"The editors have done a remarkable job of integrating the expertise of a wide range of interdisciplinary contributors ... The book is a model of clarity ... an indispensable textbook for students and information professionals alike."

—W. Boyd Rayward

Introduction to Information Science and Technology

Edited by **Charles H. Davis** and **Debora Shaw**

asis&t
ASIST Monograph Series

ogy

ıses people to look
: it. Chapters 4 and
tive representation
an identify sources
uters and network
e technologies are
ɔles of information
Chapter 10, which
:cted by the organ-
w information and
;ocial communica-
ɔn access and use;
pproach these and
:al issues and con-
:ation studies and

of information the
ıts a tiny portion of
l the rest. We hope
lly changing world

CHAPTER 1

Our World of Information

1.1. How Much Information?

Information is everywhere and in huge amounts. How much is there (can we find out)? Where does it come from? And how does all that information affect us as individuals? What can we do to find out what it's worth while providing some level of organization and control? This introductory chapter places each of us, as information producers and users, into the big picture.

In 2008, researchers Roger Bohn and James Short at the University of California–San Diego's Global Information Industry Center asked "How much information was consumed by individuals in the United States?" (2009, p. 8). They looked at only nonwork use of information, such as watching television or talking on a cell phone. Among their conclusions are the following:

- Each American spends, on average, half of each day of (11.8 hours) consuming information.
- Although we spend 41 percent of our "information time" in front of the TV, TV provides less than 35 percent of the bytes of information we consume.
- Computer and video games, because of their graphics, account for 55 percent of the information bytes we consume at home.
- Altogether, we gobbled up some 3.6 zettabytes of information at home in 2008.

How much is a zettabyte? It is 10^{21} bytes, or 1,000 exabytes. Bohn and Short (2009, p. 8) estimate that an exabyte, or 1 billion gigabytes, is the capacity of all the hard disks in home computers in Minnesota (population 5.1 million). So the nonwork information consumed in the U.S. in 2008 was equal to what could be stored on 3,600 Minnesotas' worth of hard drives. In other words, if all this information were "printed ... in books and stacked ... as tightly as possible across the United States including Alaska, the pile would be 7 feet high" (p. 13). Bohn and Short also found that radio is "a highly byte-efficient delivery mechanism." People listened to radio for 19 percent of their hours spent consuming information—this amounted to 10.6 percent of daily words received but only 0.3 percent of the total bytes of information received (p. 9).

In the 1980s, Ithiel de Sola Pool and his colleagues investigated the growth of information (measured in words) supplied by the media in the U.S. and Japan (Neuman & Pool, 1986; Pool, 1983; Pool, Inose, Takasaki, & Hurwitz, 1984). They analyzed the number of words supplied and consumed as well as the average price per word. They reported that available information was shifting from print to electronic media, the price per word was falling dramatically, and although the rate of consumption was increasing (at 3.3 percent per year), it was falling ever further behind the amount of information supplied. These findings have implications for information overload, information diversity, and the economics necessary to sustain vibrant, creative industries in journalism and popular and high culture.

Neuman, Park, and Panek (2010) extended Pool's work to cover the period from 1960 through 2005. They found a tremendous increase in the ratio of supply to demand. In 1960, 98 minutes of media were available for every 1 minute of media consumed: Choices had to be made, but the number of choices was within reason. By 2005, more than 20,000 minutes of mediated content were available for every minute consumed. This, they point out, "is *not* a human-scale cognitive challenge; it is one in which humans will inevitably turn to the increasingly intelligent digital technologies that created the abundance in the first place for help in sorting it out—search engines, TiVo's recommendation systems, collaborative filters" (p. 11). Neuman and colleagues also found a change from *push* to *pull* technologies: Traditional, one-way broadcast and publishing media push content, with the audience accepting the decisions of newspaper editors and network executives. Today, technologies are evolving to pull in audience members, who have more choice and more control than ever before over what they watch and read, and when. Search engines (especially Google) and social networking sites (e.g., YouTube, Facebook) are emerging as major influences on public opinion and popular culture.

