Wolfgang G. Stock, Mechtild Stock
Handbook of Information Science

Wolfgang G. Stock,
Mechtild Stock

Handbook of Information Science

DE GRUYTER
SAUR

Translated from the German, in cooperation with the authors, by Paul Becker.

ISBN 978-3-11-023499-2
e-ISBN 978-3-11-023500-5

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Typesetting: Michael Peschke, Berlin
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid free paper
Printed in Germany

www.degruyter.com

Preface

Dealing with information, knowledge and digital documents is one of the most important skills for people in the 21st century. In the knowledge societies of the 21st century, the use of information and communication technology (ICT), particularly the Internet, and the adoption of information services are essential for everyone—in the workplace, at school and in everyday life. ICT will be ubiquitous. Knowledge is available everywhere and at any time. People search for knowledge, and they produce and share it. The writing and reading of e-mails as well as text messages is taken for granted. We want to be well informed. We browse the World Wide Web, consult search engines and take advantage of library services. We inform our friends and colleagues about (some of) our insights via social networking, (micro-)blogging and sharing services.

Information, knowledge, documents and their users are research objects of Information Science, which is the science as well as technology dealing with the representation and retrieval of information—including the environment of said information (society, a certain enterprise, or an individual user, amongst others).

Information Science meets two main challenges. On the one hand, it analyzes the creation and representation of knowledge in information systems. On the other hand, it studies the retrieval, i.e. the searching and finding, of knowledge in such systems, e.g. search engines or library catalogs. Without information science research, there would be no elaborate search engines on the World Wide Web, no digital catalogs, no digital products offered on information markets, and no intelligent solutions in corporate knowledge management. And without information science, there would be no education in information literacy, which is the basic competence that guarantees an individual's success in the knowledge society.

This handbook focuses on the fundamental disciplines of information science, namely information retrieval, knowledge representation and informetrics. Whereas information retrieval is oriented on the search for and the retrieval of information, knowledge representation starts in the preliminary stages, during the indexing and summarization of documents. We will also discuss results from informetrics, in so far as they are relevant for information retrieval and knowledge representation. The subjects of informetrics are the measurement of information, the evaluation of information systems, as well as the users and usage of information services.

We do not discuss research into information markets and the knowledge society, as those topics are included in another current monograph (Linde & Stock, 2011).[1]

Information science is both basic research and applied research. Our scientific discipline nearly always orients itself on the applicability and technological feasibility of its results. As a matter of principle, it incorporates both the users and usage of its tools, systems and services.

---

1 Linde, F., & Stock, W.G. (2011). Information Markets. A Strategic Guideline for the I-Commerce. Berlin, New York, NY: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)

The individual chapters of our handbook all have the same structure. After the research areas and important research results have been discussed, there follows a conclusion, summarizing the most important insights gleaned, as well as a list of relevant publications that have been cited in the text.

The handbook is written from a systematic point of view. Additionally, however, we also let scholars have their say with regard to their research results. Hence, this book features many quotations from some of the most important works of information science. Quotations that were originally in the German language have been translated into English. We took care to not only portray the current state of information science, but also the historical developments that have led us there.

We endeavor to provide the reader with a well-structured, clear, and up-to-date handbook. Our goal is to give the demanding reader a compact overview of information retrieval, knowledge representation, and informetrics. The handbook addresses readers from all professions and scientific disciplines, but particularly scholars, practitioners, and students of

- Information Science,
- Library Science,
- Computer Science,
- Information Systems Research, and
- Economics and Business Administration (particularly Information Management and Knowledge Management).

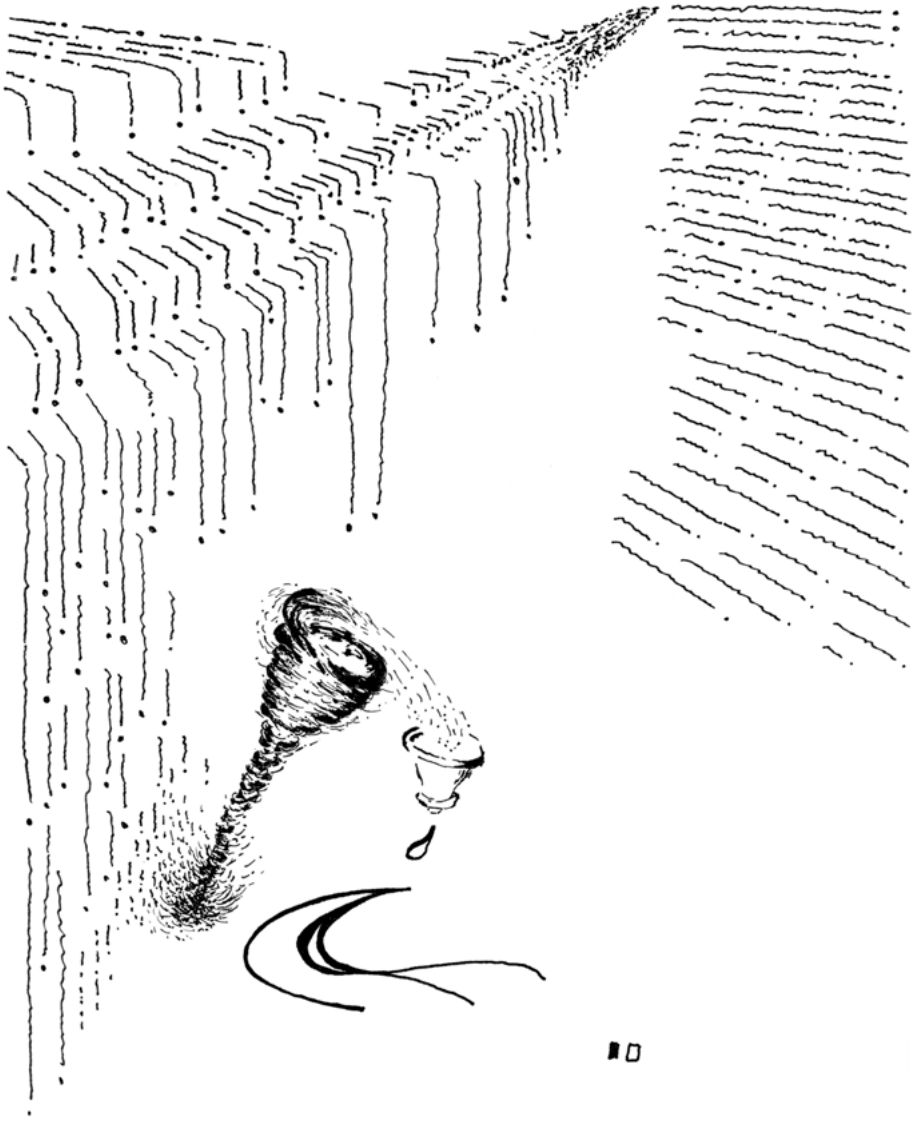
Acknowledgments

Special thanks go to our translator *Paul Becker* for his fantastic co-operation in translating our texts from German into English. Many thanks, also, to the Information Science team at the *Heinrich-Heine-University Düsseldorf*, for their many thought-provoking impulses! And finally, we would like to thank the staff at *Walter de Gruyter* publishing house for their perfect collaboration in producing this handbook.

Mechtild Stock
Wolfgang G. Stock
May, 2013

Contents

Part A
Introduction to Information Science



A.1 What is Information Science?

Defining Information Science

What is information science? How can we define this budding scientific discipline? In order to describe and delineate information science, we will examine several approaches and pursue the following lines of questioning:

- What are the fundamental determinants and tasks of information science?
- What roles are played by knowledge, information and documents?
- How has our science developed? Do the branches of information science look back on different histories?
- To which other scientific disciplines is information science closely linked?

There is no generally accepted definition of information science (Bawden & Robinson, 2012; Belkin, 1978; Yan, 2011). This is due, first of all, to the fact that this science, at around fifty years of age, is still relatively young when compared to the established disciplines (such as mathematics or physics). Secondly, information science is strongly interrelated with other disciplines, e.g. information technology (IT) and economics, each of which place great emphasis on their own definitions. Hence, it is not surprising that information science today is used for the divergent purposes of foundational research on the one hand, and applied science on the other.

Let us begin with a working definition of "information science"!

> *Information Science studies the representation, storage and supply as well as the search for and retrieval of relevant (predominantly digital) documents and knowledge (including the environment of information).*

Information science is a fundamental part of the development of the knowledge society (Webster, 1995) or the network society (Castells, 1996). In everyday life (Catts & Lau, 2008) and in professional life (Bruce, 1999), information science finds its expression in the skills of information literacy. "We already live in an era with information being the sign of our time; undoubtedly, a science about this is capable of attracting many people" (Yan, 2011, 510).

Information science is strongly related to both computer science (since digitally available knowledge is principally dependent upon information-processing machines) and to the economic sciences (knowledge is an economic good, which is traded or—sometimes—distributed for free on markets), as well as to other neighboring disciplines (such as library science, linguistics, pedagogy and science of science).

We would like to inspect the determinants of our definition of *information science* a little more closely.

- *Representation*: Knowledge contained in documents as well as the documents themselves (let us say, scientific articles, books, patents or company publications, but also websites or microblog posts) are both condensed via short content-

descriptive texts and labeled with important words or concepts with the purpose of information filtering.
- *Storage and supply:* Documents are to be processed in such a way that they are ideally structured, easily retrievable and readable and stored in digital locations, where they can be managed.
- *Search*: Information science observes users as they satisfy their information needs, it observes their query formulations in search tools and it observes the way they use the retrieved information.
- *Retrieval*: The focal points of information science are systems for researching knowledge; prominent examples include Internet search engines, but also library catalogs.
- *Relevance*: The objective is not to find "any old" information, but only the kind of knowledge that helps the user to satisfy his information needs.
- *Predominantly digital*: Since the advent of the Internet and of the commercial information industry, large areas of human knowledge are digitally available. Even though digital information represents a core theme of information science, there is also room left for non-digital information collections (e.g. in archives and libraries).
- *Documents:* Documents are texts and non-textual objects (e.g., images, music and videos, but also scientific facts, economic objects, objects in museums and galleries, real-time facts and people). And documents are both physical (e.g., printed books) and digital (e.g., blog posts).
- *Knowledge*: In information science, knowledge is regarded as something static, which is fixed in a document and stored on a memory. This storage is either digital (such as the World Wide Web), material (as on a library shelf) or psychical (like the brain of a company employee). Information, on the other hand, always contains a dynamic element; one informs (active) or is informed (passive). The production and the use of knowledge are deeply embedded in social and cultural processes; so information science has a strong cultural context (Buckland, 2012).

The possible applications of information science research results are manifold and can be encountered in a lot of different places on the information markets (Linde & Stock, 2011). By way of example, we will emphasize six types of application:
- Search engines on the Internet (a prominent example: Google),
- Web 2.0 services (such as YouTube for videos, Flickr for images, Last.fm for music or Delicious for bookmarks of websites),
- Digital library catalogs (e.g. the global project WorldCat, or catalogs relating to precisely one library, such as the catalog of the Library of Congress),
- Digital libraries (catalogs *and* their associated digital objects, such as the Perseus Digital Library for antique documents, or the ACM Digital Library for computer science literature),

- Digital information services (particularly specialist databases with a wide spectrum reaching from business and press information, via legal information, all the way to information from science, technology and medicine),
- Information services in corporate knowledge management (the counterparts of the above-mentioned five cases of Internet application, transposed to company Intranets).

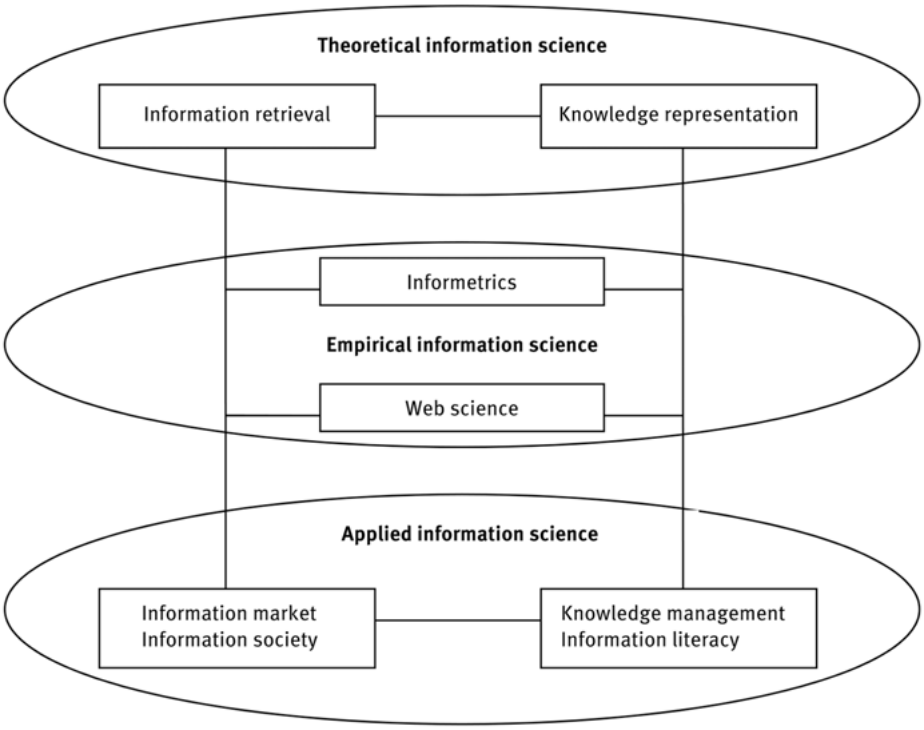


Figure A.1.1: Information Science and its Sub-Disciplines.

Information Science and Its Sub-Disciplines

Information science comprises a spectrum of five sub-disciplines:
- Information Retrieval,
- Knowledge Representation,
- Knowledge Management and Information Literacy,
- Research into the Information Society and Information Markets,
- Informetrics, including Web Science.

A central role is occupied by *Information Retrieval* (Baeza-Yates & Ribeiro-Neto, 2011; Manning, Raghavan, & Schütze, 2008; Stock, 2007), which is the science and engi-

neering of search engines (Croft, Metzler, & Strohman, 2010). Information retrieval investigates not only the technical systems, but also the information needs of the people (Cole, 2012) who use those systems (Ingwersen & Järvelin, 2005). The fundamental question of information retrieval is: how can one create technologically ideal retrieval systems in a user-optimized way?

Knowledge Representation addresses surrogates of knowledge and of documents in information systems. The theme is data describing documents and knowledge—so-called "metadata"—, such as a catalog card in a library (Taylor, 1999). Added to this are methods and tools for use during indexing and summarization (Chu, 2010; Lancaster, 2003; Stock & Stock, 2008). Essential tools are Knowledge Organization Systems (KOSs), which provide concepts and designations for indexing and retrieval. Here, the fundamental question is: how can knowledge in digital systems be represented, organized and condensed in such a way that it can be retrieved as easily as possible?

One subfield on the application side is *Knowledge Management* (Nonaka & Takeuchi, 1995), which focuses on the distribution and sharing of in-company knowledge as well as the integration of external knowledge into the corporate information systems. Another application of information science is research on *Information Literacy* (Eisenberg & Berkowitz, 1990), which means the use of (some) results of information science in everyday life, the workplace (Bruce, 1999), school (Chu, 2009) and university education (Johnston & Webber, 2003).

Research into the *Information Markets* (Linde & Stock, 2011) and the *Information Society* (Castells, 1996; Cronin, 2008; Webster, 1995) is particularly important with regard to the knowledge society. These research endeavors of information science have an extensive subject area, comprising everything from a country's information infrastructure up to the network economics, by way of the industry of information service providers.

Information science proceeds empirically—where possible—and analyzes its objects via quantitative methods. *Informetric Research* (Weber & Stock, 2006; Wolfram, 2003) comprises webometrics (research into the World Wide Web; Thelwall, 2009), scientometrics (research into the information processes in science and medicine; van Raan, 1997; Haustein, 2012), "altmetrics" as an "alternative" metrics (informetrics with regard to services in Web 2.0; Priem & Hemminger, 2010) as well as patent informetrics (Schmitz, 2010). Empirical research is also performed during the evaluation of information systems (Voorhees, 2002) as well as during user and information needs analyses (Wilson, 2000). Additionally, informetrics attempts to detect the regularities or even laws of information processes (Egghe, 2005). One cross-section of informetrics, called *Web Science*, concentrates on research into the World Wide Web, its users and their information needs (Berners-Lee et al., 2006).

