datum in a distal context. Definition 2 tells us that information is made up of more structured items. That is to say, information is the embedding of signs-in-a-proximal-context (i.e., data) in a distal context. Information, thus, adds greater structure to data. These definitions provide a first explanation for the scant interest aroused by proposals to draw more exact distinctions between data and information. In effect, in concrete cases of application, it is often difficult to distinguish precisely between distal and proximal contexts.

Conditions of knowledge. Knowledge is apparently not reducible solely to information and data. The problem is to understand 'what is lacking', what must be added to information and data in order to achieve true knowledge. My claim is that the meaning of a sign is given by the position of the sign in a field of signs (in a space). On the other hand, the content of a sign is given by the position of the item (denoted by the sign) in a field of items. Data, information, meanings and contents cover the field of knowledge. This amounts to saying that we have knowledge when we know (1) which item is denoted by which sign, (2) the item's proximal context, (3) the item's distal contexts, (4) the sign's position in the field of signs, (5) the item's position in the field of items (Poli, 2001).

Data, information, knowledge, message. I am unable to understand why data, information, knowledge and message are placed on the same level of analysis. I would suggest considering message as the "vehicle" carrying either data or information (which can be taken as synonymous). Knowledge hints to either a systematic framework (e.g., laws, rules or regularities, that is higher-order "abstractions" from data) or what somebody or some community knows ("I know that you are married"). In this latter sense knowledge presents a "subjective" side. [35] (Roberto Poli)

Data are a representation of facts or ideas in a formalized manner, and hence capable of being communicated or manipulated by some process. So: data is related to facts and machines (Holmes, 2001). Information is the meaning that a human assigns to data by means of the known conventions used in its representation. Information is related to meaning and humans (Holmes, 2001). [36] (Ronald Rousseau)

Datum is a quantifiable fact that can be repeatedly measured. **Information** is an organized collection of disparate datum. **Knowledge** is the summation of information into independent concepts and rules that can explain relationships or predict outcomes. [37] (Scott Seaman)

Data are raw evidence, unprocessed, eligible to be processed to produce knowledge. **Information** is the process of becoming informed; it is dependent on knowledge, which is processed data. **Knowledge** perceived, becomes information. Knowledge is what is known, more than data, but not yet information. Recorded knowledge may be accessed in formal ways. Unrecorded knowledge is accessible in only chaotic ways. [38] (Richard Smiraglia)

Data are discrete items of information that I would call facts on some subject or other, not necessarily set within a fully worked out framework. **Information** is facts and ideas communicated (or made available for communication). **Knowledge** is the considered product of information. Selection as to what is valid and relevant is a necessary condition of the acquisition of knowledge. [39] (Paul Sturges)

Data are facts that are the result of observation or measurement. (Landry et al., 1970). **Information** is meaningful data. Or data arranged or interpreted in a way to provide meaning. **Knowledge** is internalized or understood information that can be used to make decisions. [40] (Carol Tenopir)

Data are unprocessed, unrelated raw facts or artifacts. **Information** is data or knowledge processed into relations (between data and recipient). **Knowledge** is information scripted into relations with recipient experiences. [41] (Joanne Twining)

Data are representations of facts and raw material of information. **Information** is data organized to produce meaning. **Knowledge** is meaningful content assimilated for use. The three entities can be viewed as hierarchical in terms of complexity, data being the simplest and knowledge, the most complex of the three. Knowledge is the product of a synthesis in our mind that can be conveyed by information, as one of many forms of its externalization and socialization. [42] (Anna da Soledade Vieira)

Data are alphabetic or numeric signs, which without context do not have any meaning. **Information** is a set of symbols that represent knowledge. Information is what context creates/gives to data. It is cognitive. Normally it is understood as a new and additional element in collecting data and information for planned action. **Knowledge** is enriched information by a person's or a system's own experience. It is cognitive based. Knowledge is not transferable, but through information we can communicate about it. (Note that the highest level of information processing is the generation of wisdom, where various kinds of knowledge are communicated and integrated behind an action. [43] (Irene Wormell)

Data are artifacts that reflect a phenomenon in natural or social world in the form of figures, facts, plots, etc. **Information** is anything communicated among living things. It is one of the three mainstays supporting the survival and evolution of life, along with energy and materials. **Knowledge** is a human construct, which categorize things, record significant events, and find causal relations among things and/or events, etc. in a systematic way. [44] (Yishan Wu)

Last but not least, here are my reflections on data, information, and knowledge:

Inferential propositional knowledge. In traditional epistemology, there are three main kinds of knowledge: practical knowledge, knowledge by acquaintance, and propositional knowledge (Bernecker & Dretske, 2000). Practical knowledge refers to skills (i.e., functional abilities, such as driving a car). Knowledge by acquaintance is direct nonmediated recognition of external physical objects and organisms (e.g., "this is Albert Einstein"), or the direct recognition of inner phenomena (e.g., pain, hunger). Propositional knowledge usually comes in the form of 'knowing that''; S (subject) knows that P (proposition). It is the reflective and/or the expressed content of what a person thinks that he or she knows. Note that the contents of our reflective and/or expressed thoughts are in the form of propositions. Propositional knowledge is divided into inferential and non-inferential knowledge. Non-inferential propositional knowledge refers to direct intuitive understanding of phenomena (e.g., ""This is a true love"). Inferential knowledge is a product of inferences, such as induction and deduction. The field of information science, as well as any academic field, is composed of inferential propositional knowledge, as they are published in articles and books. This analysis is focused on defining "data," "information," and "knowledge" as they are related and implemented in inferential propositional knowledge.

