

December 2003

A Brief Historical Review of the Development of the Distinction between Data and Information in the Information Systems Literature

Robert Gray
Christopher Newport University

Follow this and additional works at: <http://aisel.aisnet.org/amcis2003>

Recommended Citation

Gray, Robert, "A Brief Historical Review of the Development of the Distinction between Data and Information in the Information Systems Literature" (2003). *AMCIS 2003 Proceedings*. 372.
<http://aisel.aisnet.org/amcis2003/372>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A BRIEF HISTORICAL REVIEW OF THE DEVELOPMENT OF THE DISTINCTION BETWEEN DATA AND INFORMATION IN THE INFORMATION SYSTEMS LITERATURE

Robert L. Gray
Christopher Newport University
rlgray@cnu.edu

Abstract

It is a commonplace among contemporary information systems professionals that the concepts of data and information are obviously distinct and clearly understood. Through a review of the historical literature, this paper shows that, in fact, the distinction is not obvious, that it is an outgrowth of work in the information systems area, and that the distinction is not clearly understood. The paper briefly notes some of the implications of this historical development for information systems theory. This paper will be of interest primarily to academics and those with an interest in the conceptual foundations and theoretical frameworks guiding Information Systems research.

Introduction: The Need for Clarity in Fundamental Concepts

It is a shared set of fundamental concepts that provide the epistemic foundations of a discipline—a shared framework for understanding, investigation, and communication. Clarity in, and general agreement about, the definitions of fundamental concepts is therefore critical to the discipline. As an academic discipline, however, Information Systems is peculiar; investigation suggests that there is, in fact, surprisingly little agreement about the meanings of a surprisingly large number of fundamental terms. This may be troublesome. For example, it has been suggested that lack of attention to basic concepts may hinder communication within the discipline (Alter, 2000). Lack of consensus about the meanings of fundamental terms may be what has led one team of researchers to describe the IS discipline, following Whitley (1984), as a “fragmented adhocracy” (Banville & Landry, 1989).

Within the discipline, few concepts would appear to be more fundamental than those of “data” and “information,” yet it has been observed that there is much confusion even about these core concepts (Checkland & Holwell, 1998). Despite the general lack of consensus or precision in the definitions of “data” and “information,” one characteristic that seems to be almost universally assumed is that the concepts are distinct. How they are distinguished depends on how they are individually defined, but it is almost universally agreed that “data” (whatever they may be) and “information” (whatever it is) are somehow different, i.e., that the terms have different referents. Some further assert not just that the terms are distinct but that understanding the distinction is important, even “vital” to understanding the discipline (Dock & Wetherbe, 1988; Mingers, 1995; Jessup & Velacich, 1999).

This paper offers a brief historical review of the development of the distinction between “data” and “information.” Although most IS researchers now take the distinction to be natural and obvious, a review of the very early literature suggests that it is not. The distinction is, in fact, an artifact, or a development of, work in the field of IS itself or in what was known to an earlier generation of researchers as “data processing.”

Scope and Significance of this Historical Review

This paper is preliminary to a broader review of the core concepts of data and information as they are actually articulated and used within the discipline. It is possible to consider those concepts from any number of different and equally valid theoretical or philosophical perspectives. As a philosopher, the author was originally trained in analytic and ordinary language philosophy, and

it is this approach to conceptual analysis rather than a body of theory that guides this analysis. Partly for reasons of space, partly because it is a perspective that has value in its own right, this paper restricts itself to the early historical development of the distinction between data and information within the IS literature. On the basis of that historical analysis, it also attempts to provide insight into several issues or questions surrounding these core concepts as follows:

1. It provides some insight into the early historical development of the discipline and its core concepts. Understanding its common historical and conceptual heritage helps to solidify and define the discipline. It has been argued that building such a cumulative tradition is essential to the development of a coherent discipline (Keen, 1980). Although there is a substantial body of historical research on the development of the computer and other information technologies, relatively little work has been done to understand the historical development of those concepts that define the Management Information Systems discipline (See Cougar, 1973; Dickson, 1981)
2. It provides a partial basis for understanding and refining the definitions of the core concepts of data and information. A cursory review of the contemporary literature will show that
 - there is no single, commonly accepted set of definitions for these terms;
 - the definitions are often articulated in ways that seem to lack conceptual rigor;
 - the definitions do not conform to ordinary usage.