Information science has strong connections to a broad range of other scientific fields. It seems to be an interdisciplinary science (Buckland, 2012, 5). It combines, among others, methods and results from computer science (e.g., algorithms of

information retrieval and knowledge representation), the social sciences (e.g., user research), the humanities (e.g., linguistics and philosophy, analyzing concepts and words), economics (e.g., endeavors on the information markets), business administration (e.g., corporate knowledge management), education (e.g., applying information services in e-learning), and engineering (e.g., the construction of search engines for the World Wide Web) (Wilson, 1996). However, information science is by no means a "mixture" of other sciences, but a science on its own right. The central concern of information science—according to Buckland (2012, 6)—is "cultural engagement", or, more precisely, "enabling people to become better informed (learning, become more knowledgeable)" (Buckland, 2012, 5). The application of information science in the normal course of life is to secure information literacy for all people. Information literate people are able to inform other people adequately (to represent, to store and to supply documents and knowledge) and they are always able to be well informed (to search for and to retrieve relevant documents and knowledge).

Information Science as Basic and Applied Research

We regard information science as both basic and applied research. In economic terms, it is an "importer" of ideas from other disciplines and an "exporter" of original results to other scientific branches (Cronin & Pearson, 1990). Information science is a basic discipline (and so an exporter of ideas) for some subfields of IT, of library science, of science policy and of the economic sciences. For instance, information science develops an innovative ranking algorithm for search engines, using user research and conceptional analyses. Computer science then takes up the suggestions and develops a workable system from them. On the other hand, information science also builds upon the results of computer science and other disciplines (e.g. linguistics) and imports their ideas. Endeavors toward the lemmatization of terms within texts, for instance, are doomed without an exact knowledge of linguistic results. The import/export-ratio of information science varies according to the partner discipline (Larivière, Sugimoto, & Cronin, 2012, 1011):

> Although LIS's (LIS: Library and Information Science, A/N) import dependency has been steadily decreasing since the mid-1990s—from 3.5 to about 1.3 in 2010 (inset)—it still has a negative balance of trade with most fields. The fields with which LIS has a positive balance of trade are from the natural, mostly medical sciences. Several of the fields with which LIS has a negative balance of trade are from the social sciences and the humanities.

Information science nearly always—even in questions of basic research—looks toward technological feasibility, incorporating user or usage as a matter of principle. It locates its object of research by investigating existing systems (such as search engines, library catalogs, Web 2.0 services and other information service providers' systems) or creating experimental systems. Information science either analyzes and

evaluates such systems or does scientific groundwork for them. Designers of information services have to consider the understanding of the systems' users, their background knowledge, tradition and language. Thus information science is by no means only a technology, but also analyzes cognitive processes of the users and other stakeholders (Ingwersen & Järvelin, 2005).

Methods of knowledge representation are regarded independently of their historical provenance, converging into a single repertoire of methods: knowledge representation typically spans computer science aspects (ontologies; Gómez-Pérez, Fernández-López, & Corcho, 2004; Gruber, 1993; Staab & Studer, Eds., 2009), originally library-oriented or documentary endeavors (nomenclatures, classifications, thesauri; Lancaster, 2003) as well as ad hoc developments in the collaborative Web (folksonomies; Peters, 2009). Knowledge representation is an important building stone on the way to the "Semantic Web" (Berners-Lee, Hendler, & Lassila, 2001) or the "Social Semantic Web" (Weller, 2010), respectively.

Information Science as the Science of Information Content

The fixed point of information science is information itself, i.e. the structured information content which expresses knowledge. According to Buckland (1991), "information" has three aspects of meaning, all of which are objects of information science:

- information as a process (one informs / is informed),
- information as knowledge (information transports knowledge),
- information as a thing (information is fixed in "informative things", i.e. in documents).

Documents are both texts (Belkin & Robertson, 1976) and non-textual objects (Buckland, 1997) such as images, music and videos, but they are also objects in science and technology (e.g., chemical substances and compounds), economic objects (such as companies or markets), objects in museums and galleries, real-time facts (e.g., weather data) as well as people (Stock, Peters, & Weller, 2010). Besides documents which are created by machines (e.g., flight-tracking systems), the majority of documents include "human recorded information" (Bawden & Robinson, 2012, 4). Documents are very important for information science (Davis & Shaw, Eds., 2011, 15):

> Before information science was termed *information science*, it was called *documentation*, and documents were considered the basic objects of study for the field.

Of less interest to information science are technological information processing (which is also an object of IT) and the organization of information activities toward the sale of content (these are subject to the economic sciences as well). Thus Rauch (1988, 26) defines our discipline via the concept of knowledge:

> For *information science* ... information is *knowledge*. More precisely: knowledge that is needed in order to deal with problematic situations. Knowledge is thus *possible* information, in a way. Information is knowledge that becomes effective and relevant for actions.

Kuhlen (2004, 5) even discusses the term "knowledge science", which in light of the sub-disciplines of "knowledge representation" and "knowledge management" would not be completely beside the point. Kuhlen, too, adheres to "information science", but emphasizes its close connection to knowledge (Kuhlen, 2004, 6):

> Information does not exist in itself. Information refers to knowledge. Information is generally understood to be a surrogate, a representation or manifestation of knowledge.

The specific content plays a subordinate role in information science; its objects are the structure and function of information and information processing.

> (For information science) it is of no importance whether the object of discussion is a new insect, for example, or an advanced method of metal working,

Mikhailov, Cernyi and Gilyarevsky (1979, 45) write. One of the "traditional" definitions of information science comes from Borko (1968, 3):

> Information science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information. This includes the investigation of information representation in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. ... It has a pure science component, which inquires into the subject without to its application, and an applied science component, which develops services and products.

"That definition has been quite stable and unvarying over at least the last 30 years," is Bates's (1999, 1044) comment on this concept definition. Borko's "pure science" can be further subdivided—as we have already seen in Figure A.1.1—into a theoretical information science, which works out the fundaments of information retrieval as well as of knowledge representation, and into an empirical information science that systematically studies information systems and users (who deal with information systems). Applied information science addresses the use of information in practice, i.e. on the information markets, within a company or administration in knowledge management and in everyday life as information literacy. One application of information science in information practice aims toward keeping all members of a company, college, city, society or humanity in general as ideally informed as possible: everyone, in the right place and at the right time, must be able to receive the right amount of

relevant knowledge, rendered comprehensibly, clearly structured and condensed to the fundamental quantity.

Whereas information science does have some broad theoretical research areas, as a whole the discipline is rather oriented towards application. Even though certain phenomena may not yet be entirely resolved on a theoretical basis, an information scientist will still go ahead and create workable systems. Search engines on the Internet provide an illustrative example. The fundamental elements of processing queries and documents are nowhere exhaustively discussed and resolved in the theory of information retrieval; even the concept of relevance, which ought to be decisive for the relevance ranking being used, is anything but clear. Yet, for all that, the search engines built on this basis operate very successfully. Looking back on the advent of information science, Henrichs (1997, 945) describes the initial sparks for the discipline, which almost exclusively stem from information practice:

> The practice of modern specialist information ... undoubtedly has a significant chronological advantage over its theory. Hence, in the beginning—and there can be no doubting this fact in this country (Germany, A/N), at least—there was practice.

As early as 1931, Ranganathan formulated five "laws" of library science. For Ranganathan, these are rules describing the processing of books in libraries. For today's information science, the information content—whether fixed in books, websites or anywhere else—is the crucial aspect:

1st Law: Information is for use.
2nd Law: Every user his or her information.
3rd Law: Every information its user.
4th Law: Save the time of the user!
5th Law: Information practice and information science are growing organisms.

The first four rules refer to the aspect of practice, which always comes to bear on information science. The fifth "Law" states that the development of information practice and information science is not over yet by a long shot, and is constantly evolving.

A Short History of Information Science and Its Sub-Disciplines

The history of searching and finding knowledge begins during the period in which human beings first began to systematically solve problems. "As we strive to better understand the world that surrounds us, and to control it, we have a voracious appetite for information," Norton (2000, 4) points out. The history of information science as a scientific discipline, however, goes back to the 1950s at its earliest (Bawden & Robinson, 2012, 8). The term "information science" was coined by Jason Farradane in the 1950s (Shapiro, 1995). Information science finally succeeded in establishing itself,

particularly in the USA, the United Kingdom and the former socialist countries, at the end of the 1960s (Schrader, 1984).

> Information science has deep historical roots accented with significant controversy and conflicting views. The concepts of this science may be at the heart of many disciplines, but the emergence of a specific discipline of information science has been limited to the twentieth century (Norton 2000, 3).

In its "prehistory" and history, we can separate five strands that each deal with partial aspects of information science.

Information Retrieval

The retrieval of information has been discussed ever since there have been libraries. Systematically organized libraries help their users to purposefully retrieve the information they desire. Let the reader imagine being in front of a well-arranged bookshelf, and discovering a perfectly relevant work at a certain place (either via the catalog or by browsing). Apart from this direct hit, there will be further relevant books to its left and right, the significance of which to the topic of interest will diminish the further the user moves away from the center. The terms "catalog" (or, today, "search engine"), "bookshelf" (or "hit list"), "browsing", "good arrangement" and "relevance" give us the basic concepts of information retrieval. The basic technology of information retrieval did not change much until the days of World War II. The advent of computers, however, changed the situation dramatically. Entirely new ways of information retrieval opened up. Experiments were made using computers as knowledge stores and elaborate retrieval languages were developed (Mooers, 1952; Luhn, 1961; Salton & McGill, 1983). Experimental retrieval systems were created in the 1960s, while commercial systems with specialist information have existed since the 1970s. Retrieval research reached its apex with the Internet search engines of the 1990s.

Knowledge Representation

What is a "good arrangement" of information? How can I represent knowledge ideally, i.e. condense it and make it retrievable via knowledge organization systems? This strand of development, too, stems from the world of libraries; it is closely related to information retrieval. Here, organization systems are at the forefront of the debate. Early evidence of such a system includes the systematics of the old library of Alexandria. Research into knowledge representation was boosted by the Decimal Classification developed by Mevil Dewey in the last quarter of the 19th century (Dewey, 1876), which was then developed further by Paul Otlet and Henri La Fontaine, finally leading to the plan of a collection and representation of world knowledge (Otlet, 1934). Over the course of the 20th century, various universal and specialized classification systems

were developed. With the triumph of computers in information practice, information scientists created new methods of knowledge representation (such as the thesaurus) as well as technologies for automatically indexing documents.

Knowledge Management and Information Literacy

In economics and business administration, information has long been discussed in the context of entrepreneurial decisions. Information always proves imperfect, and so becomes a motor for innovative competition. With conceptions for learning organizations, and, later, knowledge management (Nonaka & Takeuchi, 1995), the subject area of industrial economics and of business administration on the topic of information has broadened significantly from around 1980 onward. The objective is to share and safeguard in-company knowledge (Probst, Raub, & Romhardt, 2000) and to integrate external knowledge into an organization (Stock, 2000).

Information literacy has two main threads, namely retrieval skills and skills of creation and representation of knowledge. Retrieval literacy has its roots in library instruction and was introduced by librarians in the 1970s and 1980s (Eisenberg & Berkowitz, 1990). Since 1989, there exists a standard definition of information literacy, which was formulated by the American Library Association (Presidential Committee on Information Literacy, 1989). With the advent of Web 2.0 services, a second information literacy thread came into life: the skills of creating documents (e.g., videos for publishing on YouTube) and of representing them thematically (e.g., with tags).

Information Markets and Information Society

With the "knowledge industry" (Machlup, 1962), the "information economy" (Porat, 1977), the "postindustrial society" (Bell, 1973), the "information society" (Webster, 1995) or the "network society" (Castells, 1996), the focus on knowledge has for around 50 years been shaped by a sociological and economic perspective. In economics, information asymmetries (Akerlof, 1970) are discovered and measures to counteract the unsatisfactory aspects of markets with information asymmetries are developed. Thus, the sellers of digital information know a lot more about the quality of their product than their customers do, for example.

Originally, the information market was conceived as a very broad approach that led to the construction of a fourth economic factor (besides labor, capital and land) as well as to a fourth economic sector (besides agriculture, industry and services). Subsequently, the research area of the information markets was reduced to the sphere of digital information, which is transmitted via networks—the Internet in particular, but also mobile telephone networks (Linde & Stock, 2011). Research into the information, knowledge or network society thus discusses a very large area. Its subjects reach from "informational cities" as prototypes of cities in the knowledge society (Stock, 2011)

up to the "digital divide", which is the result of social inequalities in the knowledge society (van Dijk & Hacker, 2003).

Informetrics and Web Science

Starting in the second quarter of the 20th century, researchers discovered that the distribution of information follows certain regularities. Studies of ranking distributions—e.g. of authors in a discipline according to their numbers of publications—led to laws, of which those formulated by Samuel C. Bradford, Alfred J. Lotka or George K. Zipf have become classics (Egghe, 2005). Chronological distributions of information are described via the concept of "half-life". Since the 1970s, informetrics has established itself as information science's measurement method. Some much-noticed fields of empirical information science include scientometrics and patentometrics, since informetrics allows us to measure scientific and technological information. This leads to the possibility of making statements about the "quality" of research results, and furthermore about the performance and influence of their developers and discoverers.

From around 2005 onward, an interdisciplinary research field into the World Wide Web, called Web Science, has emerged (Hendler et al., 2008). Information science takes part in this endeavor via webometrics, altmetrics, user research as well as information needs research.

Information Science and Its Neighbors

In his classification of the position of information science, Saracevic (1999, 1052) emphasizes its relations to other disciplines. Among these, he assigns particular importance to the role of information technology and the information society in interdisciplinary work.

> First, information science is interdisciplinary in nature; however, the relations with various disciplines are changing. The interdisciplinary evolution is far from over.
> Second, information science is inexorably connected to information technology. A technological imperative is compelling and constraining the evolution of information science (...).
> Third, information science is, with many other fields, an active participant in the evolution of the information society. Information science has a strong social and human dimension, above and beyond technology.

Information science actively communicates with other scientific disciplines. It is closely related to the following:

- Computer Science,
- Economics,
- Library Science,

- (Computational) Linguistics,
- Pedagogy,
- Science of Science.

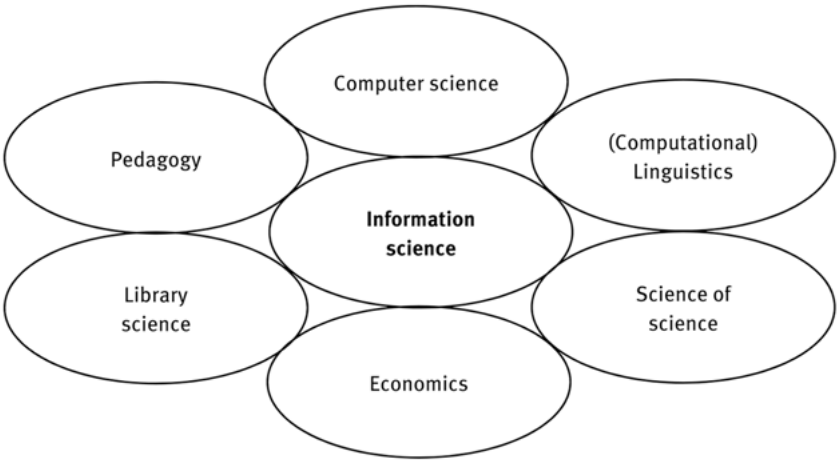


Figure A.1.2: Information Science and its Neighboring Disciplines.

Computer science provides the technological basis for information science applications. Without it, there would be no Internet and no commercial information industry. Some common research objects include information retrieval, aspects of knowledge representation (e.g., ontologies) and aspects of empirical information science, such as the evaluation of information systems. Computer science is more interested in the technology of information processing, information science in the processing of content. Unifying both perspectives suggests itself almost immediately. It has become common practice to refer to the combination between the two as Computer and Information Science—or "CIS".

On the level of enterprises, the streams of information and communication between employees as well as with their environment are of extreme importance. Organizing both information technology and the usage thereof are tasks of information systems research and of information management. Organizing the in-company knowledge base as well as the sharing and communication of information fall within the domain of knowledge management. Information management, information systems research and knowledge management together form the corporate information economy. On the industry level, we regard information as a product. Electronic information services, search engines, Web portals etc. serve to locate a market for the product called "knowledge". But this product has its idiosyncrasies, e.g. it can be copied at will, and many consumers regard it as a free common good. On the business level, lastly, network economics discusses the specifics of networks—of which the

Internet is one—and information economics asks about the buyer's and the seller's respective standards of knowledge, which already lays bare some considerable asymmetries.

The object of library science is the empirical and theoretical analysis of specific activities; among these are the collection, conservation, provision and evaluation of documents and the knowledge fixed therein. Its tools are elaborate systems for the formal and content-oriented processing of information. Topics like the creation of classification systems or information dissemination were common property of this discipline even before the term "information science" existed. This close link facilitates—especially in the United States—the development of approaches toward treating information science and library science as a single aggregate discipline, called "LIS" (Library and Information Science) (Milojević, Sugimoto, Yan, & Ding, 2011).