Subjective versus objective realms. Data (D), information (I), and knowledge (K) phenomena have two distinctive modes of existence; namely, in the subjective and in the objective realms. Correspondingly, we differentiate between subjective knowledge and objective knowledge. Note that *subjective knowledge* is equivalent here to the knowledge of the subject or the individual knower, and objective *knowledge* is equivalent here to knowledge as an object or a thing. Subjective knowledge exists in the individual's internal world (i.e., as a thought), whereas objective knowledge exists in the individual's external world (e.g., as it is published in books, presented in digital libraries, and stored in electronic devices). In this context, they are not related to arbitrariness and truthfulness, which are usually attached to the concepts of subjective knowledge and objective knowledge. To avoid confusion, I will use the terms *universal knowledge* and *collective knowledge* (i.e., knowledge in the collective realm) rather than objective knowledge. The distinction between subjective knowledge and universal knowledge differs from the distinction between private knowledge and public knowledge. *Private knowledge* is the individual's intimate knowledge. These are thoughts on contents known only to the individual, such as

intimate dreams and feelings, and "hidden agenda" (i.e., hidden goals and incentives). *Public knowledge* refers to thoughts that the individual consider as knowledge, and they are on contents known to other people as well (e.g., "2 + 2 = 4," "Paris is the capital of France").

Six distinctive concepts. Having established the distinction between the subjective and the universal domains, we are in a position to define the three key concepts data, information, and knowledge. In fact, we have six concepts to define, divided into two distinctive sets of three. One set relates to the subjective domain, and the other—to the universal domain.

D-I-K in the subjective domain. In the subjective domain, **data** are the sensory stimuli, which we perceive through our senses. **Information** is the meaning of these sensory stimuli (i.e., the empirical perception). For example, the noises that I hear are data. The meaning of these noises (e.g., a running car engine) is information. Still, there is another alternative as to how to define these two concepts—which seems even better. **Data** are sense stimuli, or their meaning (i.e., the empirical perception). Accordingly, in the example above, the loud noises, as well as the perception of a running car engine, are data. **Information** is empirical knowledge. Accordingly, in the example above, the knowledge that the engine is now on and the car is leaving is information, since it is empirically based. Information is a type of knowledge, rather than an intermediate stage between data and knowledge. **Knowledge** is a thought in the individual's mind, which is characterized by the individual's justifiable belief that it is true. It can be empirical and non-empirical, as in the case of logical and mathematical knowledge (e.g., "every triangle has three sides"), religious knowledge (e.g., "God exists"), philosophical knowledge (e.g., "Cogito ergo sum"), and the like. Note that *knowledge* is the content of a thought in the individual's mind, which is characterized by the individual's justifiable belief that it is true, while "knowing" is a state of mind which is characterized by the three conditions: (1) the individual believe that it is true, (2) S/he can justify it, and (3) It is true, or it is appear to be true.

D-I-K in the universal domain. In the universal domain, data, information, and knowledge are human artifacts. They are represented by empirical signs (i.e., signs that one can sense through his/her senses). They can take on diversified forms such as engraved signs, painted forms, printed words, digital signals, light beams, sound waves, and the like. Universal data, universal information, and universal knowledge mirror their cognitive counterparts. Meaning, in the objective domain **data** are sets of signs that represent empirical stimuli or perceptions, **information** is a set of signs, which represent empirical knowledge, and **knowledge** is a set of signs that represent the meaning (or the content) of thoughts that the individual justifiably believes that they are true.

Signs Versus meaning. Defining the D-I-K phenomena as sets of signs needs to be refined. There is a fundamental distinction between documented (i.e., written, spoken, or physically expressed) propositions and meanings. "E = MC²", "$\underline{E = MC^2}$", and "**E = MC²**" are not three different types of knowledge. These are three different sets of signs that represent the same meaning. In other words, they are three different utterances of the same knowledge. Knowledge, in the collective domain, is the meaning that is represented by written and spoken statements (i.e., sets of symbols). However, because we cannot perceive with our senses the meaning itself, which is an abstract entity, we can relate only to the sets of signs (i.e., written, spoken, or physically expressed propositions), which represent it. Apparently, it is more useful to relate to the data, information, and knowledge as sets of signs rather than as meaning and its building blocks (Zins, 2004, 2006). [45] (Chaim Zins)

Conceptual Approaches

Delimitations

Before presenting the analysis of the panel's definitions let me clarify the methodological considerations that guide me while analyzing the panel's diversified definitions. First, words can be misleading. Definitions are theory-laden. They can best be analyzed and evaluated in the context of the relevant theory. For this very reason, the definitions are grouped here according the contributing scholars rather than the defined concepts. The scholar-based organization facilitates a better understanding of the rationale and the interrelations among the three concepts, as they are understood and defined by each scholar.