It is not essential that technical definitions of common terms conform to standard usage, although when technical usage departs from common usage there should be some justification for it, if only to avoid unnecessary confusion. However, it seems *a priori* that rigor and precision in the definitions of foundation concepts are desirable characteristics in any discipline.

3. When there is a lack of agreement about the definitions of fundamental terms within a discipline, one approach to resolving the issue may be to ask to what problems or concerns the language responds. That is the approach taken in this paper. In this case, we ask “What were the concerns or observations that led earlier researchers to make the distinction and to define the terms as they did?” Answering this question will not by itself resolve contemporary definitional issues; time and technological progress may have vitiated those early concerns, but the distinction, or the manner in which it is made, might continue to be valid for different reasons.
4. Such an historical review may also have practical significance. The author is attempting in a follow-up paper to show that how this distinction is articulated provides a conceptual basis for a set of frameworks or “metaphors” for understanding the IS enterprise. Thus, for example, some ways of articulating the distinction lead naturally to the understanding of IS as a production system whose raw materials are data and whose product is information; others lead to a conceptualization of the IS enterprise that is more appropriate to a distribution system. Clearly, such different metaphors may support different perceptions of the role of the IS function within an organization and may have an impact in directing research.

In the recent literature, the distinction between data and information has been extended to include “knowledge” as a third category. Some authors actually include a fourth category, “wisdom” (Post & Anderson, 2000). Those concepts go beyond the scope of this research; the interested reader should see Alavi & Leidner’s (2001) review of the concept of “knowledge” as it is used in knowledge management research. Their review also briefly discusses the conceptual distinctions between data, information, and knowledge.

Data Versus Information: An Historical Perspective

MIS is often described (frequently to excuse some apparent or claimed deficiency) as a new or immature discipline. In fact, as a blending of interests in business systems and computer-based information technologies, a stream of literature leading to what would today be identified as MIS can be traced back for approximately fifty years, beginning very soon after the development of the contemporary stored program computer architecture. Much of this early literature is speculative. Much of it was practical and related the work of early systems designers, frequently from what were then referred to as Systems and Procedures Departments.

Given the fundamental place of such concepts as “data” and “information” in the information systems enterprise—at least from a theoretical perspective, the natural assumption might be that the distinction was obvious at the outset. However, a review of

the earliest texts suggests that no such distinction was made. What is most interesting about this is that this lack does not appear to have seriously impeded the investigation or development of information systems.

For example, Richard Canning's 1956 textbook on *Electronic Data Processing for Business and Industry* (Canning, 1956) makes no such distinction. In fact, it makes little use of the word "data" although it frequently speaks of "information" and "manipulations upon information."

A similar usage occurs in the accounting firm Haskins and Sells' early *Introduction to Data Processing by Electronics* (1955). This very early work, which was directed primarily to accounting professionals, uses "data" and "information" entirely interchangeably—sometimes in ways that seem unnatural to those of us who are familiar with contemporary usage, referring for example to "input information." Haskins and Sells' next book on the subject, *Introduction to Data Processing*, published in 1957 also uses the two words interchangeably; however, the Foreword includes the following intriguing assertion: "Data originates in the human mind. Data is information—a piece of intelligence." This seems to affirm that data and information are, in fact, the same. This view was not original or unique.

Initially copyrighted in 1955, Ned Chapin's highly influential textbook, *An Introduction to Automatic Computers*, is possibly the earliest textbook on business computing. This work appears to explicitly deny any distinction at all between "data" and "information." Chapin makes no distinction in the body of the text, using the two terms more or less interchangeably. The glossary, however, contains the following separate entries for "data" and "information:"

"Data—The same as information."

"Information—The same as data (see also dictionary definition of information)." (Chapin, 1957)

Thus, what may be the earliest relevant texts in the MIS/EDP literature appear to explicitly disavow any distinction between "data" and "information."