Search and retrieval are mainly performed via language; the knowledge—aside from some exceptions (such as images or videos)—is fixed linguistically. Computational linguistics and general linguistics are both so important for information science that relevant aspects from the different areas have been merged into a separate discipline, called information linguistics.

Pedagogy sometimes works with tools and services of Web 2.0, e.g. wikis or blogs. Additionally, in educational science learning management systems and e-portfolio systems are of great interest. Mainly in the aspects of e-learning and blended learning (Beutelspacher & Stock, 2011), there are overlaps with information science.

The analysis of scientific communication provided by empirical information science has made it possible to describe individual scientists, institutes, journals, even cities and countries. Such material is in strong demand in science of science and science policy. These results help sociology of science in researching for communities of scientists. History of science gets empirical historical source material. Science of science can check its hypotheses. And finally, science evaluation and science policy are provided with decision-relevant pointers toward the evaluation of scientific institutions.

Conclusion

- The object of information science is the representation, storage and supply as well as the search for and retrieval of relevant (predominantly digital) documents and knowledge.
- Information science consists of five main sub-disciplines: (1) information retrieval (science and engineering of search engines), (2) knowledge representation (science and engineering of the storage and representation of knowledge), (3) informetrics (including all metrics applied in information science), (4) endeavors on the information markets and the information society (researches in economics and sociology as parts of applied information science), and (5) efforts on knowledge management as well as on information literacy (researches in business administration and education).

- Some application cases of information science research results are search engines, Web 2.0 services, library catalogs, digital libraries, specialist information service suppliers as well as information systems in corporate knowledge management.
- The application of information science in the normal course of life is to secure information literacy for all people.
- Historically speaking, five main lines of development can be detected in information science: information retrieval, knowledge representation, informetrics, information markets / information society and knowledge management / information literacy. Information science has begun its existence as an independent discipline in the 1950s, and is deemed to have been firmly established since around 1970.
- Information science works between disciplines and has significant intersections with computer science, with economics, with library science, with linguistics, with pedagogy and with science of science.

Bibliography

Akerlof, G.A. (1970). The market for "lemons". Quality, uncertainty, and the market mechanism. Quarterly Journal of Economics, 84(3), 488-500.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern Information Retrieval. The Concepts and Technology behind Search. 2nd Ed. Harlow: Addison-Wesley.

Bates, M.J. (1999). The invisible substrate of information science. Journal of the American Society for Information Science, 50(12), 1043-1050.

Bawden, D., & Robinson, L. (2012). Introduction to Information Science. London: Facet.

Belkin, N.J. (1978). Information concepts for information science. Journal of Documentation, 34(1), 55-85.

Belkin, N.J., & Robertson, S.E. (1976). Information science and the phenomenon of information. Journal of the American Society for Information Science, 27(4), 197-204.

Bell, D. (1973). The Coming of the Post-Industrial Society. A Venture in Social Forecasting. New York, NY: Basic Books.

Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N., & Weitzner, D.J. (2006). A framework for web science. Foundations and Trends in Web Science, 1(1), 1-130.

Berners-Lee, T., Hendler, J.A., & Lassila, O. (2001). The semantic Web. Scientific American, 284(5), 28-37.

Beutelspacher, L., & Stock, W.G. (2011). Construction and evaluation of a blended learning platform for higher education. In R. Kwan, C. McNaught, P. Tsang, F.L. Wang, & K.C. Li (Eds.), Enhanced Learning through Technology. Education Unplugged. Mobile Technologies and Web 2.0 (pp. 109-122). Berlin, Heidelberg: Springer. (Communications in Computer and Information Science; 177).

Borko, H. (1968). Information science: What is it? American Documentation, 19(1), 3-5.

Bruce, C.S. (1999). Workplace experiences of information literacy. International Journal of Information Management, 19(1), 33-47.

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

Buckland, M.K. (1997). What is a "document"? Journal of the American Society for Information Science, 48(9), 804-809.

Buckland, M.K. (2012). What kind of science *can* information science be? Journal of the American Society for Information Science and Technology, 63(1), 1-7.

Castells, M. (1996). The Rise of the Network Society. Malden, MA: Blackwell.
Catts, R., & Lau, J. (2008). Towards Information Literacy Indicators. Paris: UNESCO.
Chu, H. (2010). Information Representation and Retrieval in the Digital Age. 2nd Ed. Medford, NJ: Information Today.
Chu, S.K.W. (2009). Inquiry project-based learning with a partnership of three types of teachers and the school librarian. Journal of the American Society for Information Science and Technology, 60(8), 1671-1686.
Cole, C. (2012). Information Need. A Theory Connecting Information Search to Knowledge Formation. Medford, NJ: Information Today.
Croft, W.B., Metzler, D., & Strohman, T. (2010). Search Engines. Information Retrieval in Practice. Boston, MA: Addison Wesley.
Cronin, B. (2008). The sociological turn in information science. Journal of Information Science, 34(4), 465-475.
Cronin, B., & Pearson, S. (1990). The export of ideas from information science. Journal of Information Science, 16(6), 381-391.
Davis, C.H., & Shaw, D. (Eds.) (2011). Introduction to Information Science and Technology. Medford, NJ: Information Today.
Dewey, M. (1876). A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library. Amherst, MA (anonymous).
Egghe, L. (2005). Power Laws in the Information Production Process. Lotkaian Informetrics. Amsterdam: Elsevier.
Eisenberg, M.B, & Berkowitz, R.E. (1990). Information Problem-Solving. The Big Six Skills Approach to Library & Information Skills Instruction. Norwood, NJ: Ablex.
Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). Ontological Engineering. London: Springer.
Gruber, T.R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199-220.
Haustein, S. (2012). Multidimensional Journal Evaluation. Analyzing Scientific Periodicals beyond the Impact Factor. Berlin, Boston, MA: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)
Hendler, J.A., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web science. An interdisciplinary approach to understanding the web. Communications of the ACM, 51(7), 60-69.
Henrichs, N. (1997). Informationswissenschaft. In W. Rehfeld, T. Seeger, & D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation (pp. 945-957). 4th Ed. München: Saur.
Ingwersen, P., & Järvelin, K. (2005). The Turn. Integration of Information Seeking and Retrieval in Context. Dordrecht: Springer.
Johnston, B., & Webber, S. (2003). Information literacy in higher education. A review and case study. Studies in Higher Education, 28(3), 335-352.
Kuhlen, R. (2004). Information. In R. Kuhlen, T. Seeger, & D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation (pp. 3-20). 5th Ed. München: Saur.
Lancaster, F.W. (2003). Indexing and Abstracting in Theory and Practice. 3rd Ed. Champaign, IL: University of Illinois.
Larivière, V., Sugimoto, C.R., & Cronin, B. (2012). A bibliometric chronicling of library and information science's first hundred years. Journal of the American Society for Information Science and Technology, 63(5), 997-1016.
Linde, F., & Stock, W.G. (2011). Information Markets. A Strategic Guideline for the I-Commerce. Berlin, New York, NY: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)

Luhn, H.P. (1961). The automatic derivation of information retrieval encodements from machine-readable texts. In A. Kent (Ed.), Information Retrieval and Machine Translation, Vol. 3, Part 2 (pp. 1021-1028). New York, NY: Interscience.

Machlup, F. (1962). The Production and Distribution of Knowledge in the United States. Princeton, NJ: Princeton University Press.

Manning, C.D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge: Cambridge University Press.

Mikhailov, A.I., Cernyi, A.I., & Gilyarevsky, R.S. (1979). Informatik. Informatik, 26(4), 42-45.

Milojević, S., Sugimoto, C.R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science. Analysis of article title words. Journal of the American Society for Information Science and Technology, 62(10), 1933-1953.

Mooers, C.N. (1952). Information retrieval viewed as temporal signalling. In Proceedings of the International Congress of Mathematicians. Cambridge, Mass., August 30 – September 6, 1950. Vol. 1 (pp. 572-573). Providence, RI: American Mathematical Society.

Nonaka, I., & Takeuchi, H. (1995). The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation. Oxford: Oxford University Press.

Norton, M.J. (2000). Introductory Concepts in Information Science. Medford, NJ: Information Today.

Otlet, P. (1934). Traité de Documentation. Bruxelles: Mundaneum.

Peters, I. (2009). Folksonomies. Indexing and Retrieval in Web 2.0. Berlin: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)

Porat, M.U. (1977). Information Economy. 9 Vols. (OT Special Publication 77-12[1] – 77-12[9]). Washington, DC: Office of Telecommunication.

Presidential Committee on Information Literacy (1989). Final Report. Washington, DC: American Library Association / Association for College & Research Libraries.

Priem, J., & Hemminger, B. (2010). Scientometrics 2.0. Toward new metrics of scholarly impact on the social Web. First Monday, 15(7).

Probst, G.J.B., Raub, S., & Romhardt, K. (2000). Managing Knowledge. Building Blocks for Success. Chichester: Wiley.

Ranganathan, S.R. (1931). The Five Laws of Library Science. Madras: Madras Library Association, London: Edward Goldston.

Rauch, W. (1988). Was ist Informationswissenschaft? Graz: Kienreich.

Salton, G., & McGill, M.J. (1983). Introduction to Modern Information Retrieval. New York, NY: McGraw-Hill.

Saracevic, T. (1999). Information Science. Journal of the American Society for Information Science, 50(12), 1051-1063.

Schmitz, J. (2010). Patentinformetrie. Analyse und Verdichtung von technischen Schutzrechtsinformationen. Frankfurt/Main: DGI.

Schrader, A.M. (1984). In search of a name. Information science and its conceptual antecedents. Library and Information Science Research, 6(3), 227-271.

Shapiro, F.R. (1995). Coinage of the term *information science*. Journal of the American Society for Information Science, 46(5), 384-385.

Staab, S., & Studer, R. (Eds.) (2009). Handbook on Ontologies. Dordrecht: Springer.

Stock, W.G. (2000). Informationswirtschaft. Management externen Wissens. München: Oldenbourg.

Stock, W.G. (2007). Information Retrieval. Informationen suchen und finden. München: Oldenbourg.

Stock, W.G. (2011). Informational cities. Analysis and construction of cities in the knowledge society. Journal of the American Society for Information Science and Technology, 62(5), 963-986.

Stock, W.G., Peters, I., & Weller, K. (2010). Social semantic corporate digital libraries. Joining knowledge representation and knowledge management. Advances in Librarianship, 32, 137-158.

Stock, W.G., & Stock, M. (2008). Wissensrepräsentation. Informationen auswerten und bereitstellen. München: Oldenbourg.
Taylor, A.G. (1999). The Organization of Information. Englewood, CO: Libraries Unlimited.
Thelwall, M. (2009). Introduction to Webometrics. Quantitative Web Research for the Social Sciences. San Rafael, CA: Morgan & Claypool.
van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. The Information Society, 19(4), 315-326.
van Raan, A.F.J. (1997). Scientometrics. State-of-the-art. Scientometrics, 38(1), 205-218.
Voorhees, E.M. (2002). The philosophy of information retrieval evaluation. Lecture Notes in Computer Science, 2406, 355-370.
Weber, S., & Stock, W.G. (2006). Facets of informetrics. Information – Wissenschaft und Praxis, 57(8), 385-389.
Webster, F. (1995). Theories of the Information Society. London: Routledge.
Weller, K. (2010). Knowledge Representation in the Social Semantic Web. Berlin: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)
Wilson, P. (1996). The future of research in our field. In J.L. Olaisen, E. Munch-Petersen, & P. Wilson (Eds.), Information Science. From the Development of the Discipline to Social Interaction (pp. 319-323). Oslo: Scandinavian University Press.
Wilson, T.D. (2000). Human information behavior. Informing Science, 3(2), 49-55.
Wolfram, D. (2003). Applied Informetrics for Information Retrieval Research. Westport, CT: Libraries Unlimited.
Yan, X.S. (2011). Information science. Its past, present and future. Information, 2(3), 510-527.

A.2 Knowledge and Information

Signals and Data

"Information" and "Knowledge" are concepts that are frequently discussed in the context of many different scientific disciplines and in daily life (Bates, 2005; Capurro & Hjørland, 2003; Lenski, 2010). They are crucial basic concepts in information science.

The concept of information was decisively influenced by Shannon (1948), during the advent of computers and of information transmission following the end of the Second World War. Shannon's interpretation of this concept was to become the foundation of telecommunications, and continues to influence all areas of computer and telecommunication technology. Wyner (2001, 56) emphasizes Shannon's importance for science and technology:

> Claude Shannon's creation in the 1940's of the subject of information theory is arguably one of the great intellectual achievements of the twentieth century. Information theory has had an important and significant influence on mathematics, particularly on probability theory and ergodic theory, and Shannon's mathematics is in its own a considerable and profound contribution to pure mathematics. But Shannon did his work primarily in the context of communication engineering, and it is in this area that it stands as a unique monument. In his classical paper of 1948 and its sequels, he formulated a model of a communication system that is distinctive for its generality as well as for its amenability to mathematical interest.

Let us take a look at Shannon's model (Figure A.2.1). In it, information is transmitted as a physical signal from an information source to a receiver via a channel.

Signals are physical entities, i.e. newsprint (on the paper you are reading), sound waves (from the traffic outside that keeps you from reading) or electromagnetic waves (from the television). The channel is prone to disruptions, and thus contains "noise" (e.g. when there is an error on one page).

An "encoder" is activated in the transmitter, between the source and the channel, transforming the digits into certain physical signals. Concurrently, a "decoder" that can interpret these signs is established between channel and receiver. Encoder and decoder must be attuned to each other in order for the signals to be successfully transmitted. As an illustrative example, let us consider a telephone conversation. Here, the information source is a person communicating a message via the microphone of their telephone. In the transmitter (the telephone), acoustic signals are transformed into electronical ones and then sent forth via telephone lines. At the other end of the line, a receiver transforms the electronical signals back into acoustic ones and sends them on to their destination (here: the person being called) via the second telephone.

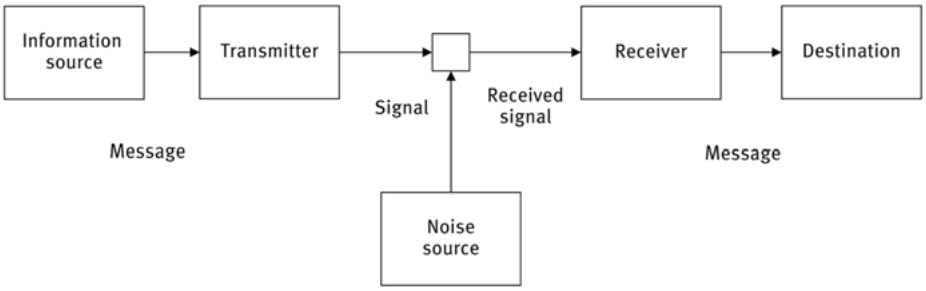


Figure A.2.1: Schema of Signal Transmission Following Shannon. Source: Shannon, 2001 [1948], 4.

The information content of a sign is calculated via this sign's probability of occurrence. Signs that occur less often receive high information content; frequently occurring ones are assigned a small value. When a sign z_i has the probability p_i of occurring, its information content I will be:

$$I(z_i) = \text{ld } 1/p_i \text{ bit or (mathematically identical):}$$
$$I(z_i) = -\text{ld } p_i \text{ bit.}$$

The logarithm dualis being used (ld) and the measure "bit" (for 'binary digit') can both be explained by the fact that Shannon considers a sign's probability of occurrence to be the result of a sequence of binary decisions. The ld calculates the exponent x in the formula $2^x = a$ (e.g. $2^3 = 8$; hence, in the reverse function, ld 8 = 3). Let us clarify the principle via an example: we would like to calculate the information content I for signs that have a probability of occurrence of 17% (like the letter *e* in the German language), of 10% (like *n*) and of 0.02% (like *q*). The sum of the probabilities of occurrence of all signs in the presupposed repertoire, which in this case is the German alphabet, must always add up to 100%. We apply the values to the formula and receive these results:

$$I(e) = -\text{ld } 0.17 \text{ bit} = -(-2.6) \text{ bit} = 2.6 \text{ bit}$$
$$I(n) = -\text{ld } 0.10 \text{ bit} = -(-3.4) \text{ bit} = 3.4 \text{ bit}$$
$$I(q) = -\text{ld } 0.0002 \text{ bit} = -(-12.3) \text{ bit} = 12.3 \text{ bit.}$$

A *q* thus has far higher information content than an *e*, since it is used far less often by comparison.

The following observation by Shannon is of the utmost importance for information science (2001 [1948], 3):

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that

> is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.

Here he is merely talking about a probability-theoretical observation of signal transmission processes; the meaning, i.e. the content of the transmitted signals, does not play any role at all (Ma, 2012, 717-719). It is the latter, however, which forms the object of information science. In so far, Shannon's information theory is of historical interest for us but it has only little significance in information science (Belkin, 1978, 66). However, it is important for us to note that any information is tied to a signal, i.e. a physical carrier, as a matter of principle.