Second, many of the 45 citations reflect systematic and comprehensive thinking and are based on solid theoretical and philosophical foundations. However, a few are incomplete, inconsistent, logically faulty, and philosophically problematic. For this very reason, the study is focused on mapping the theoretical issues that we face while formulating coherent conceptions of data, information, and knowledge, and the conceptual approaches to resolve them, rather than on evaluating the accuracy, adequacy, and coherency of the panel's definitions.

Anthropological Document

Forty-five scholars (including the researcher) shared their thoughts and formulated about 130 definitions. This collection of definitions is an invaluable "anthropological document" that documents the conceptions of D-I-K, as they are understood by leading scholars in the information science academic community. Again, the definitions are rooted in diversified theoretical grounds. Many of them reflect systematic and comprehensive thinking and are based on solid theoretical and philosophical foundations. A few, though, are incomplete, inconsistent, logically faulty, and philosophically problematic. Apparently, the definitions show that the academic community speaks in different languages. Still, they provide the basis for mapping the various conceptual approaches for defining data, information, and knowledge in the context of information science.

Metaphysical Versus Nonmetaphysical Approaches

The most basic distinction is between metaphysical and nonmetaphysical approaches. Metaphysical approaches refer

to data, information, or knowledge as metaphysical phenomena. They reflect metaphysical postulates, such as "knowledge is eternal," and "knowledge is an independent entity/object," as well as religious beliefs, such as "God knows. . . ." Obviously, for Information Science, all the panel members unanimously implement nonmetaphysical approaches. Metaphysical approaches, though, emerged mainly in theoretical reviews (see, for example, citation [8]).

Human Exclusive Versus Nonexclusive Approaches

We zoom into nonmetaphysical approaches. Nonmetaphysical approaches are divided into those exclusively centered on humans and those that ascribe the D-I-K phenomena to non-human biological (e.g., animals and plants) and/or to physical (e.g., planets, robots,) phenomena as well. Citation [20] exemplifies a non human-exclusive approach by using the phrase "organism or intelligent agent" rather than "person or human." Apparently, nearly all the panel members adopt human-exclusive approaches for defining D-I-K in the context of information science.

Human-Centered Approaches

We zoom into human-exclusive approaches. Three classifications emerge as highly relevant. The first classification is between cognitive exclusive versus nonexclusive approaches. The second classification is between "propositional" exclusive versus nonexclusive approaches. The third classification is between the subjective domain versus the objective, or rather universal domain.

Cognitive-based exclusive versus nonexclusive approaches. Human-centered approaches are divided into those that refer to D-I-K exclusively as cognitive phenomena and those that refer to D-I-K in terms of cognitive, biological, or physical phenomena, *mutatis mutandis.* The division between cognitive-based exclusive approaches and nonexclusive approaches emerges when one compares, for example, Hjørland's definition of information (see citation [25])" with Poli's definition of information (see citation [35]). Hjørland defines information in terms of biological mechanism and signals, while Poli defines information in terms of signs and meanings.

Note that the term *cognitive approaches* should be refined to *cognitive-based approaches* because it applies to human thoughts and states of mind as well as to the human artifacts that represent them (e.g., books, digital signals). Debons' definition of information exemplifies a cognitive-based approach. According to Debons, "Information represents a state of awareness (consciousness) and the physical manifestations they form" (citation [11]). Albrechsten's definition of information (see citation [1]) also exemplifies a cognitive-based approach because the contents of databases gain their status of information by relating to "meaning and human intention." Another example is Harmon's definition of data (see citation [20]). The first part ("Data is one or more kinds of

energy waves or particles (light, heat, sound, force, electromagnetic)") creates the deceptive impression that it exemplifies a physical-based approach for defining data. However, the second part of the definition ("selected by a conscious organism or intelligent agent on the basis of a pre-existing frame or inferential mechanism in the organism or agent.") makes Harmon's definition an example of a cognitive-based approach. Apparently, nearly all the panel members adopt cognitive-based approaches for defining D-I-K in the context of information science.

Propositional exclusive versus nonexclusive approaches. We zoom into cognitive-based approaches. It seems that nearly all the panel's definitions presented above explicitly or implicitly reflect propositional conceptions, although only I (see citation [45]) specifically use the term *propositional knowledge.* The concept of propositional conceptions originated from the distinction among various types of knowledge (i.e., practical knowledge, knowledge by acquaintance, and propositional knowledge, inferential and noninferential). Although the panel did not specifically refer to the various types of knowledge, a distinction should be made between focusing on propositional knowledge as against dealing with all types of knowledge. Propositional conceptions are those conceptions that refer to D-I-K exclusively in the form of propositions and their building blocks. Apparently, it seems that the propositional conceptual approaches are embedded in the cognitive approaches.