Interestingly, in a later paper, Chapin defines data differently as "the thing processed in doing data processing work....Data...are anything which takes the form of symbols" (Chapin, 1968). He does not use the term "information" in this paper, but he is careful to point out that data is distinct from "meaning" which, he says, is a "private personal matter" and that data processing operations do not add meaning to the symbols.

For the most part, other writers from the same period also seem to use the two terms interchangeably, making no explicit distinction between them (e.g., Nett & Hetzler, 1959; Burton and Mills, 1960), although the language can, nevertheless, appear surprisingly modern. For example, in an early book directed primarily to business executives on management uses of computers, Kozmetzy and Kirchner write that "An integrated business system links the event that originates an item of information with the events that occur wherever and whenever someone uses this information. The data flows can be traced directly..." (Kozmetzky & Kircher, 1956)

In another early work directed to executives, William Bell offered what was, for the time, a careful definition of "data" as "any numeric or alphabetic (sometimes called 'alpha-numeric') material." However, he offers this definition, apparently, only as a step toward a definition of "Data-processing" and makes no distinction between data and information. (Bell, 1957).

On the other hand, sometimes a distinction appears to be nascent in the language even if it is not explicitly made. For example, Laubach, in a work apparently written in 1955 reporting on a very early (1954) summary of corporate uses of ADP, writes that "What has been said about input methods and storage of *raw data* also applies to output methods and the means of storing *processed data*....If reports are to be prepared from outputs, or processed data, then devices have to be available to translate magnetic tape language into *readable information*." (Laubach, 1957, emphasis added)

The earliest evidence of an explicit distinction between data and information that I have so far been able to find comes from a text by Howard Levin on office automation. Levin clearly did not feel that there was any inherent distinction between data and information. Further, he explicitly noted that he used the words interchangeably and that "this usage will be generally satisfactory." However, he also suggested that

there may be value in carrying our concepts one step further by *constructing* a distinction between information and data. The characteristic that separates the two is *usefulness*. *Information is useful data*. Information is data that increases our knowledge with respect to achieving objectives. (Levin, 1956, emphasis added)

Whether the distinction in the subsequent data processing and information systems literature originated here is not certain; however, Levin clearly considered himself to be *constructing* a distinction, not simply acknowledging existing usage or reporting on an observation. It is also worth noting that the distinction he makes is one of utility. Information is just that subset of data that happens to be useful.

Early texts can have a substantial influence on the development of a discipline and Levin's very early construction may account for the tendency of so many subsequent writers to distinguish data and information at least partly in terms of utility. However, it would not account for the other very common notion that information is somehow derived from data as a product from raw materials, although it is worth noting that Levin also used the raw materials versus finished product analogy in the same text, observing that "Information can be considered both the raw material and finished product of office operations." Levin's purpose in this instance was just to illustrate the fundamental importance of information in business management, not to draw a distinction between data and information.

I have been unable to find another reference to a distinction between "data" and "information" in the early literature for several years—but it arises next in a most interesting way. As part of his discussion of the cost and value of data and information in a collection of lectures delivered at Dundee Technical College in mid-1958, Robert Gregory offered the following definitions:

'Data' can be defined as facts that are a matter of direct observation. Original raw data are sorted, summarized, and used to update files in order to yield processed data.

'Information' means raw or processed data that are new, accurate, timely, and useful for decision making or for controlling business operations. (Gregory, 1960)

Although these definitions would be acceptable without change to any number of contemporary writers in the MIS discipline, they do not appear to have been so uncontroversial at the time, and it appears that they may actually have occasioned some discussion at the conference. For example, the editor, J. H. Leveson, writes in the preface that:

[A] common terminology has been used. One particular difference of terminology, however, has been left unresolved. This concerns the terms 'data' and 'information.' On the one hand, Dr. [S. A.] Gill states...that 'information' is a term meaning 'any recognizable message regardless of content' and suggests that this meaning of the term is now 'well established in the theory of communication'.....Dr. [Robert H.] Gregory...defines 'information' as 'data that are new, timely, understandable, and related to some problem that a manager can deal with.' The term 'data' is also defined. It means 'facts and figures that are observed, known or available.' 'Data', he says, 'are raw material for a processing system. Information is wanted as the output.' (Leveson, 1960, p.15)

The discussion at this conference is interesting, because we next see the distinction being sharply made in a 1959 symposium on Management Information and Control Systems. The published proceedings includes the discussions that followed each of the presentations. The discussion following a paper by Alan Rowe includes a debate between Rowe and Robert Gregory which is fascinating in context of this historical investigation (Malcom & Rowe, 1960, pp. 294-295):

R. H. Gregory: The other point that has disturbed me a long while is the interchangeable use of "data" and "information." Do you differentiate between the two? Would you care to define them?