Signals transmit signs. In semiotics (which is the science of signs), signs can be observed more closely from three points of view:

- in their relations to each other (syntax),
- in the relations between signs and the objects they describe (semantics),
- in the relations between signs and their users (pragmatics).

Shannon's information theory only takes into consideration the syntactic aspect. From an information science perspective, we will call this aspect data. Correspondingly, the processing of data concerns the syntactical level of signs, which are analyzed with regard to their type (e.g. alphanumerical signs) or their structure (e.g. entry in a field named "100"), among other aspects. This is only information science from a peripheral point of view, though. This sort of task falls much more under the purview of computer science, telecommunications and telematics (as a mixed discipline formed of telecommunications and informatics). Boisot and Canals (2004, 43) discuss the difference between "data" and "information" via the example of cryptography:

> Effective cryptography protects information as it flows around the word. … Thus while the data itself can be made "public" and hence freely available, only those in possession of the "key" are in a position to extract information from it (…). Cryptography, in effect, exploits the deep differences between data and information.

In information science, we analyze information as data *and* meaning (knowledge) in context. For Floridi (2005), information consists of data, which is well-formed as well as meaningful.

Information, Knowledge and Documents

The other two sub-disciplines of semiotics, semantics and pragmatics, investigate the meaning and usage of signs. Here the concepts of "knowledge" and "information" enter the scene. Following Hardy and Meier-Oeser (2004, 855), we understand knowledge to be:

> partly a skill, i.e. the skill of conceiving of an object as it really is on the one hand, and that of successfully dealing with the objects of knowledge on the other. It is partly the epistemic state that a person occupies as a consequence of successfully performing his or her cognitive tasks, and partly the content that a cognitive person refers to when doing so, as well as the statement in which one gives linguistic expression to the result of a cognitive process.

Here, knowledge is assigned the trait of certainty from the perspective of a subject; independently of the subject, knowledge is accorded a claim for truth. We should reanalyze the above definition of knowledge and structure it simplistically. Knowledge falls into the two aspects of skill and state:

(1) Knowledge as the skill
   - (1a) of correctly comprehending an object ("to know *that*"),
   - (1b) of correctly dealing with an object ("to know *how*");

(2) Knowledge as the state
   - (2a) of a person that knows (something),
   - (2b) that which is known itself, the content,
   - (2c) its linguistic expression.

Knowledge has two fixed points: the knowing subject and what is known, in which aspects (1a), (1b) and (2a) belong to subjective knowledge and (2b) and (2c) are objective knowledge—of course, (2c) can only be objective if the linguistic expression is permanently fixed on a physical carrier, a document.

The warrantor of the distinction between subjective and objective knowledge is Popper (1972; see also Brookes, 1980; Ma, 2012, 719-720). According to him, objective knowledge is knowledge that is fixed in documents, such as articles or books, whereas subjective knowledge is the knowledge that a person has fixed in their mind. This distinction is embedded within Popper's theory of the three worlds:

- World 1: the physical world,
- World 2: the world of conscious experience,
- World 3: the world of content.

World 1 is material reality; this world contains the signals that are known as data or documents when bundled accordingly. Subjective knowledge (World 2) is clearly fixated on objective knowledge (World 3): practically all subjective knowledge (World 2) depends upon World 3. It is also clear, conversely, that objective knowledge develops from knowledge that was formerly subjective. There is a close interrelationship between the two: subjective knowledge orients itself upon objective knowledge, while objective knowledge results from the subjective. And without the physical world—World 1—"nothing goes", as Brookes (1980, 127) emphasizes:

> Though these three worlds are independent, they also interact. As human living on Earth, we are part of the physical world, dependent for our continued existence on heat and light from the Sun ... and so on. Though our mentalities we are also part of World 2. In reporting the ideas Popper has recorded in his book, I have been calling on the resources of World 3. Books and all other

> artifacts are also physical entities, bits of World 1, shaped by humans to be exosomatic stores of knowledge which have an existence as physical things independent of those who created them.

Objective knowledge, or our point (2b) above, exists independently of people (in Popper's World 3). Semiotically speaking, we are in the area of semantics, which is known for disregarding the user of the sign, the subject. How does the transition from World 3 to World 2, or within World 2, take place? The (formless) content from World 3 must be "cast into a mould", put into a *form*, that can be understood in World 2. The same procedure is performed when attempting to transmit knowledge between two subjects in World 2. We are thus dealing with in*FORM*ation, if you'll excuse the pun. It is not possible to transmit knowledge "as such", we need a carrier (the "form") to set the knowledge in motion. "Information is a thing—knowledge is not," Jones (2010) declares.

Indeed the word information, etymologically speaking, has exactly this origin: according to Capurro (1978) and Lenski (2010, 80), the Latin language uses "informatio" and "informo" to mean education and formation, respectively. (Information is thus related to schooling; for a long time, the "informator" referred to a home tutor.) Information makes content move, thus also having a pragmatic component (Rauch, 2004). The groundbreaking work of economic research into the information market is "The Production and Distribution of Knowledge in the United States" (1962) by Machlup. Machlup was one of the first to formulate knowledge as static and information as dynamic. Knowledge is not transmitted: what is sent and received is always information. Machlup (1962, 15) defines:

> to *inform* is an activity by which knowledge is conveyed; to *know* may be the result of having been informed. "Information" as the act of informing is designed to produce a state of knowing in someone's mind. "Information" as that which is being communicated becomes identical with "knowledge" in the sense of which is known. Thus, the difference lies not in the nouns when they refer to *what* one knows or is informed about; it lies in the nouns only when they are to refer to the *act* of informing and the *state* of knowing, respectively.

Kuhlen (1995, 34) proceeds similarly: "*Information is knowledge in action*." Kuhlen continues:

> Correspondingly, from an information science perspective we are interested in methods, procedures, systems, forms of organization and their respective underlying conditions. With the help of these, we can work out information for current problem solutions from socially produced knowledge, or produce new knowledge from information, respectively. The process of working out information does not leave knowledge in its *raw state*; rather, it should be regarded as a process of transformation or, with a certain valuation, of refinement ... This transformation of knowledge into information is called the creation of informational added value.

Knowledge has nothing to do with signals, since it exists independently of them, whereas information is necessarily tied to these physical carriers.

Bates (2005; 2006) distinguishes three levels of information and knowledge in regard to matter and energy ("information 1"; e.g., genetic information), to beings, especially to men ("information 2") and to meaning and understanding ("knowledge"). These levels are similar to the three worlds of Popper. Bates (2006, 1042) defines

> Information 1: The pattern of organization of matter and energy.
> Information 2: Some patterns of organization of matter and energy by a living being (or its constituent parts).
> Knowledge: Information given meaning and integrated with other contents of understanding.

If we put knowledge into carriers of Popper's World 1 (e.g., into a book or a Web page), the meaning is lost. A reader of the book or the Web page has to create his understanding of the meaning on his own. Bates (2006) emphasizes:

> *Knowledge* in inanimate objects, such as books, is really only information 1, a pattern of organization of matter and energy. When we die, our personal knowledge dies with us. When an entire civilization dies, then it may be impossible to make sense out of all the information 1 left behind; that is, to turn it into information 2 and then knowledge.

According to Buckland (1991b), information has four aspects, of which three fall under the auspices of information science and one (at least additionally) to computer science. Information is either tangible or intangible, it can either be physically grasped or not; in Popper's sense, it is either tied to World 1 or it is not. Additionally, information is always either a process or a state. This two-fold distinction leads to the following four-field chart (modified following Buckland, 1991b, 352):

Information	*Intangible*	*Tangible*
State	Knowledge	Information as thing
Process	Information as process	Information processing

Information is tangible as a thing in so far as the things, i.e. the documents, are always tied to signals, and thus to World 1. They are to be regarded—at least over a certain period of time—as stable. Processes for dealing with information can also be physically described and must thus be allocated to World 1. Here Buckland locates the domain of information processing and data processing, and hence a discipline that is more a part of computer science. Knowledge is intangible; it falls either to World 2 (as subjective knowledge) or World 3 (as objective knowledge). In the process of informing or being informed (information as process), Buckland abstracts from the physical signal and only observes the transitions of knowledge from a sender to a receiver. Buckland is also aware that such transitions always occur in the physical world (Buckland, 1991b, 352):

> Knowledge … can be represented, just as an event can be filmed. However, the representation is no more knowledge than the film is the event. Any such representation is necessarily in tangible forms (sign, signal, data, text, film, etc.) and so *representations* of knowledge (and of events) are necessarily "information-as-thing".

Here the great significance of documents (in Buckland's words: information as thing) for information science becomes clear, since they are what shelters knowledge. In the documents, knowledge (which is not tangible as such) is embedded. For Mooers (1952, 572), the knowledge is a "message", and the documents containing those messages are the "channels":

> (A) "channel" is the physical document left in storage which contains the message.

As knowledge is not directly tangible, how can it then be fixed and recognized? Apparently, this requires a subject to perform a knowledge act: knowledge is not simply given but is acquired on the basis of the respective subject's foreknowledge.

How can knowledge be recorded and presented, and what are its exact characteristics? In the following sections, we will investigate the manifold aspects of knowledge. At first, we will concentrate on the question: can only texts contain knowledge, or are other document types, such as pictures, capable of this as well?

Non-Textually Fixed Knowledge

Knowledge fixed in documents does not exclusively have to be grounded in texts. Non-textual documents—images, videos, music—also contain knowledge. We will demonstrate this on the example of pictures and Panofsky's theory (2006). The reader should call to mind Leonardo da Vinci's famous painting *The Last Supper*. According to Panofsky, there exist three levels on which to draw knowledge from the painting. An Australian bushman—as an example for "everyman"—will look at the picture and describe its content as "13 persons either sitting at or standing behind one side of a long table." For Panofsky, this semantic level is "pre-iconographical"; it only presupposes practical experiences and a familiarity with certain objects and events on the part of the viewer. On the second semantic level—the "iconographical" level—the viewer additionally requires foreknowledge concerning cultural traditions and certain literary sources, as well as familiarity with the thematic environment of the painting.

Armed with this foreknowledge, a first viewer will identify 13 men, one of which being Jesus Christ and the other 12 his disciples. Another viewer, with different foreknowledge (perhaps nourished by Dan Brown's novel *The Da Vinci Code*), however, sees 12 men and a woman, by replacing the depiction of John—to Jesus' left as seen from the viewer's perspective—with that of Mary of Magdala. On this level, it also

depends upon who is looking at the picture, or more precisely, what background knowledge the viewer has of the picture. On the iconographical level, one recognizes the last supper of Christ with his disciples (or with 11 disciples and a woman).

The third, or "iconological" knowledge level can be reached via expert knowledge. Our exemplary image is now interpreted as a particularly successful work of Leonardo's (due, among other aspects, to the precise rendering of perspective), as a textbook example of the culture of the Italian Renaissance, or as an expression of the Christian faith. Panofsky (1975, 50) summarizes the three levels in tabular form:

> I *Primary*, or *natural* subject—(A) factual, (B) expressive—that forms the world of *artistic motifs*.
> II *Secondary*, or *conventional* subject, which forms the world of *pictures*, *anecdotes* and *allegories*.
> III *Actual Meaning* or *content*, forming the world of "*symbolic*" *values*.

In correspondence to these three levels of knowledge, there are the three interpretative acts of (I) pre-iconographical description, (II) iconographical analysis and (III) iconological interpretation. The three levels of pictorial knowledge thus described should, analogously, be distinguishable in other media (films and music). In certain texts, too, particularly in fiction, it seems appropriate to divide access to knowledge in three. The story of the fox and the grapes (by Aesop) contains the primary knowledge that there is a fox who sees some grapes dangling above him, unreachable, and that the fox then rejects them, citing their sour taste. The secondary knowledge sees the fable as an example of the way people develop resentment out of impotence; and level three literary-historically places the text into the context of antique fables.

Knowing That and Knowing How

A simple initial approximation regards knowledge as instances of true propositions. The following definition of knowledge (Chisholm, 1977, 138) holds firm in some varieties of epistemology:

> *h is known* by *S* =Df *h* is accepted by *S*; *h* is true; and *h* is nondefectively evident for *S*.

h is a proposition and *S* a subject; *=Df* means "equals by definition". Hence, Chisholm demands that the subject accepts the proposition *h* (as true), which is in fact the case (objectively speaking) and that this is so not merely through a happy coincidence, but precisely "nondefectively evident". Only when all three determinants (acceptance, truth, evidence) are present can knowledge be seen as well and truly established. In the absence of one of these aspects, such a statement can still be communicated—as information—but it would be an error (when truth and evidence are absent), a supposition (if acceptance and evidence are given, but the truth value is undecided) or a lie (when none of the three aspects applies).

Is knowledge always tied to statements, as Chisholm's epistemology suggests? To the contrary: the propositional view of knowledge falls short, as Ryle (1946, 8) points out:

> It is a ruinous but popular mistake to suppose that intelligence operates only in the production and manipulation of propositions, i.e., that only in ratiocinating are we rational.

Propositions describe "knowing that" while neglecting "knowing how". But knowledge is not exclusively expressed in propositions; and knowing how is most often (and not only for Ryle) the more important aspect of knowledge by far. Know-how is knowledge about how to do certain things. Ryle distinguishes between two forms of know-how: (a) purely physical knowledge, and (b) knowledge whose execution is guided by rules and principles (which can be reconstructed). Ryle (1946, 8) describes the first, physical variant of know-how as follows:

> *(a)* When a person knows how to do things of a certain sort (e.g., make good jokes, conduct battles or behave at funerals), his knowledge is actualised or exercised in what he does. It is not exercised (...) in the propounding of propositions or in saying "Yes" to those propounded by others. His intelligence is exhibited by deeds, not by internal or external dicta.

In the second variant of know-how, knowledge is grounded in principles, which can be described at least in their essence (Ryle, 1946, 8):

> *(b)* When a person knows how to do things of a certain sort (e.g., cook omelettes, design dresses or persuade juries), his performance is in some way governed by principles, rules, canons, standards or criteria. (...) It is always possible in principle, if not in practice, to explain why he tends to succeed, that is, to state the reasons for his actions.

According to Ryle, it is not possible to trace such "implicit" know-how (Ryle, 1946, 7) to know-that, and thus to propositions. In the know-how variant (b), it is not impossible to reconstruct and objectify the implicit knowledge via rules etc.; this is hardly the case in variant (a).

Subjective Implicit Knowledge

Polanyi enlarges upon Ryle's observations on know-how. According to Polanyi, people dispose of more knowledge than they are capable of directly and comprehensibly communicating to others. Polanyi's (1967, 4) famous formulation is this:

> I shall consider human knowledge by starting from the fact that *we can know more than we can tell*. This fact seems obvious enough; but it is not easy to say exactly what it means.

Implicit, or tacit, knowledge consists of various facets (Polanyi, 1967, 29):

> The things that we know in this way included problems and hunches, physiognomies and skills, the use of tools, probes, and denotative language.

This implicit knowledge is basically "embedded" within the body of the individual, and they use it in the same way that they normally use their body (Polanyi, 1967, X and XI):

> (The structure of tacit knowing) shows that all thought contains components of which we are subsidiary aware in the focal content of our thinking, and that all thought dwells in its subsidiaries, as if they were parts of our body. ...(S)ubsidiaries are used as we use our body.

We must strictly differentiate between the meaning that the person might ascribe to the (implicit) knowledge, and the object bearing this meaning. Let the person carry an object in their hand, say: a tool; this object is thus close to the person (proximal). The meaning of this object, on the other hand, may be very far from the person (distal)—it does not even have to be known in the first place. The meaning of the object (unexpressed or inexpressible) arises from its usage. Polanyi (1967, 13) introduces the two aspects of implicit knowledge via an example:

> This is so ... when we use a tool. We are attending to the meaning of its impact on our hands in terms of its effect on the things to which we are applying it. We may call this the *semantic aspect* of tacit knowledge. All meaning tends to be displayed *away from ourselves*, and that is in fact my justification for using the terms "proximal" and "distal" to describe the first and second terms of tacit knowledge.

Explicit knowledge can also contain implicit components. The author of a document has certain talents, is well acquainted with certain subjects and is steeped in traditions, all of which put together makes up his or her personal knowledge (Polanyi, 1958). Since these implicit factors are unknown to the reader, the communication of personal knowledge—even in written and thus explicit form—is always prone to errors (Polanyi, 1958, 207):

> Though these ubiquitous tacit endorsements of our words may always turn out to be mistaken, we must accept this risk if we are ever to say anything.

Here Polanyi touches upon hermeneutical aspects (e.g. understanding and preunderstanding).