1.2. Where Does Information Come From?

Philosopher Karl Popper (1979) found it useful to use a metaphor of three "worlds" to describe how knowledge exists and develops:

- World 1: the physical world
- World 2: subjective reality (how we see or experience the world)
- World 3: objective knowledge (accumulated and scientific knowledge)

Science, Popper says, is a process that takes place in all three worlds: In World 1, events happen; in World 2, we try to make sense of them; and in

World 3, we try to explain the events while others react to these explanations and try to improve on them. Thus, we bring the three worlds together to create information (or awareness) through a never-ending process that produces knowledge. Along the way we create tools and technologies that help this process.

To take a less philosophical, more practical view, information reaches us from records of historical events, scientific knowledge, religious or cultural knowledge, art and literature, personal knowledge and records, documentation of governments or organizations, business, commerce, and many other sources.

Information may arrive prepackaged from a variety of sources. Publishers, government agencies, and other organizations produce formal products such as books, journals, and databases. Individuals package information in email, blogs, wikis, and other forms. Various institutions handle these packages. Libraries customarily deal with books, journals, video and audio recordings, microforms, databases, and even manuscripts, papyri, and clay tablets. Archives typically house governmental records, personal papers, and manuscripts. Databases (some commercially compiled and others available for free on the internet) also provide access to books, journals, webpages, blogs, videos, and other sources.

All of these various "packages" of information can be considered to be information systems (micro and macro) created to achieve some purpose. They may also be considered to be (micro and macro) communications systems, so that the information in them can be satisfactorily transferred: from the package to someone who wants the information or from one package to another package. However, all communications systems have potential problems. Information science seeks to analyze, design, and evaluate these systems in order to understand and improve how they function.

1.3. The Effects: Information Overload

The world is filled with information, and we acquire it in various ways:

1. We discover it through our physical, mental, and emotional senses.
2. We seek it by asking questions and searching for it.
3. We obtain it through feedback from other people and from various types of information systems.
4. We organize it (in our heads and in our files) and may make new information.

For centuries people have noted (or complained) that there is too much information in the world. In 1755 French encyclopedist Denis Diderot wrote that the increase in published material would eventually make it easier to rediscover facts from observing nature than to find information "hidden away in an immense multitude of bound volumes" (Diderot, 1975/1755, pp. 234–235). Alvin Toffler (1970) described the technological and structural changes in society in his book *Future Shock*, which helped to popularize the term *information overload*, meaning having so much information that it is difficult to set priorities or make decisions. Richard Saul Wurman (1989) observed that people respond with *information anxiety* to this inability to cope with the perceived flood of information.

Consider the ideas of information overload and information anxiety on a personal level. Thinking historically, compare the amount of information (and the systems for accessing that information) available to you today for succeeding in college or finding a job with that available to your parents and your grandparents. Are you, your family, and your country better off (financially, psychologically, or in other ways) because you can know almost instantly what is happening around the world (say in Baghdad, Moscow, or Mumbai)?

1.4. Evaluating Information

As we attempt to screen information and reduce the amount with which we must contend, we ask two basic questions about information: its value (what is it worth?) and its quality (is it any good?). Ultimately, the value estimation must be considered in light of the cost of the information, which brings us to the familiar question of the relation between costs and benefits.

Calculating a cost-benefit ratio is not easy because there are many aspects of cost and because the notion of benefit may be difficult to assess. Costs are typically of two types. Fixed costs, which are moderately easy to determine, include labor (salaries), equipment, supplies, and software. Variable costs are more difficult to determine. Examples include delays by others involved, unexpected breakdowns (for whatever reason), and mistakes or errors. On the other side of the cost-benefit ratio, the following questions can be used to determine the benefit of information or an information system:

1. Did it save time?
2. Did it enhance effectiveness?
3. Did it give us an advantage over the competition?
4. Did it save money in the short run and the long run?
5. Did it help avoid costs of some type?