The Mainstream of the Field

At this point, we can characterize the most common conceptual approach for defining data, information, and knowledge in the context of information science. Undoubtedly, the most common conceptual approach that represents the mainstream of the field is characterized as the non-metaphysical, human-centered, cognitive-based, propositional approach.

Models for defining D-I-K. We zoom into non-metaphysical, human-centered, cognitive-based, propositional approaches. The third division, which is the division between the subjective domain (SD) versus the universal domain (UD), establishes the theoretical ground for formulating generic models for defining D-I-K. The division between D-I-K in the subjective domain, namely, D-I-K as inner phenomena bound in the mind of the individual knower versus the universal domain (UD), namely, D-I-K as external phenomena to the mind of the individual knower was presented earlier in this article (see citation [45], for further discussion see Zins, 2004, 2006). The different combinations of D-I-K phenomena in the universal domain and in the subjective domain establish the ground for formulating different models. By analyzing the 45 citations one can realize that in most citations data are characterized as phenomena in the universal domain, and knowledge is characterized as phenomena in the subjective domain, though in many cases these interpretations are not exclusive. This significantly limits

Model 1		Model 2		Model 3		Model 4		Model 5	
<u>UD</u>	<u>SD</u>	<u>UD</u>	<u>SD</u>	<u>UD</u>	<u>SD</u>	<u>UD</u>	<u>SD</u>	<u>UD</u>	<u>SD</u>
D		D		D		D	D	D	D
I			I	I	I	I	I	I	I
	K		K	K	K		K	K	K

FIG. 1.   Four models for defining data (D)–information (I)–knowledge (K).

the number of optional models for defining D-I-K that can present the mainstream of the field. To be precise, I spotted at least five different models (see Figure 1).

1. The first model is **UD**: D-I; **SD**: K; meaning: D-I are external phenomena; K are internal phenomena. This model is the most common one. The model is implemented in citations [17], [40], and [43]. It underlies the rationale of the name *Information Science*; that is, Information Science is focused on exploring data and information, which are seen external phenomena. It does not explore knowledge, which are seen as internal phenomena.
2. The second model is **UD**: D; **SD**: I-K; meaning: D are external phenomena; I-K are internal phenomena. Citations [5] and [20] exemplify the model.
3. The third model is **UD**: D-I-K; **SD**: I-K; meaning: D are external phenomena; I-K phenomena can be in both domains, external or internal. Citation [6] exemplifies the model.
4. The fourth model is **UD**: D-I; **SD**: D-I-K; meaning: D-I phenomena can be in both domains, external or internal; K phenomena are internal. Citation [1] exemplifies the model.
5. The fifth model is **UD**: D-I-K; **SD**: D-I-K; meaning: D-I-K phenomena can be in both domains, universal (i.e., external) or subjective (i.e., internal). Citations [11] and [45] exemplify the model.

The reader may refine my analysis, and may revise my interpretation of the exemplary citations. Still, formulating comprehensive and systematic definitions of data, information, and knowledge requires reflection on these two domains (S-U) and their key role in shaping our conceptions on these three constitutive concepts (D-I-K) of information science.

A Concluding Remark

This study maps the major issues on the agenda of scholars engaged in exploring and substantiating the foundations of Information Science. Conceptual approaches were identified and formulated for defining data, information, and knowledge. This might help the reader to a better understanding of the issues and the considerations involved in establishing the foundations of Information Science; however, by no means does it replace the personal quest to ground one's positions on solid theoretical foundations

Acknowledgments

References

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

Bernecker, S., & Dretske, F. (Eds.). (2000). Knowledge: Readings in contemporary epistemology. Oxford: Oxford University Press.

Blair, D.C. (2002). Knowledge management: Hype, hope or help? Journal of the American Society for Information Science and Technology, 53(12), 1019–1028.

Bohm, D. (1980). Wholeness and the implicate order. New York: Routledge & Kegan Paul.

Borgmann, A. (1999). Holding on to reality: The nature of information at the turn of the millennium. Chicago: University of Chicago Press.

Buckland, M. (1991a). Information and information systems. New York: Greenwood Press.

Buckland, M. (1991b). Information as thing. Journal of the American Society of Information Science, 42(5), 351–360.

Capurro, R. (1978). Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs [A contribution to the etymological and conceptual history of the concept of information]. München: Saur Verlag.

Capurro, R., & Hjørland, B. (2003). The concept of information. Annual Review of Information Science and Technology, 37(8), 343–411.

Davis, G.B., & Olson, M.H. (1985). Management information systems. New York: McGraw Hill.

Debons, A., Horne, E., & Cronenweth, S. (1988). Information science: An integrated view. New York: G.K. Hall.

Dreyfus, H. (2001) On the Internet. New York: Routledge.

Ess, C. (2003). Liberal arts and distance education: Can Socratic virtue ($\alpha\rho\varepsilon\tau\varepsilon$) and Confucius' exemplary person (junzi) be taught online? Arts and Humanities in Higher Education, 2(2), 117–137.