How can implicit knowledge be transmitted? In the case of distal implicit knowledge, transmission is nearly impossible, whereas proximal implicit knowledge, according to Polanyi, is communicated via two methods. On the one hand, it is possible to "physically" transmit knowledge via demonstration and imitation of the relevant activities (Polanyi, 1967, 30):

> The performer co-ordinates his moves by dwelling in them as parts of his body, while the watcher tries to correlate these moves by seeking to dwell in them from outside. He dwells in these moves by interiorizing them. By such exploratory indwelling the pupil gets the feel of a master's skill and may learn to rival him.

The second option consists of intellectually appropriating the implicit knowledge of another person by imitation, or more precisely, by rethinking. Polanyi (1967, 30) uses the example of chess players:

> Chess players enter into a master's spirit by rehearsing the games he played, to discover what he had in mind.

According to Polanyi, it is important to create "coherence" between the bearer of implicit knowledge and the person who wants to acquire it—be it of the physical or intellectual kind (Polanyi, 1967, 30):

> In one case we meet a person skillfully using his body and, in the other, a person cleverly using his mind.

For implicit knowledge to be communicated without a hitch, it must be transformed into words, models, numbers etc. that can be understood by anyone. In other words, it must be "externalized" (ideally: stored in written documents). Externalized knowledge—and only this—is accessible to knowledge representation, and hence information science. In practice, according to Nonaka and Takeuchi (1995), this is not achieved via clearly defined concepts and statements that can be understood by anyone, but only via rather metaphorical and vague expressions. Should it prove impossible to comprehensibly externalize a person's implicit knowledge, we will be left with four more ways toward at least conserving some aspects of their knowledge:

- firstly, we regard the person themselves as a document which is then described via metadata (e.g. in an expert database),
- secondly, we draw upon those artifacts created by the person themselves and use them to approximate their initial knowledge (as in the chess player example above), or we attempt—as in Ryle's scenario (b)—to reconstruct the rules governing the lines of action,
- thirdly, we "apprentice" to the person (and physically appropriate their knowledge),
- fourthly, it is possible to compile information profiles of the person via their preferences (e.g. of reading certain documents), and to extrapolate their implicit knowledge on this basis (with the caveat of great insecurity).

Variant (1) amounts to "Yellow Pages", which ultimately become useless once the person is no longer reachable (e.g. after leaving his company), variant (2) tries to make do with documenting the artifacts (via images, videos and descriptions) and hopes that will be possible for another person to reconstruct the original knowledge

on their basis. Variant (4) is, at best, useful as a complement to the expert database. The ideal path is probably variant (3), which involves being introduced to the subject area under the guidance of the individual in question. Here, Nonaka and Takeuchi (1995) talk about "socialization". Memmi (2004, 876) finds some clear words on the subject:

> In short, know-how and expertise are only accessible through contact with the appropriate individuals. Find the right people to talk to or to work with, and you can start acquiring their knowledge. Otherwise there is simply very little you could do (watching videos of expert behavior is a poor substitute indeed).

Socialization is an established method in knowledge management, but it has nothing to do with digital information and very little with information science (since in this instance, knowledge is *not* being represented at all, but directly communicated from person to person). From the purview of information science, then, implicit knowledge creates a problem. We can approach it in the context of knowledge representation (using methods 1, 2 and 4), but we cannot conclusively resolve it. Implicit knowledge already creates problems on the level of interpersonal relationships; if we additionally require an information system in knowledge representation, the problem will become even greater. Thus, Reeves and Shipman (1996, 24) emphasize:

> Humans make excellent use of tacit knowledge. Anaphora, ellipses, unstated shared understanding are all used in the service of our collaborative relationships. But when human-human collaboration becomes human-computer-human collaboration, tacit knowledge becomes a problem.

Knowledge Management

In Nonaka's and Takeuchi's (1995) approach, we are confronted with four transitions from knowledge to knowledge:

- from the implicit knowledge of one person to that of another (socialization),
- from the implicit knowledge of a person to explicit knowledge (externalization)—which includes the serious problems mentioned above,
- from explicit knowledge to the implicit knowledge of a person (internalization, e.g. by learning),
- from explicit knowledge to explicit knowledge (combination).

Information science finds its domain in externalization (where possible) and combination. Explicit, and thus person-independent, knowledge is conserved and represented in its objective interconnections. For Nonaka, Toyama and Konno (2000), the ideal process of knowledge creation in organizations runs, in a spiral form, through all four forms of knowledge transmission and leads to the SECI Model (*S*ocialization, *E*xternalization, *C*ombination, *I*nternalization) (see Figure A.2.2). In order to exchange

knowledge, the actors must occupy the same space (Japanese: "ba"), comparable to the "World Horizon" of hermeneutics (Nonaka, Toyama, & Konno, 14):

> *ba* is here defined as a shared context in which knowledge is shared, created and utilised. In knowledge creation, generation and regeneration of ba is the key, as ba provides the energy, quality and place to perform the individual conversions and to move along the knowledge spiral.

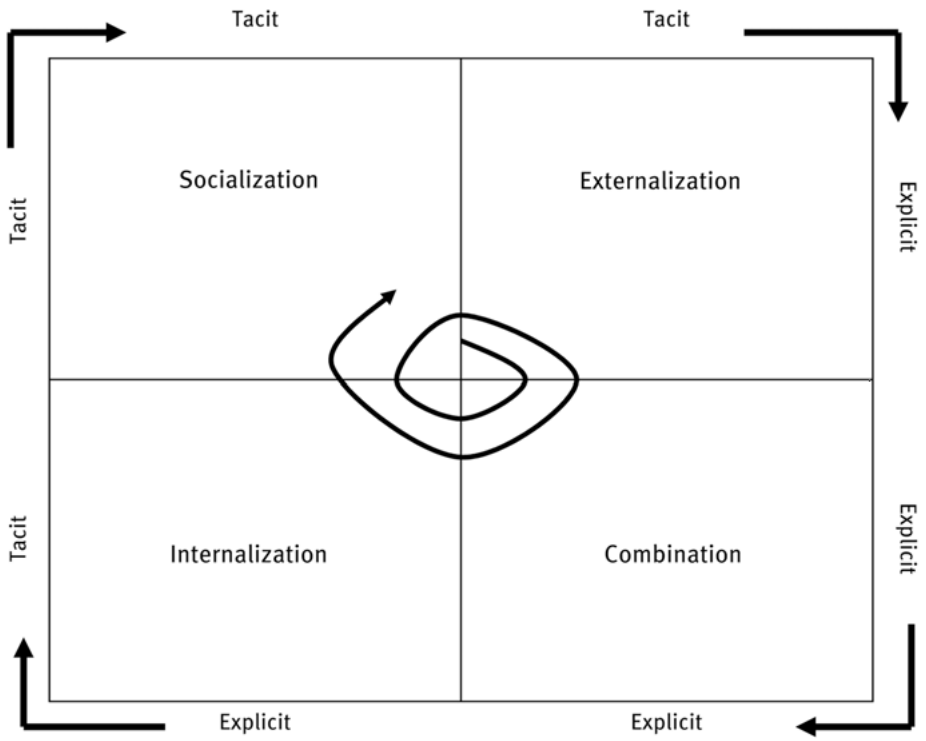


Figure A.2.2: Knowledge Spiral in the SECI Model. Source: Modified from Nonaka, Toyama & Konno, 2000, 12.

In enterprises, several building blocks are required in order to adequately perform knowledge management (Figure A.2.3). According to Probst, Raub and Romhardt (2000), we must distinguish between the work units (at the bottom) and the strategic level. The latter dictates which goals should be striven for in the first place. Then, the first order of business is to identify the knowledge required by the company. This can already be available internally (to be called up from knowledge conservation) or externally (on the Web, in a database, from an expert or wherever) and will then be acquired. If no knowledge is available, the objective will be to create it by one's self (e.g. via research and development). Knowledge is meant to reach its correct recipi-

ents. To accomplish this, the producers must first be ready to share their knowledge, and secondly, the system must be rendered capable of adequately addressing, i.e. distributing, said knowledge. Both knowledge compiled in enterprises and crucially important external knowledge must be kept easily accessible. The central aspect is the building block of Knowledge Use: the recipient translates the knowledge into actions, closes knowledge gaps, prepares for decisions or lets the system notify him—in the sense of an early-warning system—about unexpected developments. Finally, Knowledge Measurement evaluates the operative cycle and sets the results in relation to the set goals.

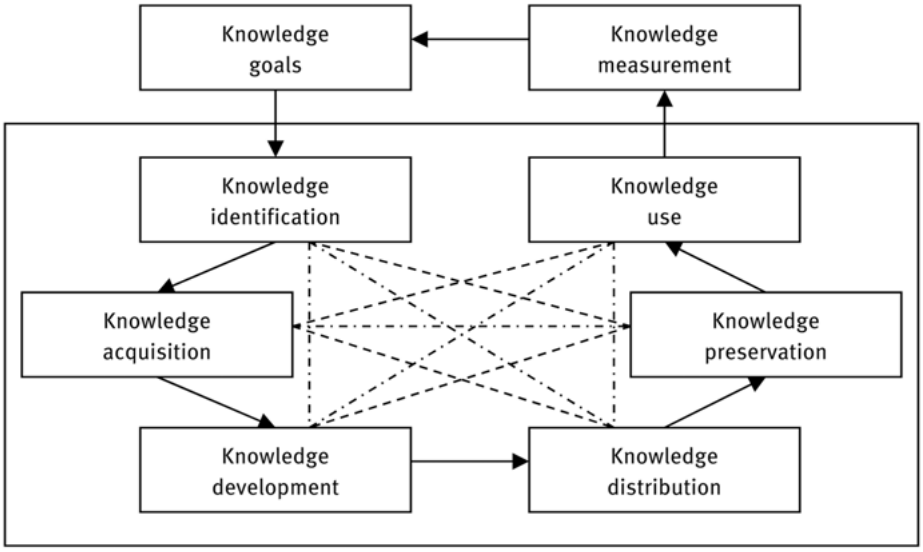


Figure A.2.3: Building Blocks of Knowledge Management Following Probst et al. Source: Probst, Raub, & Romhardt, 2000, 34.

Types of Knowledge

For Spinner (1994, 22), knowledge has become so important—we need only consider the "knowledge society"—that he places the knowledge order on the same level as the economic order and the legal order of a society:

> The outstanding importance of these *three fundamental orders* for all areas of society results, ... in the case of the *knowledge order*, from the function of scientific-technological progress as the most important productive force, as well as from extra-scientific information as a mass medium of communication and control, i.e. as a means of entertainment and administration.

Such a knowledge order contains "knowledge of all kinds, in every conceivable quantity and quality" (Spinner, 1994, 24), which is granted by the triad of form, content and expression, as well as via the validity claim. Studying the form involves the logical form of statements and their area of application, where Spinner distinguishes dichotomously between general statements (statements in theories or laws, e.g. "All objects, given gravitation, fall downward") and singular existential statements with reference to place and time ("My pencil falls downward in our office on December 24th, 2007"). Knowledge, according to Spinner, is devoid of content if it has only slight informational value (e.g. entertainment or advertisement). On the other hand, it is "informative" in the case of high informational value (e.g. in scientific statements or news broadcasts). The expression of knowledge aims for the distinction between implicit (tacit knowledge, physical know-how) and explicit (fixed in documents). The application of knowledge can also be considered as an epistemic "additional qualification" (Spinner, 2002, 21). An application can be apodictic if the knowledge concerned is presupposed to be true (as in dogmas), or hypothetical if its truth value is called into question.

Can all knowledge be equally well structured, or represented? Certainly not. A well-formulated scientific theory, published in an academic journal, should be more easily representable than a microblog on Twitter or a rather subliminal message in an advertising banner. Spinner (2000, 21) writes on this subject:

> We can see that there are knowledge types, knowledge stores and knowledge characteristics that can be ordered and those that resist being ordered by looking at the species-rich knowledge landscape. This landscape reaches from implicit notions and unarticulated ideas all the way to fully formulated ('coded') legal texts and explicit ('holy') texts that are set down word for word.

Normal-Science Knowledge

Let us restrict our purview to scientific knowledge for a moment. According to Kuhn's (1962) deliberations, science does not advance gradually but starts its research—all over again, as it were—in the wake of upheavals, or scientific revolutions. Kuhn describes the periods between these upheavals as "normal sciences". We must understand these normal sciences as a type of research that is accepted by scientists of certain schools of thought as valid with regard to their own work. Research is based on scientific accomplishments of the past and serves as a basis for science; it represents commonly accepted theories which are applied via observation and experimentation. The scientific community is held together by a "paradigm", which must be sensational in order to draw many advocates and adherents, but which also offers its qualified proponents the option of solving various problems posed by the paradigm itself. The researchers within a normal science form a community that orients itself on the common paradigm, and which disposes of a common terminology dictated by the

paradigm. If there is agreement within the community of experts as to the specialist language being used, it is easily possible—at least in theory—to represent the specialist language via exactly one system of a knowledge order. Kuhn (1962, 76) emphasizes the positive effects of such a procedure of not substituting established tools:

> So long as the tools a paradigm supplies continue to prove capable of solving the problems it defines, science moves fastest and penetrates most deeply.

For the record: In normal sciences, specific knowledge orders can be worked out for the benefit of science. But there is more than just the normal sciences. Sometimes, anomalies arise—observations that do not fit the paradigm. If these anomalies pile up, the scientific community enters a critical situation in which the authority of the paradigm slackens. If the scientists encounter a new paradigm during such a crisis, a scientific revolution is to be expected—a paradigm shift, a change in perspectives, the formation of a new scientific community. The new paradigm must offer predictions that differ from those of the old (anomalous) one, otherwise it would be unattractive for the researchers. For Kuhn, however, this also means that the old paradigm and the new one are logically irreconcilable: the new one supplants the old. Both paradigms are "incommensurable", i.e. they cannot be compared to each other. For the knowledge order of the old paradigm, this means that it has become as useless as the paradigm itself, and must be replaced.

There are sciences that have not (yet) found their way toward becoming normal sciences. These are in their so-called "pre-paradigmatic" phase, in which individual schools of thought, or "lone wolves", attempt to solve the prevalent problems via their own respective terminologies. Since there is no binding terminology in such cases, it is impossible to build up a unified knowledge order.

Brier (2008, 422) emphasizes the importance of different languages in scientific fields for information practice:

> (A) bibliographical system such as BIOSIS will only function well within a community of biologists. This means that both their producers and the users must be biologists—and so must be the indexers.

If for example a chemist is looking for biological information, he cannot use his chemical terminology but must speak the language of biology (Brier, 2008, 424):

> (C)hemists must use the correct biological name for a plant in order to find articles about a chemical substance it produces.

Kuhn's observations only claim to hold true for scientific paradigms. However, in spite of his caution, it seems that they can be generalized for all types of knowledge. It is only possible to create a binding knowledge order when all representatives of a community (apart from scientists, these may include all employees of a company, or

the users of a subject-specific Web service) speak a common language, which then makes up the basis of the knowledge order. If there is no common paradigm (e.g. in enterprises: the researchers have *their* paradigm, the marketing experts an alternative one), the construction of *one single* knowledge order from the perspective of exactly one group makes little sense. "Revolutionary changes" that lead to a paradigm shift require equally revolutionary changes in the knowledge order. It is a great challenge in information science "how to map semantic fields of concepts and their signifying contexts into our systems" (Brier, 2008, 424).

Information as Knowledge Put in Motion

Let us turn back to Figure A.2.1, in which we are discussing the process of signal transmission. If we want to put knowledge "into a form", or in motion, we cannot do so by disregarding this physical process. Information is thus fundamentally a unit made up of two components: the document as signal, and the content as knowledge. For the purposes of information science, we must enhance Shannon's schema by adding the knowledge component. The triad of sender—channel—receiver is preserved; added to it are, on the sender's side, the knowledge he is referring to and wants to put in motion, and on the receiver's side, the knowledge as he understands it. The problem: there is no direct contact between the knowledge that is meant by a sender and that which is understood by the receiver; the transmission always takes a detour via encoding, channel (including noise) and decoding. And this route, represented schematically in Figure A.2.4, is extremely prone to disruptions.

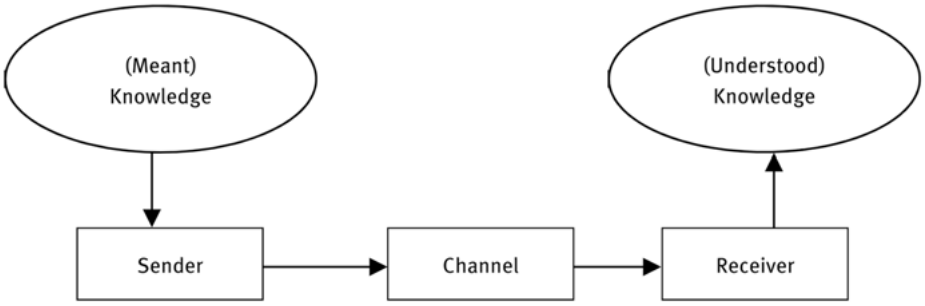


Figure A.2.4: Simple Information Transmission.