Quality is the
Information scien
mine the quality o

- Accuracy
- Timelines
- Age and o
- Complete
- Source av
- Ease of ur
- Trustwortl

From the persp
the following criter

- Scope of c
- Authority:
- Objectivity
- Accuracy:
- Timeliness

Evaluating infor
information. Useful
lowing:

1. Identify an
2. Discover th
3. Determine
4. Establish tl
5. Verify factu

It is usually muc
subjective (opinion-

1.5. Managin

We are said to be livir
agree as to when it b

Quality is the second aspect we consider in evaluating information. Information scientists often consider the following factors in order to determine the quality of information:

- Accuracy
- Timeliness
- Age and obsolescence
- Completeness
- Source availability and ease of use
- Ease of understanding
- Trustworthiness of source

From the perspective of the legal research service Virtual Chase (2008a), the following criteria are valuable for assessing the quality of information:

- Scope of coverage: Is it inclusive or limited?
- Authority: Who said it?
- Objectivity: Is it limited or is there no bias?
- Accuracy: Has it been checked or verified?
- Timeliness: Is it out of date or up-to-date?

Evaluating information quality is especially important for web-based information. Useful steps identified by Virtual Chase (2008b) include the following:

1. Identify and check the source.
2. Discover the source's expertise.
3. Determine the level of objectivity.
4. Establish the date of publication.
5. Verify factual statements.

It is usually much easier to evaluate so-called factual information than subjective (opinion-based) information.

1.5 Managing Your Information

We are said to be living in an information society—even though historians disagree as to when it began and definitions of information vary. We can easily see some impacts of information on society (such as information overload, described in section 1.3), but others are hard to identify. We cannot say what the economic impacts of information are because we have trouble differentiating an information worker from a noninformation worker. This complication holds for many products today: Which are information intensive and which are not? Even if it is difficult to define, the notion of an information society is so common that we need at least a brief list of its major characteristics:

- Major changes occur in information technologies.
- Large portions of the economy deal with information.
- Many occupations now are information intensive.
- Information networks are a major feature of our lives.
- Information available for our use is extensive—and continually growing.

Who manages this information? Information professionals! And who are they? Their professional titles include database managers, webmasters, information systems staff, librarians, systems librarians, records managers, archivists, and many more.

What happens to information after it is created? A large portion is destroyed (by plan), such as online course materials that are removed after a specified time. Quite a bit of electronic information self-destructs, such as the data we generate while playing a video game. Some information is saved in archives (which may be personal, corporate, or governmental), and some is stored in libraries (and may eventually be destroyed or discarded). The web retains some information; for example, the Internet Archive's Wayback Machine (www.archive.org) shows earlier versions of websites. And some information is destroyed.

What can you do about the impact information has on you? Some options include managing it better; using new technologies to improve your control; creating better indexes, classification systems, and archival systems; and just getting rid of the information you no longer need.

References

Bohn, R. E., & Short, J. E. (2009). *How much information? 2009 report on American consumers.* San Diego, CA: Global Information Industry Center, University of California–San Diego. Retrieved November 11, 2010, from hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf.

Diderot, D. (1975–1995). Encyclopédie. In *Oeuvres complètes* (H. Dieckmann et al., Eds.) (vol. 7, pp. 174–262). Paris: Hermann. Original work published in 1755.

Neuman, W. R., Park
An empirical ass
MI: University
com/Flow_of_In

Neuman, W. R., & P
Rokeach & M. C
Hills, CA: Sage.

Pool, I. S. (1983). Tra

Pool, I. S., Inose, H.,
United States an

Popper, K. (1979). *Ob*
University Press.

Toffler, A. (1970). *Fut*

Virtual Chase. (2008
from virtualchas

Virtual Chase. (2008
2010, from virtua

Wurman, R. S. (1989)

science means, and the chapter concludes with an account of the field's intellectual and historical roots.

2.2. Defining Terms

Information

Defining *information* is an obvious first step in understanding information science and technology. Buckland (1991) observed that *things* can be informative. A tree stump contains in its rings information about its age, as well as information about the climate during the tree's lifetime. In similar ways, *anything* can be informative.