Ess, C. (2004). Computing in philosophy and religion. In S. Schreibman, R.G. Siemens, & J. Unsworth (Eds.), A companion to digital humanities (pp. 132–142). Oxford: Blackwell.

Hjørland, B. (2000). Documents, memory institutions, and information science. Journal of Documentation, 56(1), 27–41.

Hjørland, B. (2002). Principia informatica. Foundational theory of information and principles of information services. In H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), Emerging frameworks and methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4) (pp. 109–121). Greenwood Village, CO: Libraries Unlimited.

Holmes, N. (2001). The great term robbery. Computer, 34(5), 94–96.

Karpatschof, B. (2000). Human activity—Contributions to the anthropological sciences—From a perspective of activity theory. Copenhagen: Dansk Psykologisk Forlag.

Landry, B.C., Mathis, B.A., Meara, N.M., Rush, J.E., & Young, C.E. (1970). Definition of some basic terms in computer and information science, Journal of the American Society for Information Science, 24(5), 328–342.

Landry, B.C., & Rush, J.E. (1970). Toward a theory of indexing II. Journal of the American Society for Information Science 21, 358–367.

Le Coadic, Y.F. (2004). La science de l'information [Information science] Collection Que sais-je? (No. 2873). Paris: PUF.

Luhmann, N. (1996). Soziale Systeme. Frankfurt am Main: Suhrkamp.

Menou, M.J. (1995). The impact of information (Part 2): Concepts of information and its value. Information Processing and Management, 31(4), 479–490.

Morris, C.W. (1938). Foundations of the theory of signs. Chicago: The University of Chicago Press.

O'Leary, S.D., & Brasher, B.E. (1996). The unknown God of the Internet: Religious communication from the ancient agora to the virtual forum. In C. Ess (Ed.), Philosophical perspectives on computer-mediated communication (pp. 233–269). Albany, NY: State University of New York Press.

Pagels, H. (1988). The dreams of reason. New York: Simon and Schuster.

Peirce, C.S. (1931). Collected papers of Charles Sanders Peirce. C. Hartshorne & P. Weiss (Eds.) (Vol. I–VI). Cambridge, MA: Harvard University Press.

Peirce, C.S. (1958). Writings of Charles S. Peirce. A chronological edition. A.W. Burke (Ed.) (Vol. VII–VIII). Bloomington: Indiana University Press.

Poli, R. (2001). ALWIS. Ontology for knowledge engineers. Unpublished doctoral dissertation, University of Utrecht, the Netherlands.

Porat, M.V., & Rubin, M. (1977). The information economy: Definition and measurement (OT Special publication, Vol. 1, pp. 77–120). Washington DC: Office of Telecommunications, U.S. Department of Commerce.

Rush, J.E., & Davis, C.H. (2006). Guide to information science and technology. Manuscript in preparation.

Šmajs, J., & Krob, J. (2003). Evoluční ontologie [Evolutionary ontology]. Brno: Masaryk University.

Sowell, T. (1996). Knowledge and decisions. New York: Basic Books.

Stonier, T. (1993). The wealth of information. London: Thames/Methuen.

Stonier, T. (1997). Information and meaning—An evolutionary perspective. Berlin: Springer.

Wellisch, H.H. (1996). Abstracting, indexing, classification, thesaurus. construction: A glossary. Port Aransas, TX: American Society of Indexers.

Wersig, G., & Neveling, U. (1975). Terminology of documentation: A selection of 1200 basic terms. Paris: The UNESCO Press.

Yovits, M.C., & Ernst, R.L. (1970). Generalized information systems: Consequences for information transfer. In H.B. Pepinsky (Ed.), People and information (pp. 1–31). Elmsford, NY: Pergamon Press.

Zins, C. (2004). Knowledge mapping: An epistemological perspective. Knowledge Organization, 31(1), 49–54.

Zins, C. (2006). Redefining information science: From information science to knowledge science. Journal of Documentation, 62(4), 447–461.

Zins, C. (2007a). Conceptions of information science. Journal of the American Society for Information Science and Technology, 58(3), 335–350.

Zins, C. (2007b). Knowledge map of information science. Journal of the American Society for Information Science and Technology, 58(4), 526–535.

Zins, C. (in press). Classification schemes of information science: Twenty-eight scholars map the field. Journal of the American Society for Information Science and Technology.