Apart from the physical problems of information transmission, several further bundles of problems must be successfully dealt with in order for the understood knowledge to at least approximately match the meant knowledge. First, we need to deal with the characters being used. Sender and receiver must dispose of the same character set, or employ a translator. (As a self-experiment, you could try speaking BAPHA out

loud. If this appears to yield no meaning at the first attempt, let us enlighten you that it includes Cyrillic characters.) Secondly, the language being used should be spoken fluently by both sender and receiver. A piece of "hot information", written in Bulgarian, will be useless to someone who does not speak the language. Thirdly, natural languages are by no means unambiguous. (Another experiment: what are you thinking of when you hear or see the word JAVA? There are at least three possible interpretations: programming language, island and coffee.) Fourthly, there must be a common "world horizon", i.e. a certain socio-cultural background. Imagine the problems of an Eskimo in explaining to an Arab from Bahrain the various forms of snow. When information transmission is successful, the knowledge (carried by the information) meets the pre-existing knowledge of the receiver and changes it.

According to Brookes (1980, 131), the connection between knowledge and transmitted information can be expressed via a (pseudo-mathematical) equation. Knowledge (K) is understood as a structure (S) of concepts and statements; the transmitted information (Δ I) carries a small excerpt from the world of knowledge. The equation goes:

$$K[S] + \Delta I = K[S + \Delta S].$$

The knowledge structure of the receiver is modified via Δ I. This effects a change to the structure itself, as signified by the Δ S. The special case of Δ S = 0 is not ruled out. The same Δ I, received by different receivers with different K[S] than each other, can effect different structural changes Δ S. The process of information differs "from person to person and from situation to situation" (Saab & Riss, 2011, 2245). So the "understanding of a person's knowledge structure" (Cool & Belkin, 2011, 11) is essential for information science.

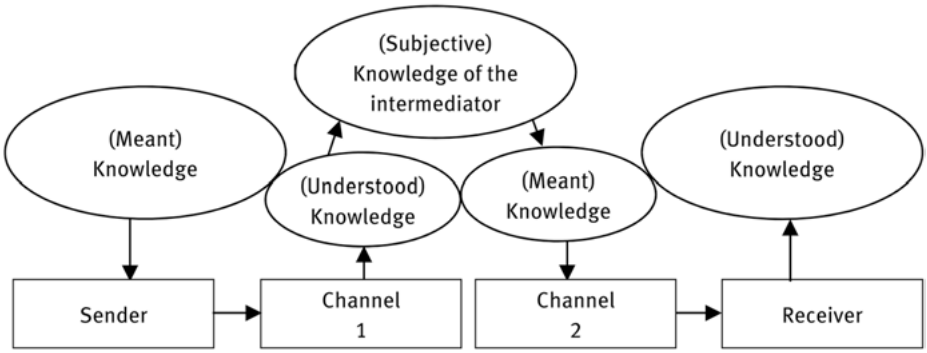


Figure A.2.5: Information Transmission with Human Intermediator.

The intermediation is set between instances of the chain of information transmission; this is where the specific informational added value provided by information science is being developed.

When interposing a human information intermediator, the blurry areas between the knowledge that is meant and that which is understood are widened due to the addition of the intermediator's subjective knowledge. Let us discuss the problems via an example: Over the course of this chapter, we will talk about Bacon's "Knowledge is Power". Bacon (Sender) has fixed his knowledge in several texts (Channel 1), from which we can hardly work out exhaustively what he "really" meant (meant knowledge). An information intermediator (in this case, we will assume this role) has read Bacon's works (Channel 1) and, ideally, understood them—on the basis of the existing knowledge base (this knowledge base also contains other sources, e.g. Schmidt's (1967) essay and further background knowledge). An opinion about Bacon's "Knowledge is Power" has been developed and fixed in this book (Channel 2). This is where you, the reader (Receiver), come into play. On the basis of your foreknowledge, you understand our interpretation of Bacon's "Knowledge is Power". How much of what Bacon actually meant would have reached you in its unadulterated form? To clarify: Information science attempts to minimize such blurry areas via suitable methods and tools of knowledge representation and information retrieval, and to provide for a (possibly) ideal information flow.

From Kuhlen we learned that the objective in information science is not only to put knowledge in motion, but also to refine it with informational added value. We now incorporate into our schema of information transmission a human intermediator with his or her own subjective knowledge (Figure A.2.5) on the one hand, and an automatic mechanical intermediator (Figure A.2.6) on the other.

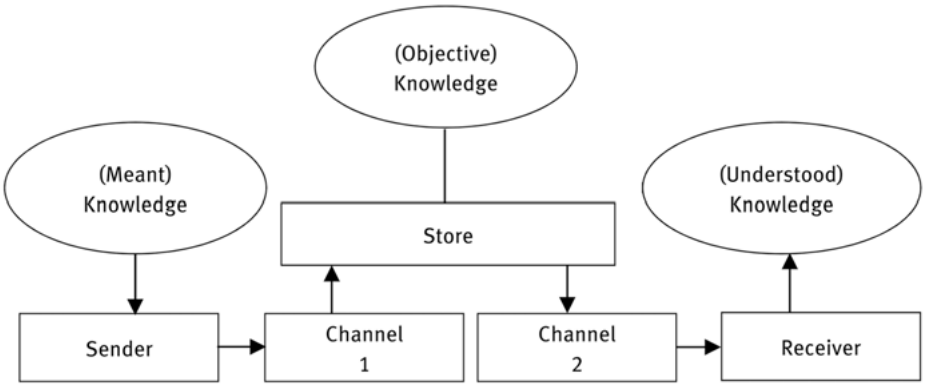


Figure A.2.6: Information Transmission with Mechanical Intermediation.

The situation is slightly different in the case of mechanical intermediation. Instead of a human, a machine is set between sender and receiver; say, a search engine on

the Internet. The problems posed by the human intermediator's (meant and understood) knowledge no longer exist: the machine stores—in Popper's sense—objective knowledge. The questions are now: How does it store this knowledge? And: How does it then yield it (i.e. in what order)? If the receiver wants to search comfortably and successfully, he must (at least to some extent) grasp the techniques used by search engines. Information science provides such techniques of Information Indexing and also makes sure users find it as easy as possible to search and retrieve knowledge. Is it possible to replace a subjective bearer of knowledge—say, an employee with valuable know-how or know-that—with an objective one in such a way that the knowledge is still accessible after the employee has left the company? The answer is: yes, in principle, only such a current employee must dispose of a knowledge structure that allows him to adequately understand the stored knowledge.

What are the characteristics of information? Does information have anything to do with truth? Is information always new? "Knowledge is power", Bacon says. Are there any relations between knowledge and power?

Information and Truth

Knowledge has a truth claim. Is this also the case for information, if information is what sets this knowledge in motion? Apart from knowledge, there are further, related forms of dealing with objects. Thus we speak of belief when we subjectively think something is true that cannot be objectively explained, of conjecture when we introduce something as new without (subjective or objective) proof, or of lies when something is clearly not true but is communicated regardless. If beliefs, conjectures or lies are put in motion, are they not information? For predominantly practical reasons, it would prove very difficult, if not impossible, for information science to check every piece of informational content for its truth value. "Information is not responsible for truth value," Kuhlen (1995, 41) points out (for a contrary opinion cf. Budd, 2011). Buckland (1991a, 50) remarks, "we are unable to say confidently of anything that it could not be information;" and Latham (2012, 51) adds, "even untrue, incorrect or unseen information is information." The task of checking the truth value of the knowledge, rather, must be delegated to the user (Receiver). He then decides whether the information retrieved represents knowledge, conjecture or untruth. At the same time, though, it must of course be required of the information's user to be competent enough to perform this task without any problems, i.e. he is well trained in information literacy.

Information and Novelty

Does information have anything to do with novelty? Must knowledge always be new to a receiver in order to become information? And conversely, is every new knowledge always information? Let us start off with the last question and imagine we are at the airport, planning to board a plane to Graz, Austria. There is a message over the PA system: "Last call for boarding Flight 123 to Rome!" This is new to us; we find it to be of no interest at all, though, and will probably hardly even notice it before immediately forgetting we heard it at all. Now another announcement: "Flight 456 to Graz will board half an hour late due to foggy weather." This is also new, and it does interest us; we will modify our actions (e.g. by going for a bite to eat before boarding) accordingly. This time, the announcement is useful information to us.

We now come to the first question. Staying with the airport example (and sitting in the airport restaurant by now), the repeated announcement "Flight 456 to Graz will board half an hour late due to foggy weather" is not new to us, but it does affect our actions, as it allows us to continue eating in peace (after all, the fog could have lifted unexpectedly in the meantime). By now it should have become clear that novelty does not have to be an absolute requirement for information; it is still a relative requirement, though. By the fourth or fifth identical announcement, however, the knowledge being communicated loses a lot of its value for us. A certain quantum of repetition is acceptable—it is sometimes even desirable, as a confirmation of what is already known—but too much redundancy reduces the information's relevance to our actions. In this context, von Weizsäcker (1974) identifies the vertices of first occurrence and repetition as the components of the pragmatic aspect of information: the "correct" degree of informedness lies between the two extremes.

"Knowledge is Power"

What does information yield its receiver? The knowledge that is transmitted via information meets the user's pre-existing knowledge and leads to the user having more know-how or know-that than previously. Such knowledge enhancement is not in service of purely intellectual pursuits (at least not always); rather, it serves to better understand or be able to do something in order to gain advantages vis-à-vis others—e.g. one's competition. Knowledge and having knowledge is thus related to power. The classical dictum "Knowledge is Power" has been ascribed to Bacon. In his "Novum Organum" (Bacon, 2000 [1620], 33), we read:

> Human knowledge and human power come to the same thing, because ignorance of cause frustrates effect. For Nature is conquered only by obedience; and that which in thought is a cause, is like a rule in practice.

A further reference in the "New Organon" shows the significance of science and technology ("active tendency") to power (Bacon, 2000 [1620], 103):

> Although the road to human knowledge and the road to human power are very close and almost the same, yet because of the destructive and inveterate habit of losing oneself in abstraction, it is altogether safer to raise the sciences from the beginning on foundations which have an active tendency, and let the active tendency itself mark and set bounds to the contemplative part.

Bacon's context is clear: he is talking about scientific knowledge, and about the power that allows us to reign over nature with the help of this (technological) knowledge. The aspect of knowledge as power over other people cannot be explicitly found in Bacon, but it is—according to Schmidt (1967, 484-485)—an implication hidden in the text.

> The power that Bacon speaks of is the power of man to command nature; we must, however, pay attention to the dialectical nuance that man is not outside of nature, but a part of it. Every victory of man over nature is, by implication, a victory of man over himself. Power over other people can only be gained because every man is subject to natural conditions.

Uncertainty

What is the power that can be obtained through knowledge, and which is what we talk about in the context of information science? Ideally, a receiver will put the transmitted information to use. On the basis of the experienced "discrepancy between the information regarded as necessary and that which is actually present" (Wittmann, 1959, 28), uncertainty arises, and must be reduced. The information obtained—thus Wittmann's (1959, 14) famous definition—is "purpose-oriented knowledge". For Belkin (1980) such "anomalies" lead to "anomalous states of knowledge" (ASK), which are the bases of a person's information need. Belkin, Oddy and Brooks (1982, 62) note:

> The ASK hypothesis is that an information need arises from a recognized anomaly in the user's state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly.

Wersig (1974, 70) also assumes a "problematic situation" that leads to "uncertainty". Wersig (1974, 74) even defines the concept of information with reference to uncertainty:

> Information (...) is the reduction of uncertainty due to processes of communication.

Kuhlthau (2004, 92) introduces the Uncertainty Principle:

> Uncertainty is a cognitive state that commonly causes affective symptoms of anxiety and lack of confidence. Uncertainty and anxiety can be expected in the early stages of the information search process. The affective symptoms of uncertainty, confusion, and frustration are associated with vague unclear thoughts about a topic or question. As knowledge states shift to more clearly focused thoughts, a parallel shift occurs in feelings of increased confidence. Uncertainty due to a lack of understanding, a gap in meaning, or a limited construction initiates the process of information seeking.

What does "purpose-oriented knowledge" mean? What is "correct" information, information that fulfills its purpose and reduces uncertainty? Information helps us master "problematic situations". Such situations include preparations for decision-making, knowledge gaps and early-warning scenarios.

In the first two aspects, the information deficit is known: decisions must be prepared with reference to the most comprehensive information available, knowledge gaps must be filled. Both situations are closely linked and, for the most part, coincide. In the third case, it is the information that alerts us to the critical situation in the first place. The early-warning system signals dangers (e.g. posed by the competition, such as the market entry of new competitors). The "correct" internal knowledge (including the "correct" bearers of knowledge, i.e. the employees), combined with the "correct" internally used and externally acquired knowledge shows up business opportunities for the company and at the same time warns of risks. In any case, it leads to conscious action (or forbearance). This is the practical goal of all users' information activities: to translate knowledge into action via information. For a specific information user in a specific situation (Henrichs, 2004), information is action-relevant knowledge—beyond true or false, new or confirmed. Similarly, Kuhlen (2004, 15) states:

> Corresponding to the pragmatic interpretation, information is that quantity of knowledge which is required in current action situations, but which the current actor generally does not personally dispose of or at the very least does not have a direct line to.

Once an actor has assembled the quantity of knowledge that he needs in order to perform his action, he is—in Kuhlen's sense—"knowledge-autonomous". Under the term "information autonomy", Kuhlen summarizes the activities of gathering relevant knowledge (besides mastery of the relevant retrieval techniques) as well as assessing the truth and innovation values of the knowledge retrieved.

To put it in the sense of Bacon's adage, we are granted the ability to act appropriately on the basis of knowledge if and only if we have informational autonomy. The totality of a person's skills for acting information-autonomous (e.g. basic IT and smartphone skills, skills for mastering information retrieval and skills for controlling the uploading and indexing of documents) is called information literacy (see Ch. A.5).

The power of information literacy is constrained. Even information that is ideally "correct" is generally not enough to reduce any and all uncertainty during the decision-making process. The information basis for making decisions can only be smaller or greater. However, increasing the size of the information basis brings enormous competitive advantages in economic life, compared to the competition. Imperfect information and the respective relative uncertainty are key aspects of information practice: in an environment that as a matter of principle cannot be comprehensively recorded, and which is in constant flux to boot, it is important for individuals, enterprises, and even for entire regions and countries, to reduce the uncertainties resulting from this state of affairs as far as possible. However, complete information and thus the complete reduction of uncertainty can never be attained. Rather, information creates a state of creative uncertainty. The perception and assessment of uncertainty lead to information being used as a strategic weapon. By disposing of and refining knowledge, companies and individuals can stand to gain advantages over their competitors. The competitive advantage is given for as along as any party monopolizes certain information, or at least for as long as the information is asymmetrically distributed in favor of our person, our company or our country. The other competitors must necessarily follow suit, until there is once more an information equilibrium in a certain field. Then, information in other fields or new information in the original field is used to work out a new information advantage etc. Information thus takes center stage in the arena of innovative competition.

Production of Informational Added Value

Information activities process knowledge, present it in a user-friendly manner, represent it via condensation and the allocation of information filters, and prepare it for easy and comprehensive search and retrieval. All aspects that go beyond the original knowledge are informational added values, which can be found in private as well as public information goods. In addition to *knowing how* and *knowing that* the informational value added is *knowing about*. Information activities lead to knowledge *about* documents and *about* the knowledge fixed in the documents. For Buckland (2012), *knowing about* is more concerned with information science than *knowing how* and *knowing that*. Buckland (2012, 5) writes,

> in everyday life we depend heavily and more and more on second-hand knowledge. We can determine little of what we need to know by ourselves, at first hand, from direct experience. We have to depend on others, largely through documents. … In this flood of information, we have to select and have to decide what to trust. What we believe about a document influences our use of it, and more importantly, our use of documents influences what we believe.

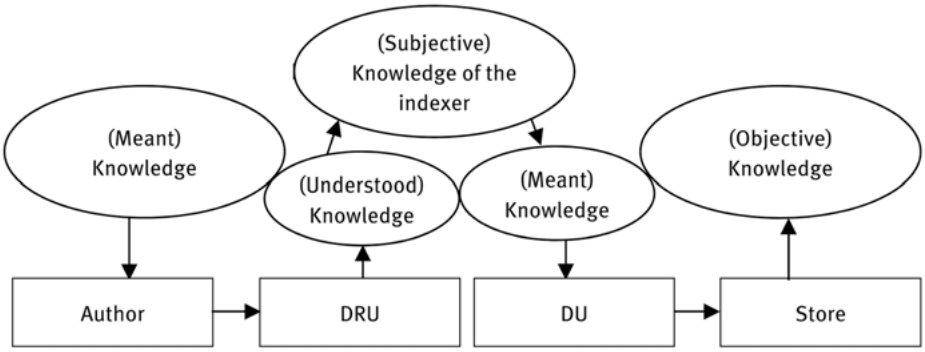


Figure A.2.7: Intermediation—Phase I: Information Indexing (here: with Human Indexer). DRU: Documentary Reference Unit, DU: Documentary Unit.