Some theorists hold that information is an objective phenomenon; others say that it depends on the receiver. Parker (1974) took the objective approach: "Information is the pattern of organization of matter and energy" (p. 10). Bateson (1972) took the subjective view: Information is "a difference that makes a difference" (for somebody or something or from a specific perspective; p. 453).

Note, first, the similarities between the two perspectives. Both the objective and the subjective views agree that any "pattern of organization of matter and energy" may inform somebody and thus be considered information. Information is thus a very broad term that is not limited to text or human products.

The basic difference should also be mentioned. If information is objective, then the representation of information is independent of context and purpose. If, on the other hand, information is understood as subjective, then its representation in information systems must consider who is to be informed and about what. These two perspectives, which Ellis (1992) labeled the *physical* and *cognitive paradigms*, have both provided useful bases for thought and development of information science.

Recently, researchers have added a third perspective: the *socio-cognitive approach*, which holds that individual cognitive or subjective understanding is conditioned by society and culture. Hjørland (1997), for example, holds that information "users should be seen as individuals in concrete situations in social organizations and domains of knowledge" (p. 111).

Some writers contrast information with the notions of *data* and *knowledge*.

Data

Data is the plural of *datum*, derived from Latin *dare* (to give); hence, data is "something given." Some style manuals insist that data be used only in the plural; it may, however, be used as a collective noun: a plural noun used in the singular to denote a set of items. Machlup (1983) wrote:

> Data are the things given to the analyst, investigator, or problem-solver; they may be numbers, words, sentences, records, assumptions—just anything given, no matter in what form and of what origin. This used to be well known to scholars in most fields: Some wanted the word *data* to refer to facts, especially to instrument-readings; others to assumptions. Scholars with a hypothetico-deductive bent wanted data to mean the given set of assumptions; those with an empirical bent wanted data to mean the records, or protocol statements, representing the findings of observation, qualitative or quantitative. … One can probably find quotations supporting all possible combinations of the three terms [*data*, *information*, and *knowledge*] or of the concepts they are supposed to denote. Each is said to be a specific type of each of the others, or an input for producing each of the others, or an output of processing each of the others. Now, data from the point of view of the programmers, operators, and users of the computer, need not be data in any other sense. (pp. 646–647)

Spang-Hanssen (2001) discussed data as well in a 1970 speech:

> Information about some physical property of a material is actually incomplete without information about the precision of the data and about the conditions under which these data were obtained. Moreover, various investigations of a property have often led to different results that cannot be compared and evaluated apart from information about their background. An empirical fact has always a history and a perhaps not too certain future. This history and future can be known only through information from particular documents, i.e. by document retrieval. The so-called fact retrieval centers seem to be just information centers that keep their information sources—i.e. their documents—exclusively to themselves.

We may conclude that what is considered data is relative: What some consider the given (or input), others may consider the output. From the perspective of information science, it is important to represent and communicate not just data but also its background, its reception, and the theoretical assumptions connected with data, which makes the concepts of *knowledge* and *document* important.

Knowledge

The classical definition goes back to Plato: Knowledge is verified true belief. This definition is problematic, however, because knowledge is always open to modification and revision, so that very little (or nothing) can be considered knowledge in Plato's sense. This is why pragmatic and materialist theories consider the concept of knowledge in relation to human practice: Knowledge expands the actors' possibilities to act and to adjust to the world in which they live. Pragmatism and materialism consider human practice the final criterion of knowledge and see experimentation as an integrated component.

The *Oxford English Dictionary* definitions of knowledge include 1) "skill or expertise acquired in a particular subject ... through learning;" 2) "that which is known;" and 3) "being aware or cognizant of a fact, state of affairs, etc." (OED Online, "knowledge").

The Data-Information-Knowledge-Wisdom Hierarchy

Ackoff (1989) saw the *information pyramid* (Figure 2.1) as a progression:

1. Data are facts that result from observations.
2. Information is collections of facts provided with context.
3. Knowledge is generated when people supply meaning to information.
4. Wisdom results from shared insights and knowledge.