Appendix A

The Panel

Dr. Hanne Albrechtsen, Institute of Knowledge Sharing, Copenhagen, Denmark; **Prof. Elsa Barber**, University of Buenos Aires, Argentina; **Prof. Aldo de Albuquerque Barreto**, Brazilian Institute for Information in Science and Technology, Brazil; **Prof. Shifra Baruchson–Arbib**, Bar Ilan University, Ramat-Gan, Israel; **Prof. Clare Beghtol**, University of Toronto, Canada; **Prof. Maria Teresa Biagetti**, University of Rome 1, Italy; **Prof. Michael Buckland**, University of California, Berkeley, CA; **Mr. Manfred Bundschuh**, University of Applied Sciences, Cologne, Germany; **Dr. Quentin L. Burrell**, Isle of Man International Business School, Isle of Man; **Dr. Paola Capitani**, Working Group Semantic Web, Italy; **Prof. Rafael Capurro**, University of Applied Sciences, Stuttgart, Germany; **Prof. Thomas A. Childers**, Drexel University, Philadelphia, PA; **Prof. Charles H. Davis**, Indiana University; **Prof. Anthony Debons**, University of Pittsburgh, Pittsburgh, PA; **Prof. Gordana Dodig-Crnkovic**, Mälardalen University, Västerås/Eskilstuna, Sweden; **Prof. Henri Dou**, University of Aix-Marseille III, France; **Prof. Nicolae Dragulanescu**, Polytechnics University of Bucharest, Romania; **Prof. Carl Drott**, Drexel University, Philadelphia, PA; **Prof. Luciana Duranti**, University of British Columbia, Canada; **Prof. Hamid Ekbia**, University of Redlands, Redlands, CA; **Prof. Charles Ess**, Drury University, Springfield, MO; **Prof. Raya Fidel**, University of Washington, Seattle, WA; **Prof. Thomas J. Froehlich**, Kent State University, Kent, OH; **Mr. Alan Gilchrist**, Cura Consortium and TFPL, London, UK; **Dr. H.M. Gladney**, HMG Consulting, McDonald, PA; **Prof. Glynn Harmon**, University of Texas at Austin, Austin, TX; **Dr. Donald Hawkins**, Information Today, Medford, NJ; **Prof. Caroline Haythornthwaite**, University of Illinois at Urbana Champaign, Urbana, IL; **Mr. Ken Herold**, Hamilton College, Clinton, NY; **Prof. William Hersh**, Oregon Health & Science University, Portland, OR; **Prof. Birger Hjorland**, Royal School of Library and Information Science, Copenhagen, Denmark; **Ms. Sarah Holmes,*** the Publishing Project, USA; **Prof. Ian Johnson,*** the Robert Gordon University, Aberdeen, UK; **Prof. Wallace Koehler**, Valdosta State University, Valdosta, GA; **Prof. Donald Kraft**, Louisiana State University, Baton Rouge, LO; **Prof. Yves François Le Coadic**, National Technical University, Lyon, France; **Dr. Jo Link-Pezet**, Urfist, and University of Social Sciences, France; **Mr. Michal Lorenz**, Masaryk University in Brno, Czech Republic; **Prof. Ia McIlwaine**, University College, London, UK; **Prof. Michel J. Menou**, Knowledge and ICT management consultant, France; **Prof. Haidar Moukdad**, Dalhousie University, Halifax, Nova Scotia, Canada; **Mr. Dennis Nicholson**, Strathclyde University, UK; **Prof. Charles Oppenheim**, Loughborough University, Leicestershire, UK; **Prof. Lena Vania Pinheiro**, Brazilian Institute for Information in Science and Technology, Brazil; **Prof. Maria Pinto**, University of Granada, Spain; **Prof. Roberto Poli**, University of Trento, Italy; **Prof. Ronald Rousseau**, KHBO, and University of Antwerp, Belgium; **Dr. Silvia Schenkolewski–Kroll**, Bar Ilan University, Ramat-Gan, Israel; **Mr. Scott Seaman,*** University of Colorado, Boulder, CO; **Prof. Richard Smiraglia**, Long Island University, Brookville, NY; **Prof. Paul Sturges**, Loughborough University, Leicestershire, UK; **Prof. Carol Tenopir**, University of Tennessee, Knoxville, TN; **Dr. Joanne Twining**, Intertwining.org, a virtual information

---

*An observer (i.e., those panel members who did not strictly meet the criteria for the panel selection and terms of participation).

consultancy, USA; **Prof. Anna da Soledade Vieira**, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **Dr. Julian Warner**, Queen's University of Belfast, UK; **Prof. Irene Wormell**, Swedish School of Library and Information Science in Borås, Sweden; **Prof. Yishan Wu**, Institute of Scientific and Technical Information of China (ISTIC), Beijing, China.

Appendix B

Excerpts From the Three Questionnaires on Data, Information, and Knowledge

Knowledge Map of Information Science: Issues,
Principles, Implications
(First Round, December 15, 2003)
. . .
3: Data, Information, Knowledge

Three related concepts, "data," information," and "knowledge," emerge in the context of information science. The academic and professional IS literature supports diversified meanings for each concept. We begin the conceptual analysis by trying to define these concepts.

Data. Data (the plural form of the Latin word *datum*, which means "the given").

Question 3.1
What are "data"? (Please, define the concept; Refer to theoretical background. Thanks)

Answer 3.1
Data are (or datum is). . .

Information.

Question 3.2
What is "information"? (Please define the concept; refer to theoretical background)

Answer 3.2
Information is. . .

Knowledge.

Question 3.3
What is "knowledge"? (Please define the concept; refer to theoretical background.)

Answer 3.3
Knowledge is. . .

Interrelations. "Data," "information," and "knowledge" are interrelated. Discussions among scholars focus on the nature of the relations among these key concepts, as well as on their meanings.