A. J. Rowe: I may be able to clarify the difference by referring to an example. Knowing that part number XYZ is in some stage of completion, I would consider information rather than data.

R. H. Gregory: I will describe my thoughts in a workable definition. "Data" means raw facts, all the observations you care to make, the whole file. From the data available, those parts that meet certain criteria, such as timeliness, unusualness, and relevance (and relevance hasn't been touched upon here), we select a set of data which I prefer to call "information." Perhaps you prefer merely to call it a subset of the data that is timely, has worthiness, or is relevant to a situation.

A. J. Rowe: I would like to characterize information in relationship to the action to be taken, and I think that is how you have characterized it. In some cases it might be the raw data itself. In other cases it might be summarization of data, but the criterion is that it be the basis for action. Isn't that what you would say?

R. H. Gregory: I am tempted to consider the element of further usage, but one encounters the definitional problem similar to the physicist's question: Can there be *noise* in the desert but no *sound* because no one heard it. You say information exists only if the data are used.

A.J. Rowe: I am not saying one actually has to take action, but I am saying information is what is needed to take action.

R. H. Gregory: Your definitions are quite particular to the environment; there is information to the extent there is a decision maker that could make use of it. What if that decision maker does not exist?

A. J. Rowe: Well then, I would not call it information.

Gregory's later textbook is probably the first influential textbook to clearly define data and information differently and to emphasize the importance of the distinction. Writing in the Preface, the authors observe that "At the heart of this book is a distinction between data and information: *data are the raw material from which management must distill information.*" (Gregory & Van Horn, 1960, p. v, emphasis added) They subsequently give the following actual definitions: "The word 'data' might be used to cover all the facts that are obtained. Another word, such as 'information' is needed to cover the particular facts that management wants to know." (Gregory & Van Horn, 1960, p. 4) This appears to have been a popular text; it was in its fourth printing by 1962.

It would thus appear that Gregory actively championed the distinction and that we may owe much of our current tendency to distinguish the two concepts partly to his publications.

This manufacturing analogy, viewing data as, in effect, raw materials which are converted into information or from which information is derived is often repeated and frequently employed in the subsequent literature. It has had a substantial influence on the development of management information systems theory. For example, the idea that "the basic responsibility" of a management information system is "the conversion of *data* into *information*" appears very shortly after the publication of Gregory's book in James Gallagher's early text on management information systems (Gallagher, 1961).

The distinction did not catch on immediately with everyone, however. Most textbooks published in the 1960's did not make a distinction (e.g., Awad, 1965; Brooks & Iverson, 1963; Elliott & Wasley, 1965; Davis, 1965; Dearden & MacFarlan, 1966) although at least one mid-1960's textbook observed that a distinction based on usefulness was "often made:"

a distinction is often made between data and information, namely, that data is the raw material from which information is derived. According to this concept, the significant characteristic that separates data from information is usefulness..... The conversion of data to information is a primary function of data processing. (Arnold *et al*, 1966)

In their 1963 text, Schmidt and Meyers offer the following definition of "data" but make no distinction between data and information.

"Historically the word 'data' meant counts and measures that arose in statistical investigations. With the advent of the computer the word 'data' has been further extended to include any kind and any form of information processed."

One distinction that sometimes appears in the modern literature is between data as input and information as output (Licker, 1997). In a textbook prepared in 1963 by a group of industry representatives under the auspices of the Data Processing Management Association this distinction appears as "Facts (data) are manipulated (processed) to create information (output) which provides answers to specific problems." (Awad & DPMA, 1966) On this view, the essential distinction between information and data lies in the process: the one is output, or created, from the other.