For example:

1. Contributors to the World Wide Web post results from their empirical research in the sciences or new insights into historical events or literary research. These contributors create links from one page to another. Each link could be considered a piece of data (a datum).
2. By tracing the links, we can create a structure or map of the web. This organized collection of links is information.
3. Analyzing the map of the web allows us to see areas of dense linking and identify sites that receive links from many others (hubs). This knowledge of web structure results from our understanding of the information.
4. The occurrence of many links to a website is often believed to indicate that the site has value and potential utility for other web users. Search engines such as Google take advantage of this "wisdom."

ge is verified true belief.
wledge is always open to
ing) can be considered
and materialist theories
an practice: Knowledge
the world in which they
actice the final criterion
ed component.
ledge include 1) "skill or
earning;" 2) "that which
ct, state of affairs, etc."

m Hierarchy

1) as a progression:

with context.

meaning to

owledge.

ults from their
sights into historical
rs create links from
sidered a piece of data

e or map of the web.
ion.

e areas of dense
om many others
lts from our

often believed to
utility for other web
lvantage of this

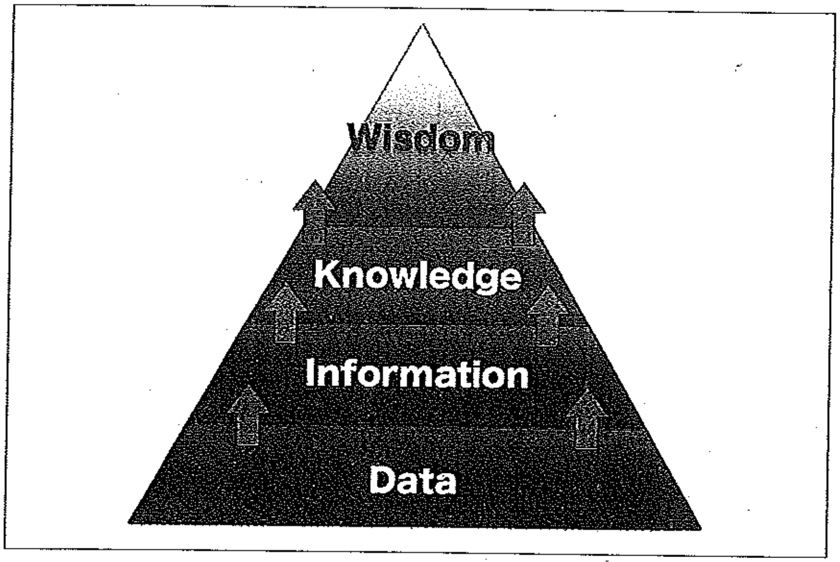


Figure 2.1 Data-information-knowledge-wisdom pyramid

Braganza (2004) suggested a top-down perspective rather than the traditional bottom-up approach. Rather than assuming that data is the basic unit of information and knowledge, information professionals, in order to provide more useful insights into information work, should consider beginning with a focus on the creation and communication of knowledge.

The pyramid model does a reasonable job of reflecting the evolution of thinking about the concept of information. Early research focused on the base of the pyramid: how to send, receive, and manipulate bits of data. The Shannon-Weaver model of communication (Figure 2.2) from 1949 shows this focus (and is discussed in more detail in Chapter 14).

In the 1940s, Claude Shannon was working for the telephone company at Bell Labs. While investigating how to transmit a message both efficiently and effectively, he realized that noise, from any source, could keep the destination (the person on the other end of the telephone line) from receiving the message that the information source had sent. Shannon's analysis also demonstrated that there is a theoretical limit to bandwidth.