In an initial, rough approximation of theoretical information science, we distinguish between three phases of information transmission (Figures A.2.7 through A.2.9). In Phase 1, the objective is to represent knowledge fixed by an author in a document (e.g. a book, a research article, a patent document or a company-internal memo). This document is called a "documentary reference unit" (DRU). Then comes an information specialist or—in the case of automatic indexing—an information system, reads this unit and represents its content (via a short summary) and its main topics (via some keywords) in a so-called "documentary unit" (DU), also called "surrogate". In the ideal scenario, this DU (as an initial informational added value) is saved in an information store together with the DRU. Warner (2010, 10) calls the activities of knowledge representation "description labor" as the labor involved "in transforming objects into searchable descriptions". This process, sketched just above, proves to be far more complex than assumed and is what forms the central object of the second part of this book.

The DU may be sent to interested users by the respective information system itself, as a push service; until then, it remains in the knowledge store awaiting its users and usage. One day, a user will feel an information need and start searching, or let an information expert do his search for him. Via its pull service, the information system allows users to systematically search through the documentary units (with the informational added value contained within them) and also to simply browse through the system. The language of the searcher is then aligned with the languages of the DUs and DRUs, and the retrieval system arranges the list of DUs according to relevance. The searcher will select some relevant DUs and through them acquire the DRUs, i.e. the original documents, which he will then transmit to his client (of course client and searcher can be the same person). For Warner (2010, 10) the processes of information retrieval are connected with "search labor", which is understood as "the human labor expended in searching systems". Information retrieval is where the second amount

of informational added values is worked out. The entire first part of this book is dedicated to this subject.

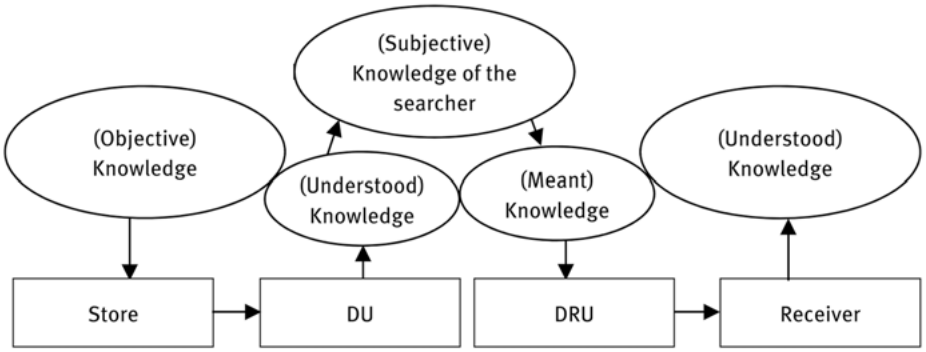


Figure A.2.8: Intermediation—Phase II: Information Retrieval.

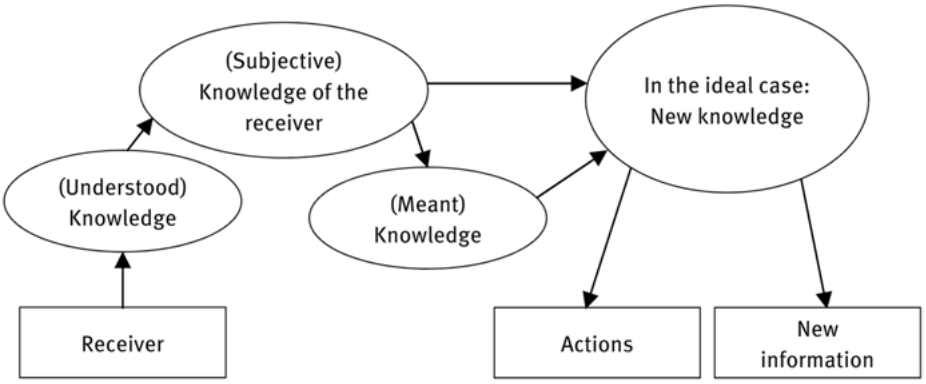


Figure A.2.9: Intermediation—Phase III: Further Processing of Retrieved Information.

Figure A.2.9 seamlessly picks up where Figure A.2.8 left off. The receiver has received the relevant documentary reference units containing his action-relevant knowledge. The transmitted information meets the receiver's state of knowledge and, in the ideal case, merges with it to create new knowledge. The new knowledge leads to the desired actions and—if the actor reports on it—new information.

Conclusion

- "Information" must be conceptually differentiated from the related terms "signal", "data" and "knowledge". The basic model of communication—developed by Shannon—takes into consideration sender, channel and receiver, where the channel is used to transmit physical signals. Sender and receiver respectively encode and decode the signals into signs.
- General semiotics distinguishes between relations of signs among each other (syntax), the meaning of signs (semantics) and the usage of signs (pragmatics). When we neglect semantics and pragmatics and concentrate on the syntax of transmitted signs, we are speaking of "data"; when we regard transmitted signs from a semantic and pragmatic perspective, we are looking at "knowledge" and "information", respectively.
- Knowledge is subjective, and thus forms part of Popper's World 2, when a user disposes of it (know-that and know-how). Knowledge is objective and thus an aspect of World 3 in so far as the content is stored user-independently (in books or digital databases).
- Knowledge is thus fixed either in a human consciousness (as subjective knowledge) or another store (as objective knowledge). Knowledge as such cannot be transmitted. To transmit it, a physical form is required, i.e. an inFORMation. Information is knowledge put in motion, where the concept of information unites the two aspects of signal and knowledge in itself.
- Knowledge is present not only in texts, but also in non-textual documents (images, videos, music). Panofsky distinguishes between three semantic levels of interpretation: the pre-iconographical, the iconographical and the iconological level.
- Knowledge—as one of the core concepts of knowledge representation—allows for different perspectives. Knowledge can be regarded as both (true, known) statements (know-that) and as the knowledge of how to do certain things (know-how).
- According to Polanyi, implicit knowledge is pretty much physically embedded in a person, and hence can be objectified (externalized) only with difficulty.
- Knowledge management means the way knowledge is dealt with in organizations. According to Nonaka and Takeuchi, the knowledge spiral consisting of socialization, externalization, combination, and internalization (SECI) leads to success, even if externalization (transforming implicit knowledge into explicit knowledge) can hardly be achieved comprehensively. Probst et al. use several building blocks of corporate knowledge management.
- According to Spinner's knowledge theory, we are confronted by a multitude of different types of knowledge, amounts of knowledge and qualities of knowledge, which all require their own methods of knowledge representation.
- Only knowledge in normal sciences (according to Kuhn) can be put into a knowledge organization system. In normal sciences, there are discipline-specific special languages (e.g., the terminologies of chemists or of biologists).
- Information does not (in contrast to knowledge) have a truth claim, and neither does it claim to be absolutely new; ideal informedness is achieved between the extreme poles of novelty and repeated confirmation.
- The meant knowledge of a sender does not have to match the understood knowledge of a receiver. If we interpose further instances between sender and receiver (besides signal transmission via the channel), i.e. human and mechanical intermediators, the problems of adequately communicating knowledge via information will become worse.
- Information science allows knowledge to be put in motion, but it also facilitates the development of specific added values (knowledge about), which, at the very least, reduce problems during the transferral of knowledge.
- "Knowledge is Power" refers, in Bacon, to man's power over nature. Thinking further, it can also be taken to refer to power over other people. The negative consequences of "Knowledge is Power"

can be counteracted by granting everybody the opportunity of requesting knowledge. The goal is the subjects' information autonomy, guaranteed by their information literacy.
- In problematic situations, when there is uncertainty, purpose-oriented knowledge will be researched in order to consolidate decisions, close knowledge gaps and detect early-warning signals. The goal of the information-practical value chain is to translate knowledge into concrete actions via information. Information, however, is never "perfect" and cannot reduce uncertainties to zero. What it can do is to create relative advantages over others.
- Humans as well as institutions require information autonomy in order to be able to research and apply the most action-relevant knowledge possible.
- Both information science and information practice work out informational added values (knowledge about). This is accomplished over the course of professional intermediation, both in the input area of storage (knowledge representation) and during the output process (information retrieval).

Bibliography

Bacon, F. (2000 [1620]). The New Organon. Cambridge: Cambridge University Press. (Latin original: 1620).

Bates, M.J. (2005). Information and knowledge. An evolutionary framework for information science. Information Research, 10(4), paper 239.

Bates, M.J. (2006). Fundamental forms of information. Journal of the American Society for Information Science and Technology, 57(8), 1033-1045.

Belkin, N.J. (1978). Information concepts for information science. Journal of Documentation, 34(1), 55-85.

Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133-143.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval. Journal of Documentation, 38(2), 61-71 (part 1), 38(3), 145-164 (part 2).

Boisot, M., & Canals, A. (2004). Data, information and knowledge. Have we got it right? Journal of Evolutionary Economics, 14(1), 43-67.

Brier, S. (2008). Cybersemiotics. Why Information is not enough. Toronto, ON: Univ. of Toronto Press.

Brookes, B.C. (1980). The foundations of information science. Part I. Philosophical aspects. Journal of Information Science, 2(3-4), 125-133.

Buckland, M.K. (1991a). Information and Information Systems. New York, NY: Praeger.

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

Buckland, M.K. (2012). What kind of science *can* information science be? Journal of the American Society for Information Science and Technology, 63(1), 1-7.

Budd, J.M. (2011). Meaning, truth, and information. Prolegomena to a theory. Journal of Documentation, 67(1), 56-74.

Capurro, R. (1978). Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs. München: Saur.

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

Chisholm, R.M. (1977). Theory of Knowledge. Englewook Cliffs, NJ: Prentice-Hall.

Cool, C., & Belkin, N.J. (2011). Interactive information retrieval. History and background. In I. Ruthven & D. Kelly (Eds.), Interactive Information Seeking, Behaviour and Retrieval (pp. 1-14). London: Facet.

Floridi, L. (2005). Is information meaningful data? Philosophy and Phenomenological Research, 70(2), 351-370.

Hardy, J., & Meier-Oeser, S. (2004). Wissen. In Historisches Wörterbuch der Philosophie. Vol. 12 (pp. 855-856). Darmstadt: Wissenschaftliche Buchgesellschaft; Basel: Schwabe.

Henrichs, N. (2004). Was heißt „handlungsrelevantes Wissen"? In R. Hammwöhner, M. Rittberger, & W. Semar (Eds.), Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen (pp. 95-107). Konstanz: UVK.

Jones, W. (2010). No knowledge but through information. First Monday, 15(9).

Kuhlen, R. (1995). Informationsmarkt. Chancen und Risiken der Kommerzialisierung von Wissen. Konstanz: UVK. (Schriften zur Informationswissenschaft; 15).

Kuhlen, R. (2004). Information. In R. Kuhlen, T. Seeger, & D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation (pp. 3-20). 5th Ed. München: Saur.

Kuhlthau, C.C. (2004). Seeking Meaning. A Process Approach to Library and Information Services. 2nd Ed. Westport, CT: Libraries Unlimited.

Kuhn, T.S. (1962). The Structure of Scientific Revolutions. Chicago, IL: University of Chicago Press.

Latham, K.F. (2012). Museum object as document. Using Buckland's information concepts to understanding museum experiences. Journal of Documentation, 68(1), 45-71.

Lenski, W. (2010). Information. A conceptual investigation. Information, 1(2), 74-118.

Ma, L. (2012). Meanings of information. The assumptions and research consequences of three foundational LIS theories. Journal of the American Society for Information Science and Technology, 63(4), 716-723.

Machlup, F. (1962). The Production and Distribution of Knowledge in the United States. Princeton, NJ: Princeton University Press.

Memmi, D. (2004). Towards tacit information retrieval. In RIAO 2004 Conference Proceedings (pp. 874-884). Paris: Le Centre de Hautes Études Internationales d'Informatique Documentaire / C.I.D.

Mooers, C.N. (1952). Information retrieval viewed as temporal signalling. In Proceedings of the International Congress of Mathematicians. Cambridge, Mass., August 30 – September 6, 1950. Vol. 1 (pp. 572-573). Providence, RI: American Mathematical Society.

Nonaka, I., & Takeuchi, H. (1995). The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation. Oxford: Oxford University Press.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, *Ba* and Leadership. A unified model of dynamic knowledge creation. Long Range Planning, 33(1), 5-34.

Panofsky, E. (1975). Sinn und Deutung in der bildenden Kunst. Köln: DuMont.

Panofsky, E. (2006). Ikonographie und Ikonologie. Köln: DuMont.

Polanyi, M. (1958). Personal Knowledge. Towards a Post-Critical Philosophy. London: Routledge & Kegan Paul.

Polanyi, M. (1967). The Tacit Dimension. Garden City, NY: Doubleday (Anchors Books).

Popper, K.R. (1972). Objective Knowledge. An Evolutionary Approach. Oxford: Clarendon.

Probst, G.J.B., Raub, S., & Romhardt, K. (2000). Managing Knowledge. Buildings Blocks for Success. Chichester: Wiley.

Rauch, W. (2004). Die Dynamisierung des Informationsbegriffs. In R. Hammwöhner, M. Rittberger, & W. Semar (Eds.), Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen (pp. 109-117). Konstanz: UVK.

Reeves, B.N., & Shipman, F. (1996). Tacit knowledge: Icebergs in collaborative design. SIGOIS Bulletin, 17(3), 24-33.

Ryle, G. (1946). Knowing how and knowing that. Proceedings of the Aristotelian Society, 46, 1-16.
Saab, D.J., & Riss, U.V. (2011). Information as ontologization. Journal of the American Society for Information Science and Technology, 62(11), 2236-2246.
Schmidt, G. (1967). Ist Wissen Macht? Kantstudien, 58, 481-498.
Shannon, C. (2001 [1948]). A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review, 5(1), 3-55. (Original: 1948).
Spinner, H.F. (1994). Die Wissensordnung. Ein Leitkonzept für die dritte Grundordnung des Informationszeitalters. Opladen: Leske + Budrich.
Spinner, H.F. (2000). Ordnungen des Wissens: Wissensorganisation, Wissensrepräsentation, Wissensordnung. In Proceedings der 6. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation (pp. 3-23). Würzburg: Ergon.
Spinner, H.F. (2002). Das modulare Wissenskonzept des Karlsruher Ansatzes der integrierten Wissensforschung – Zur Grundlegung der allgemeinen Wissenstheorie für 'Wissen aller Arten, in jeder Menge und Güte'. In K. Weber, M. Nagenborg, & H.F. Spinner (Eds.), Wissensarten, Wissensordnungen, Wissensregime (pp. 13-46). Opladen: Leske + Budrich.
Warner, J. (2010). Human Information Retrieval. Cambridge, MA, London: MIT Press.
Weizsäcker, E.v. (1974). Erstmaligkeit und Bestätigung als Komponenten der Pragmatischen Information. In E. v. Weizsäcker, Offene Systeme I (pp. 82-113). Stuttgart: Klett-Cotta.
Wersig, G. (1974). Information – Kommunikation – Dokumentation. Darmstadt: Wissenschaftliche Buchgesellschaft.
Wittmann, W. (1959). Unternehmung und unvollkommene Information. Köln, Opladen: Westdeutscher Verlag.
Wyner, A.D. (2001). The significance of Shannon's work. ACM SIGMOBILE Mobile Computing and Communications Review, 5(1), 56-57.

A.3 Information and Understanding

Information Hermeneutics

> Hermeneutics is the art of hemeneuein, i.e. of proclaiming, interpreting, explaining and expounding. <Hermes> is the name of the Greek messenger of the Gods, who passed the messages of Zeus and his fellow immortals on to mankind. His proclamations are obviously more than mere communication. In explaining the divine commands, he translates them into the language of the mortals, making them understandable. The virtue of H(ermeneutics) always fundamentally consists in transposing a complex of meaning from another "world" into one's own. This also holds for the original meaning of hermeneia, which is "the statement of thought". This concept of statement itself is ambivalent, comprising the facets of utterance, explanation, interpretation and translation (Gadamer, 1974, 1061-1062).

Without Hermes, there would have been no understanding between the two worlds of gods and men. Hermes is the mediator. Hermeneutics is the science of (correct) understanding. An important source on hermeneutics, for us, is Gadamer's "Truth and Method" (1975), which is significantly influenced by the philosophy of Heidegger (1962 [1927]).