Sequential order. Many scholars claim that data, information, and knowledge are part of a sequential order. Data are the raw material for information, and information is the raw material for knowledge. However, if this is the case, then "information science" should explore data (information's building blocks) and information, but not knowledge, which is an entity of a higher order.

Nevertheless, it seems that information science does explore knowledge since it includes two sub-fields, "knowledge organization", and "knowledge management". I am confused. Should we refute the sequential order? Should we change the name of the field from "Information Science" to "Knowledge Science"? Or should we perhaps exclude the fields of knowledge organization and knowledge management from information science?

Question 3.4 Are data, information, and knowledge part of a sequential order?

(Please explain.) **If yes, please explain how it is that "knowledge organization" and "knowledge management" are sub-fields of information science?**

Answer 3.4

Knowledge vs. information. Another common view is that knowledge is not conveyed by information. Knowledge is the product of a synthesis in our mind. If this is the case, we should exclude the fields of knowledge organization and knowledge management from information science. Besides, is Albert Einstein's famous equation "$E = MC^2$" (which is printed on my computer screen) information or knowledge? Is "$2 + 2 = 4$" information or knowledge?

Question 3.5 Is knowledge not conveyed by information? (Please explain and elaborate).

Answer 3.5

Synonyms. The alternative view that "information" and "knowledge" are synonyms is problematic too. If "information" and "knowledge" are synonyms, should not we use the term "knowledge science" rather than "information science"?

Question 3.6 Are "information" and "knowledge" synonyms? If yes, how do you explain the name "information science"? (Please explain and elaborate.)

Answer 3.6

The researcher's views. At this point, I present my conceptions to the panel. If you want to receive a detailed paper, please contact me.

Propositional knowledge. In traditional epistemology there are three kinds of knowledge: practical knowledge (i.e., skills), knowledge by acquaintance (i.e., knowing a

person or a thing), and propositional knowledge (i.e., in the form of propositions). Propositional knowledge is divided into inferential and non-inferential. Inferential knowledge is a product of inferences, such as induction and deduction. We are zooming in on inferential propositional knowledge. **Information science, like all academic fields, is composed of inferential propositional knowledge.**

Two approaches. There are two basic approaches to define "knowledge": in the subjective domain (i.e., as a thought in the subject's mind) and in the objective domain (i.e., as an object). Note that the terms "subjective" and "objective" are **not** used here as we use them in our daily life. "Subjective" means 'existing in the mind' (not 'arbitrary'). "Objective" means 'existing as an independent object (or a thing)' (not 'unbiased').

The subjective domain. The first approach conditions the knowledge in the individual's (or subject's) mind. Knowledge is a thought. It is characterized as "a justified true belief." Generally, we can identify subjective propositional knowledge by the certainty of the individual that his/her own thoughts are true, and by his/her ability to base this certainty on a sound justification (e.g., experiments, observations, and logical inferences).

(Note that in the subjective realm "knowledge" is the <u>content</u> of a justified true thought, while "knowing" is the <u>state of mind</u> that is characterized by three conditions: justification, belief, and truth.)

The objective domain. The second approach ascribes an independent objective existence to knowledge. Knowledge is a collection of concepts, arguments, and rules of inference. They are true and exist independently of the subjective knowledge of the knowing individual. This is the case, for example, of arguments published in books.

The field of information science, like any academic field, is composed of objective propositional knowledge, as it is recorded, documented, and represented in the professional and the academic literature. This is what we explore and map in this collective research enterprise.

Mutual dependency. Paradoxically, the subjective and the objective domains are complementary. On the one hand, objective knowledge is the product of outputting (externalizing, recording, or documenting) subjective knowledge. (One might say, "this questionnaire is an output of my brain".) On the other hand, <u>the realization of objective knowledge necessitates the consciousness of at least one individual knower.</u> This is crucial. The term "objective domain" is equivalent here to "collective domain". Objective knowledge is collective, in the phenomenological sense, not in the metaphysical sense.

Six concepts. Having established the distinction between the subjective and the objective domains, we have six concepts to define, divided into two distinctive sets of three. One set relates to the subjective domain, the other to the objective (i.e., collective) domain.

The subjective domain. In the subjective domain, "data" and "information" acquire two alternative meanings. The first option:

"**Data**" are the sensory stimuli that we perceive through our senses. "**Information**" is the meaning of these sensory stimuli (i.e., the empirical perception). Example: The noises that I hear are data. The meaning of these noises, for example, a running car engine, is information.

The second option (which I personally prefer):

"**Data**" are the sense stimuli, or their meaning (i.e., the empirical perception). Accordingly, in the example above the perception of a running car engine, as well as the noises of a running car engine, are data.

"**Information**" is empirical knowledge. Accordingly, in the example above the knowledge that the engine is <u>now</u> on is information, since it is empirically based. As one can see, <u>information is a type of knowledge (i.e., empirical knowledge), rather than an intermediate stage between data and knowledge.</u>

"**Knowledge**", as mentioned above, is a thought in the individual's mind, which is characterized by the individual's justifiable belief that it is true. It can be empirical (e.g., "It is a rainy day") and non-empirical, as in the case of logical and mathematical knowledge (e.g., "Every triangle has three sides"), religious knowledge (e.g., "God exists"), philosophical knowledge (e.g., "Cogito ergo sum"), and the like.