The notion that information conveys or somehow consists in "meaning" is often found in the contemporary literature (Boland, 1987; Checkland & Holwell, 1998; McLeod & Schell, 2001). This notion also has precursors in the early literature. The idea that "Information is data...recorded, classified, organized, related or interpreted within context to convey *meaning*" appears in a monograph by Blumenthal (1969; emphasis added). At about the same time, Dippel & House (1969) noted that "it is important to understand the distinction between data as the raw material and information as the end product," observing that

data are not useful until they have been introduced to some processing activity for further evaluation and analysis. Such processing transforms data into information. By definition, then, data are not information until they have been processed to a point where they are useful and meaningful to the information-system user.

This observation appears to combine several notions: information is the *product* of *processing* input data to add or create *utility* and *meaning*. Absent any of those three elements, data would not, apparently, provide information.

The distinction between data and information appears to have become established as a basic tenet of information systems theory during this time frame. Thereafter, most introductory texts that the author has examined contain some reference to a distinction based on some combination of the elements of usefulness, context, process, and meaning.

Conclusion

Although most IS researchers now make a distinction between data and information and take that distinction to be natural, obvious and somehow fundamental to the discipline, what this brief historical review shows is that the distinction is neither necessarily natural nor obvious. It is an outgrowth and a deliberate construct of early work in the fields of information systems and data processing.

This observation has historic interest, but it may have other research implications as well. If there is a lack of agreement about the definitions of fundamental terms within a discipline, one approach to resolving the issue would be to ask to what problems or concerns the language responds. In this case, we would ask what were the concerns or observations that led earlier researchers to make the distinction. What we find in this case is that there does not appear to have been any clearly formulated rationale. Levin introduced the distinction between data and information only as a possible convenience, a linguistic construct that *may be* of value, but without any referent or purpose that would allow us by ostention or by analysis to give precision to the distinction or to correct errors in subsequent usage. Even if the distinction were originally made only as a possible linguistic convenience, it does not follow that it has no real significance. Having been made and accepted in common parlance, the distinction forms a framework in terms of which subsequent theory may be articulated, so that its ultimate influence may be far greater than its original motivation would warrant.

Later writers took up the distinction but defined it differently or with a slightly different emphasis and ultimately attached far greater significance to it than was originally intended. In any case, we can see in these early writings the beginnings of several distinct themes that presumably lead to the different definitions in use today.

Stipulated definitions cannot be said to be correct or incorrect. Writers are free to use words as they wish. However, it may be worthwhile as a matter of future research to examine the definitions of “data” and “information” that are in contemporary use as well as alternatives that have been proposed by researchers in the area (e.g., Hirschheim, et al, 1995) to determine whether those definitions are themselves useful, meaningful, and helpful or whether there might be some more generally acceptable set of definitions that would be more useful in providing a foundation and a framework for increasing our understanding and knowledge of information systems.

References

- Alavi, Maryam and Leidner, Dorothy E. “Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues.” *MIS Quarterly* (25:1), 2001, pp. 107-136.
- Alter, Steven. “Same Words, Different Meanings: Are Basic IS/IT Concepts our Self-Imposed Tower of Babel?” *Communications of the AIS* (3:10), 2000.
- Arnold, Robert R., Hill, Harold C., and Nichols, Aylmer V. *Introduction to Data Processing*. Wiley, New York, 1966..
- Awad, Elias M. *Business Data Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1965.
- Banville, C. and Landry, M. “Can the IS field be Disciplined?” in R. Galliers (ed.), *Information Systems Research: issues, methodology and practical guidelines*, Blackwell, Oxford, 1992
- Bell, William D. *A Management Guide to Electronic Computers*. McGraw-Hill, New York, 1957.
- Blumenthal, Sherman C. *Management Information Systems: A Framework for Planning and Development*. Prentice-Hall, Englewood Cliffs, NJ, 1969.