In 1949, Shannon and Warren Weaver wrote *The Mathematical Theory of Communication*, which demonstrates how redundancy helps to compensate for noise in the transmission of a message. If you are directing a colleague to a site on the web, you might give the internet protocol (IP) address, such as 209.85.129.99. However, if you mistype just one numeral, your friend might be directed to the wrong location. By using a URL in natural language, your friend can compensate for errors (noise) that might creep into the message:

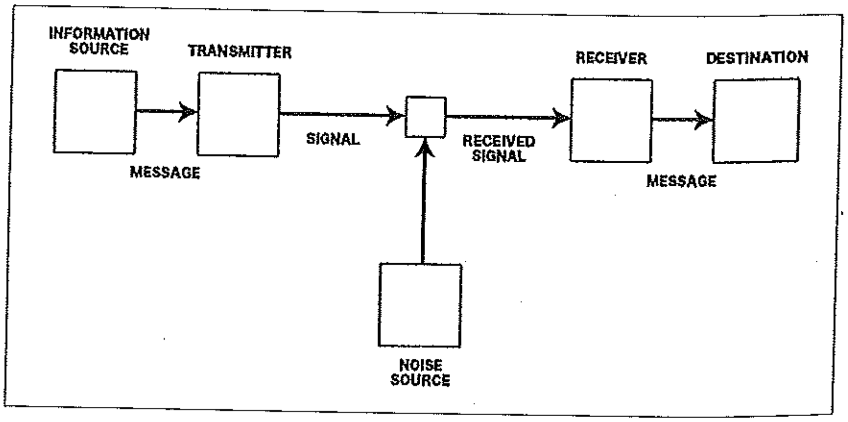


Figure 2.2 Shannon-Weaver model of communication

www.google.com is the redundant, human-understandable version of an IP address.

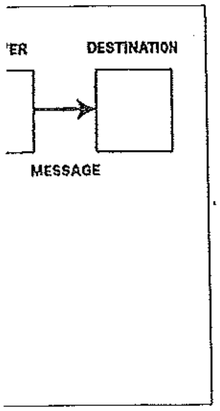
Shannon and Weaver (1949/1964) identified three aspects of information:

1. Technical aspects, concerned with problems of transmission
2. Semantic aspects, related to the meaning and truth of a message
3. Influential aspects, concerned with how a message affects human behavior

The definition of information used in their mathematical theory considers only the first level, the technical concerns in transmission; this is the base of the information pyramid. The Shannon-Weaver model has been criticized for its *conduit metaphor*, emphasizing the channel, rather than the source and destination of the message.

Interest in the semantic aspects of information had gained ground by the 1980s. Researchers noted that different people had different understandings of the same item of information. This led to research on the cognitive aspects of information (the mental processes of knowing), including how people assess information (Machlup & Mansfield, 1983). Requiring consideration of the human perspective means that information can no longer be understood objectively; what is informative will depend on the person assessing the meaning and truth of a message, as Shannon and Weaver would say. Recent research has also considered social aspects of how information is understood. This approach notes that how each of us understands an item of potential information is influenced by our social environment: societal conventions (such as language), history, and interactions with other people. This view has been labeled the *socio-cognitive approach to information* (Hjørland, 2002).

ıgy



ıble version of an IP

ıects of information:

transmission

ruth of a message

ssage affects human

ıtical theory consid-
ssion; this is the base
el has been criticized
ther than the source

gained ground by the
rent understandings
the cognitive aspects
cluding how people
ring consideration of
onger be understood
person assessing the
er would say. Recent
nation is understood.
an item of potential
societal conventions
people. This view has
n (Hjørland, 2002).

2.3. Disseminating Information

Documents

Before information science was termed *information science*, it was called *documentation*, and documents were considered the basic objects of study for the field. Buckland (1991) has described the history of the concept of *document* in information science. Early in the 20th century, researchers felt a need for a generic term for the object of their work: not only texts, but also natural objects, artifacts, models, objects reflecting human activities, objects of art, and human ideas. The term document (or *documenting unit*) was used with a special meaning in order to include informative physical objects. Buckland noted that the word document comes from Latin *docere*, meaning to teach or inform, and the suffix -*ment*, meaning a tool. Originally, then, the word meant a tool for teaching or informing, whether through lecture, experience, or text. Only later was it narrowed to mean objects carrying texts.