The objective domain. Objective data, objective information, and objective knowledge mirror their cognitive counterparts. They are represented by empirical symbols, and can have diversified forms such as engraved signs, painted forms, printed words, digital signals, light beams, sound waves, and the like.

"**Data**" are sets of symbols that represent empirical perceptions.

"**Information**" is a set of symbols that represent empirical knowledge.

"**Knowledge**" is a set of symbols that represent thoughts that the individual justifiably believes are true.

Question 3.7
Do you accept these conceptions? If you have comments, observations, or critical reflections, please share them with the panel. Thanks.
Answer 3.7

Question 3.8
If you have different and elaborate conceptions, please share them with the panel. Thanks.
Answer 3.8

. . .
Knowledge Map of Information Science:Issues, Principles, Implications
(Third Round, October 8, 2004)
. . .
2: Data, Information, Knowledge, Message

Data, Information, Knowledge, Message. In the first and the second rounds I received **20 + pages** of definitions.

While analyzing the responses I found inconsistencies among the definitions, the conceptions of IS and the IS classification schemes (see section 4).

However, if one analyzes the 20 + pages, one can identify and formulate several distinct models for defining each of these four key concepts.

Question 2.1 If you want to revise your definitions, please do so. Thanks.

Answer 2.1
A. Data is..
B. Information is..
C. Knowledge is..
A. Message is..

. . .
7: Selected Responses

In this section I present selected responses on various topics.

I received hundreds of detailed answers. I will relate to all of them in future publications. Evidently, I can present here only few of your invaluable contributions.

Ad hoc definitions.
7.1
A comment on "ad hoc definitions": not all data in my view are empirical perceptions. For example within computer science "input data" can be anything like: names, numbers (totally unrelated to any empirical perceptions, like series of prime numbers and similar).

Raw data (sometimes called source data or atomic data) is data that have not been processed for use. They might be the result of empirical perceptions as well as chosen sets of symbols which are to be processed to obtain some kind of information. An example is the computer program input data. They can be any set of symbols chosen for "information processing". Here we explain the concept "data" with another concept "information processing" that is not defined. Circularity in definitions seems to be unavoidable. Within computer science circularity (or recursivity) is acceptable as long as it ends somewhere in some trivial base case. To find the analogous situation for those definitions one might need more time. . .

If "Information" is a set of symbols that represent **empirical knowledge**, so that information is knowledge representation, we use the concept of knowledge that is higher order to explain information that is a more basic concept. Again there must be a way to include non-empirical knowledge. Information derived from some empirical knowledge might still be only an information, and nevertheless non-empirical. Again within computer science there is an abundance of examples of usage of the term "information" not meaning any directly empirical knowledge.

It seems to me that the way of usage defines what is to be seen as data /information/ knowledge. I can imagine that the same set of symbols can play any of those roles depending on the usage.

Researcher's comment: Thank you for your clarifications. I will elaborate the ad hoc definitions in the final analysis of the panel responses.

Information & truth.
7.2
I wonder if in the definition of "information" you have any constraints on the truth value of information. Sometimes claims of the necessity of the strong definition of information are made (Luciano Floridi), i.e. the information must necessarily be TRUE in order to qualify as information. How do you view that question?

Researcher's comment: Since "information" is defined here as empirical knowledge, and "knowledge" is defined as "a justified true belief," information must be perceived by the informed person, at the relevant time, to be true. Evidently, s/he might be wrong.

Clarifications. Finally, I would like to clarify several issues raised by the panel.

1. **Popper's World 3.** I am not "Popperian." In fact, I am a phenomenologist. Generally, I follow Edmund Husserl's phenomenology.
2. **Subjectivity vs. Objectivity.** We always know the objective through our subjective mind. Meaning is formed subjectively by individuals.
3. **Symbols vs. meaning.** There is a fundamental distinction between documented (i.e., written, spoken or physically expressed) propositions and meaning. "$E = MC_2$", " $\underline{E = MC^2}$", and "**E = MC²**" are not three different 'knowledges' (pl. of knowledge). These are three different sets of symbols (or characters) that represent the same meaning. In other words, these are three different utterances of the same knowledge. Knowledge, in the collective domain, is the meaning, which is represented by written and spoken statements (i.e., sets of symbols). However, since we cannot perceive with our senses the meaning itself, we can relate only to the sets of symbols, which represent the meaning. Note, however, that although the knower ascribes a universal status to the meaning, s/he cannot be certain if it really exists outside his/her own mind [As I noted above, I am not "Popperian." Actually, I hold an agnostic position: 'I don't know']. Apparently, it is more fruitful to define "D", "I", "K", and "M" as sets of symbols rather than as meanings.

Question 7.1 If you have critical reflections on the responses, please let me know.

Thanks.

Answer 7.1