- Boland, Richard J., Jr. "The In-Formation of Information Systems in R. J. Boland, Jr. and R. A. Hirschheim (eds.) *Critical Issues in Information Systems Research*, John Wiley & Sons, 1987.
- Brooks, Frederick P., Jr. and Iverson, Kenneth E. *Automatic Data Processing*, John Wiley, New York, 1963.
- Burton, A. J. and Mills, R. G. *Electronic Computers and their Business Applications*, Ernest Benn, London, 1960.
- Canning, Richard G. *Electronic Data Processing for Business and Industry*, John Wiley & Sons, New York, 1956.
- Chapin, Ned. *An Introduction to Automatic Computers*, Van Nostrand Co., Princeton, NJ, 1957.
- Chapin, Ned. "A Deeper Look at Data," *Proceedings of the ACM National Conference*, 1968, pp. 631-638.
- Cougar, J.D. "Evolution of Business System Analysis Techniques," *ACM Computing Surveys* (5,3), 1973, pp. 167-198
- Davis, Gordon B. *An Introduction to Electronic Computers*, McGraw-Hill, New York, 1965.
- Dickson, Gary W. "Management Information Systems: Evolution and Status" in Marshall C. Yovits (ed.), *Advances in Computers*, vol. 20., Academic Press, New York, 1981
- Dock, V. Thomas and Wetherbe, James C. *Computer Information Systems for Business*, West Publishing, St. Paul, MN, 1988.
- Elliott, C. Orville and Wasley, Robert S. *Business Information Processing Systems*. Irwin, Homewood, IL, 1965.
- Checkland, Peter and Holwell, Sue *Information, Systems, and Information Systems: Making Sense of the Field*, John Wiley & Sons, Chichester, 1998.
- Gallagher, James D. *Management Information Systems and the Computer*, AMA, New York, 1961.
- Gregory, R. H. "Economics of Computers," Chapter 13 in J. H. Leveson (ed.), *Electronic Business Machines*, Philosophical Library, New York, 1960, pp. 177-197.
- Gregory, Robert H. and Van Horn, Richard L. *Automatic Data-Processing Systems: Principles and Procedures*, Wadsworth Publishing Co., Belmont, CA, 1960.
- Haskins & Sells. *Data Processing by Electronics*, Haskins & Sells, 1955.
- Haskins & Sells. *Introduction to Data Processing: An Outline of Basic Data Processing Operations and Methods*, Haskins & Sells, 1957.
- Hirschheim, Rudy, Klein, Heinz K. and Lyytinen, Kalle. *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations*, Wadsworth Publishing Co., Cambridge, 1995.
- Jessup, Leonard M and Valacich, Joseph S. *Information Systems Foundations*, Que, Indianapolis, IN, 1999
- Keen, Peter G. W. "MIS Research: Reference Disciplines and a Cumulative Tradition" in E McLean (ed.), *Proceedings of the First International Conference on Information Systems*, Philadelphia, PA, 1980, pp. 9-18.
- Kozmetsky, George and Kircher, Paul. *Electronic Computers and Management Control*, McGraw-Hill, New York, 1956.
- Laubach, Peter B. *Company Investigations of Automatic Data Processing*, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1957.
- Levin, Howard S. *Office Work and Automation*, John Wiley & Sons, New York, 1956.
- Leveson, J. H. *Electronic Business Machines*, Philosophical Library, New York, 1960.
- Licker, Paul S. *Management Information Systems: A Strategic Leadership Approach*, Dryden Press, Fort Worth, TX, 1997.
- Malcom, Donald G., Rowe, Alan J. and McConnell, Lorimer F. (eds.), *Management Control Systems*, Wiley, New York, 1960.
- Mingers, J. C. "Information and Meaning: Foundations for an Intersubjective Account," *Information Systems Journal* (5), 1995, pp. 285-306
- Nett, Roger and Hetzler, Stanley A. *An Introduction to Electronic Data Processing*, The Free Press, Glencoe, IL, 1959.
- Post, Gerald V. and Anderson, David L. *Management Information Systems: Solving Business Problems with Information Technology* (2nd ed.), Irwin McGraw-Hill, 2000
- Schmidt, Richard N. and Meyers, William E. *Electronic Business Data Processing*. Holt, Rinehart and Winston, New York, 1963.
- Whitley, R. "The Development of Management Studies as a Fragmented Adhocracy," *Social Science Information* (23), 1984, pp. 775-818.