In information science today, the concept of document is understood as "any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental" (Briet, 1951, p. 7, as cited in Buckland, 1991, p. 355).

Information and Communication Technologies

For the past half century we have used the term *information technology* to note the use of computer hardware and software for handling information. Kline (2004) traces the origin of the term to the business world, where "management information systems" were developed in the 1960s. The term information and communication technologies (ICTs) has been adopted more recently, acknowledging the increasing importance of telephones, cable, and satellite transmission in effective use of information technologies. Figure 2.3 uses data from the U.S. Census to show the rates of adoption for several ICTs.

Alan Kay, who worked at Atari, Xerox, Apple, and Disney, has defined technology as "anything that was invented after you were born." His assertion expresses the common feeling that new technologies are being introduced and adopted with increasing speed. Figure 2.3 suggests that U.S. residents born in the 21st century will view broadband internet as a natural part of life, but it will always be a "technology" for many of their parents. As we design, use, and evaluate systems that rely on ICTs, we should be aware of new developments and also mindful of the "long tail" of technologies that some users will approach as novel or challenging even as others accept them as an inseparable part of life.

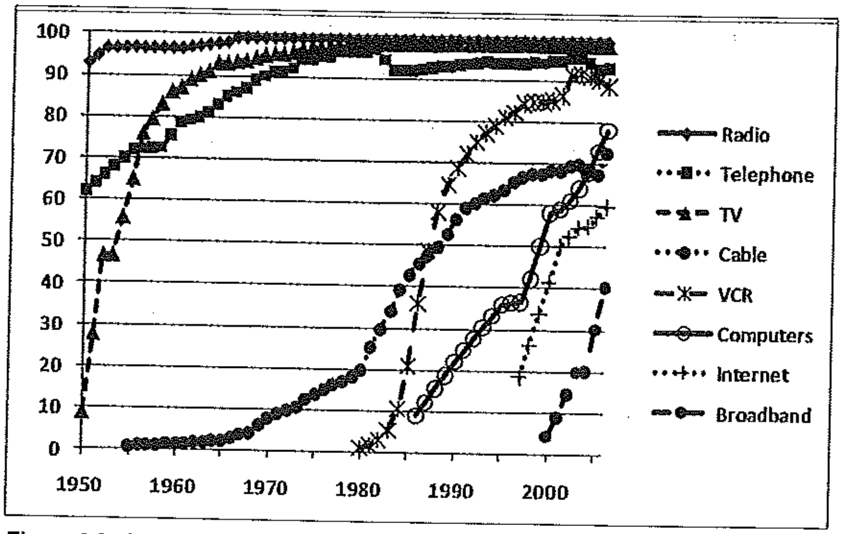


Figure 2.3 Percentage of U.S. households using information technologies, 1950–2006

2.4. Information Science

Information science emerged as the name for this field in about 1960. The Institute of Information Scientists was established in 1958; the American Documentation Institute changed its name to the American Society for Information Science in 1968, and in 2000 changed its name again to the American Society for Information Science and Technology.

Borko's (1968) definition of information science provides a list of tasks the field should address: "the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information" (p. 3). *The Online Dictionary for Library and Information Science* (Reitz, 2007) uses similar terms: "The systematic study and analysis of the sources, development, collection, organization, dissemination, evaluation, use, and management of information in all its forms, including the channels (formal and informal) and technology used in its communication."

Sometimes the plural, information sciences, is used. Machlup and Mansfield (1983), for example, suggested that one should speak about the information sciences as one speaks of the social sciences.

The term *informatics*, which was proposed independently by Walter F. Bauer and Phillipe Dreyfus in 1962 (Bauer, 1996), has a similar meaning. Redmond-Neal and Hlava (2005) say it "represents the conjunction of information science and information technology" (p. 63). Reitz (2007) continues, "It is the formal study of information, including its structure, properties, uses, and functions in society; the people who use it; and in particular the