What does information science activity have in common with that of Hermes? Looking at the sheer unstructured amount of documents on the one hand, and contrasting this opaque mass with the specific information need of an individual, it is impossible to reach a satisfactory understanding without a mediator of some kind. Information science helps create a "bridge" between documents and information needs. The complex of meaning—hermeneutically speaking—is transposed from one "world" to the other; figuratively speaking, from a text to a query. The problems and perspectives of hermeneutics, cognitive sciences and other disciplines on human cognition are manifold. In the context of this book, we will only emphasize some fundamental points of discussion that concern the area of information.

If we neglect the world outside of information, the social periphery, we will run the danger of remaining in permanent tunnel vision (Brown & Duguid, 2000, 5). A linear, purely goal-oriented path of information science can easily lead into a dead end, since a stable knowledge organization system and stable bibliographical records, respectively, will one day cease to correspond to the requirements of the day. A holistic point of view, on the other hand, opens up one's field of vision.

Information retrieval concentrates on the searching for and finding of information. In knowledge representation, the evaluation and presentation of information is at the foreground. Both areas of information science only serve as means to a particular end: their focus is the potential user who is supposed to find the processed information, and the knowledge behind it, useful. Even the "best" indexing of content is of no use to the user if he cannot himself interpret or understand and explain it. Information science, with all its methods and applications, interacts dynamically with its

agents, the users and authors. We must not forget, here, that all the technologies, methods, tools and tasks of information science being used are founded, once more, by human activity.

Interaction can only come about when one recognizes what the other means and wishes, and when one declares oneself ready to accommodate him or her. This is accomplished via language. Linguistic ability alone is not enough to make one understandable to another, however. What has influenced and continues to influence every single human being has been learned via interaction with his environment, culture, and past. Hence, language is not neutral but is always used from a certain socio-cultural viewpoint. Gadamer (1975) calls this the "horizon". Language is not formed and used without understanding and pre-judgments. Otherwise, people would talk past one another. A stream of communication is thus based on a commonly understood language used within a culture. However, the language changes over time, due to individual usage, and man likewise changes via/alongside his use of language.

This circumstance represents a particular challenge for the analysis and usage of a language in time, space and interaction. Since information science deals with the form and content of different languages, in various times and places as well as different areas of knowledge and cultural spheres, it must do justice to this diversity and the constant change underlying it.

Heraclitus already recognized: "Panta rhei—Everything is in a state of flux" (Simplicius, 1954, citing Heraclitus). Plato writes, "Heraclitus, I think, says that everything moves and nothing rests" (Plato, 1998, Cratylus 402a). Except for natural laws and mathematical formulas, perhaps, everything in the world is in constant change. Language is a part of this world. Information (and its interpretations) streams like a river, like water endlessly spreading itself into countless directions, coming apart into separate streams before flowing back together. Even if the stream of information appears to run the same course for a long time, we will never see the same exact stream containing exactly the same information. So every knowledge organization system and every subject description in bibliographical records cannot be stable over the course of time, but are objects of possible change (Gust von Loh, Stock, & Stock, 2009). For Buckland (2012, 160)

> (i)t is a good example of a problem that is endemic in indexes and categorization systems: Linguistic expressions are necessarily culturally grounded, and, for that reason, in conflict with the need to have stable, unambiguous marks to enable library systems to perform efficiently. A static, effective subject indexing vocabulary is a contradiction in terms.

Water constantly seeks a way to keep flowing. It never flows uphill, as this would contravene nature. Rather, it is in interaction with nature. It follows certain requirements. What are the hermeneutic and cognitive bases, respectively, that resonate throughout the entire information process, and which we do not think about explicitly and tend to ignore? They are fundaments of complex build, which form the activities of both

information professionals and users. We are dealing with a holistic perspective on information science, where man is incorporated into the information process with all of his characteristics and activities.

Four large complexes of human characteristics combine in a feedback loop (Figure A.3.1):

- Horizons of the author / information professional / user,
- Understanding of the documents (interpretation),
- Readiness to cooperate (commitment),
- Practice and innovation.

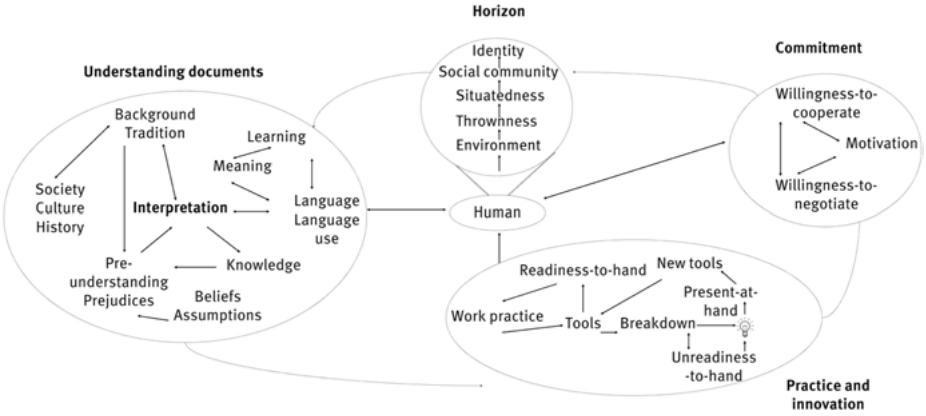


Figure A.3.1: Feedback Loop of Understanding Information.

Understanding

Man is born into the world, into an environment, without getting a say in the matter or the option of doing anything against it, a condition which Heidegger (1962 [1927]) calls "Thrownness". There is no escaping from this environment. Man is not alone—this he would not survive—but he comes in contact with others. Accordingly, he must follow his own instincts while always being tied to the situation he finds himself in. His agency is a goal-oriented one, the goal being self-preservation. He learns to (continue to) exist by becoming a member of a social community. Embedded into this community, he develops his identity via his actions. The individual thus represents his standpoint from his own perspective.

Understanding, like the creation of knowledge, is never done independently of the respective context. In Japanese philosophy, this "shared context" is called "ba". Ba is that place where information is understood as knowledge. Nonaka, Toyama and Konno (2000, 14) write:

> *Ba* is a time-space nexus, or as Heidegger expressed it, a locationality that simultaneously includes space and time. It is a concept that unifies physical space such as an office space, virtual space such as e-mail, and mental space such as shared ideals.

The thoughts and feelings of individuals never exist outside of reality. For Heidegger, subject and existence (being-in-the-world) form a fundamental unit, which is ontologically grounded and cannot be further called into question. The world as such is explored by the individual via his understanding, which is present in human beings from the beginning. There is a multitude of possibilities in the world, and man carries the key—to put it metaphorically—that allows him to picture them. This key is understanding, or—seen from the cognitive standpoint—human cognition. Understanding, according to Heidegger, is what makes humans human. Accordingly, every activity in information science is irrevocably built on this fundamental understanding. Understanding is amenable to development, which relates it to interpretation. Interpretation means the act of processing the possibilities that have been developed through understanding. When interpreting something as something, that which is already understood is joined by something new and the whole is brought into a relation. Since the pre-opinion of the interpreter always forms a part of the text's interpretation, this process is not merely one of unconditionally registering the given data. In analogy to the way the interpreter approaches the text via his horizon, the text itself is founded by the horizon of the author. Pre-judgments and prejudices, without which no understanding would come about at all, are the respective bases of any interpretation. Interpretation is what grants a text its meaning.

History, culture and society form human background and tradition in the past and in the present. The *Dasein* of man is closely related to "man's historicity" (Day, 2010, 175). Knowledge—as the result of interpretation—depends upon earlier experience and on man's situatedness in a tradition. The constant learning process plays an important role in the usage and development of a language. In order for interpretation not to be influenced by arbitrary ideas, man—shaped as he is by pre-opinion and language use—directs his focus onto the thing itself.

A text is not read without understanding and pre-judgment (Gadamer, 1975). The draft or the expectation of meaning one approaches the text with bears witness to a certain openness, in the sense that one is prepared to accept the text in the first place. Understanding means knowing something and only then to sound another's opinion. The movement of understanding is not static but circular: from the whole to the part and back to the whole. Gadamer emphasizes that the hermeneutic circle is neither subjective nor objective or formal in nature—instead, it is determined via the commonality that links man to tradition, and hence to each other via the use of language. The circle (Gadamer, 1975, 293)

> describes understanding as the interplay of the movement of tradition and the movement of the interpreter. The anticipation of meaning that governs our understanding of a text is not an act of

> of subjectivity, but proceeds from the commonality that binds us to the tradition. But this commonality is constantly being formed in our relation to tradition.

The objective is to regard the interval as a positive and productive possibility, and not as a reproductive characteristic. Understanding as a process of effective history means to take into account one's own historicality. When we deal with communications, we do so from a certain standpoint, our "horizon". Having a horizon means (Gadamer, 1975, 301-302):

> A person who has a horizon knows the relative significance of everything within this horizon, whether it is near or far, great or small. Similarly, working out the hermeneutical situation means acquiring the right horizon of inquiry for the questions evoked by the encounter with tradition.

We move dynamically in and with the horizon. The past is always involved in the horizon of the present. In understanding, finally, there is a "fusion of horizons"—an interpreter wants to understand the communication contained in the text, and to do so relates the text to hermeneutical experience. This communication is language in the purest meaning of the word, since it speaks for itself in the same way as an interlocutor (you) does. A *you*, however, is not an object but something that relates to another. Hence, one must admit the communication's claim to having something to say.

A forwarded text that becomes subject to interpretation asks a question of the interpreter. A text, with all the subjects addressed within it, can only be understood if one grasps the question answered by the text. This does not merely involve comprehending an outside opinion, but rather the necessity of setting the subjectivity of the text in relation to one's own thinking (fusion of horizons).

How does a user deal with a documentary unit that has been created with the help of one of the methods of knowledge representation (e.g. classification or thesaurus) and while using a specific tool (e.g. the International Patent Classification)? In the ideal scenario, he will try to understand and explain both the surrogate *and* the original document.

Understanding, as a form of cognition, means understanding something *as something*. Processing the documentary unit as well as reading the original document involves all hermeneutical aspects:

- the presence of a corresponding key,
- concentration on the forwarded text (i.e. the thing),
- here: reconstruction of the question answered by the text,
- understanding the parts necessitates understanding the whole, *and* understanding the whole necessitates understanding the parts (hermeneutic circle),
- text and reader are in different horizons, which are fused in the understanding,
- in this, the reader's own pre-understanding is to be estimated positively—as prejudgment—as long as it contributes dynamically (in the hermeneutic circle) to the fusion of horizons.

Hermeneutics and Information Architecture

Of central importance to our subject is the conjunction of the results obtained by Winograd and Flores (1986) with regard to the architecture and design of information systems and, simultaneously, to the use of methods of knowledge representation and information retrieval. The focus is not on some computer programs—what is being analyzed, rather, are the decisive fundamentals influencing the works and innovations achieved while using computers. Here, the basis is formed by man's social background and his language as a tool of communication; computers only aid real communication in the form of a medium, or a tool.

Winograd and Flores argue as follows: tradition and interpretation influence one another, since man—as a social animal—is never independent of his current and historically influenced environment. Any individual that understands the world interprets it. These interpretations are based on pre-judgment and pre-understanding, respectively, which always contain the assumptions drawn upon by the individual in question. We live with language. Language, finally, is learned via the activities of interpretation. Language itself is changed by virtue of being used by individuals, and individuals change by using it. Hence, man is determined via his cultural background and can never be free of pre-judgments. Concerning the inevitability of the hermeneutic circle according to Gadamer, we read (Winograd & Flores, 1986, 30):

> The meaning of an individual text is contextual, depending on the moment of interpretation and the horizon brought to it by the interpreter. But that horizon is itself the product of a history of interactions in language, interactions which themselves represent texts that had to be understood in the light of pre-understanding. What we understand is based on what we already know, and what we already know comes from being able to understand.

According to Heidegger, we deal with things in everyday life without thinking about the existence of the tool we are using. This tool is ready-to-hand (*Zuhandenheit* in Heidegger's original German text). We have been thrown into the world and are thus coerced into acting pragmatically. At the moment when the stream of action is interrupted by whatever circumstances, man begins to interpret. The tool has become useless. It is unready-to-hand (*Unzuhandenheit*). We search for solutions. Only now do the characteristics of the tool come to light, and the individual becomes aware of them. He realizes the tool's *Vorhandenheit*—it is now present-at-hand.

Such a "breakdown" with regard to language plays a fundamental role in all areas of life, including—Winograd and Flores maintain—in computer design and management and, related to our own problem area, in knowledge representation and information retrieval. A breakdown creates the framework about which something can be said, and language as a human act creates the space of our world (Winograd & Flores, 1986, 78):

> It is only when a breakdown occurs that we become aware of the fact that 'things' in our world exist not as a result of individual acts of cognition but through our active participation in a domain of discourse and mutual concern.

Man gives meaning to the world in which he lives together with other men. However, we must not understand language as a series of statements about objective reality, but rather as human acts that rest on conventions regarding the background that informs the respective meaning. Background knowledge and expectations guide one's interpretation. Information retrieval knows this problem all too well, as there exist, in many places, different words for one and the same concept (synonyms), or anaphora, i.e. expressions that point to other terms (e.g. pronouns to their nouns). Without interpretation, there can be no meaning. Interpretation, meaning and situation, respectively tradition, cannot be regarded in isolation here, either. According to Winograd and Flores, the phenomena of background and interpretation permeate our entire everyday life.

When understanding situations or sentences, man links them to suppositions or expectations similar to them. The basis of understanding as a process is memory. The origin of human language as an activity is not the ability to reflect on the world, but the ability to enter commitments. The process is an artistic development that goes far beyond a mere accumulation of facts.

Concerning the question whether computers also enter such connections and comprehend commands in this way, Winograd and Flores emphasize that computers and their programs are nothing other than actively structured media of communication that are "fed" by the programmer. Winograd and Flores (1986, 123) write:

> Of course there is a commitment, but it is that of the programmer, not the program. If I write something and mail it to you, you are not tempted to see the paper as exhibiting language behavior. It is a medium through which you and I interact. If I write a complex computer program that responds to things you type, the situation is still the same—the program is still a medium through which my commitments to you are conveyed. ... Nonetheless, it must be stressed that we are engaging in a particularly dangerous form of blindness if we see the computer—rather than the people who program it—as doing the understanding.

Hence, we should not equate the technologies of artificial intelligence with the understanding of human thought or human language. The design of computer-based systems only facilitates human tasks and interaction. The act of designing itself is ascribed an ontological position: since man generally uses tools as part of his "manness", in the same way the programmer uses his computer. The indexer uses the respective programs of the computer, the methods of indexing as well as the specific tools. By using tools, we ask and are told what it means to be human. The programmer designs the language of the world in which the user works; or, at least, he tries to create the world that regards or could regard the user. During this creative dynamic process, a fundamental role is played by breakdowns as conspicuous incidents,

imperfections and the like. They necessitate and sometimes cause a counteraction to fix the problem. True to the motto "mistakes are for learning", breakdowns help analyze human activity and create new options or changes.

Analysis of Cognitive Work

The main characteristic of the analysis of cognitive work is that it observes human action in the midst of, and in dependence of, its environment. CWA (Cognitive Work Analysis) was developed in the early 1980s by Danish researchers at the Risø National Laboratory in Roskilde (Rasmussen, Pejtersen, & Goodstein, 1994; see also: Vicente, 1999). Where human work requires decisions to be made, it is designated as cognitive work. Fidel (2006) grounds the necessity of empirical work analyses in information science on the following: since information systems are created for the reason of supporting people in their activities, the architecture and design of these systems should consequently be based on those activities as performed by them. In CWA, the actors are referred to in this context as carriers of activities. Factors that necessarily shape these activities are called "constraints". It often happens in the day-to-day that one and the same person performs different activities at different times in the public and private spheres, respectively, and correspondingly has different information needs (Fidel, 2006, 6):

> This diversity introduces the idea that each type of activity may require its own information system.

CWA, with the help of work analyses of certain user groups, strives to answer the following questions: what are the constraints formed by the different activities, and how do these constraints affect the activities?

Proponents view the requirement of CWA as based in various different areas. For instance, Mai (2006) relates the approach of contextual analysis to knowledge representation—more specifically, to the development of controlled vocabularies (including classifications, semantic networks and up to complex thesauri). According to Mai, the developers of classification systems can use analysis to gain an insight into which factors (can) influence the work of the actor (Mai, 2006, 18):

> The outcome is not a prescription of what actors should do (a normative approach) or a detailed description of what they actually do (a descriptive approach), but an analysis of constraints that shape the domain and context.

The structure of the analysis of cognitive work is displayed graphically in Figure A.3.2.

At the center of the dimensions is the actor with his respective education, technological experience, preference for certain information tools, knowledge about the

specific area of application and values he adheres to. Hence, six further dimensions that influence and shape the activities are analyzed.

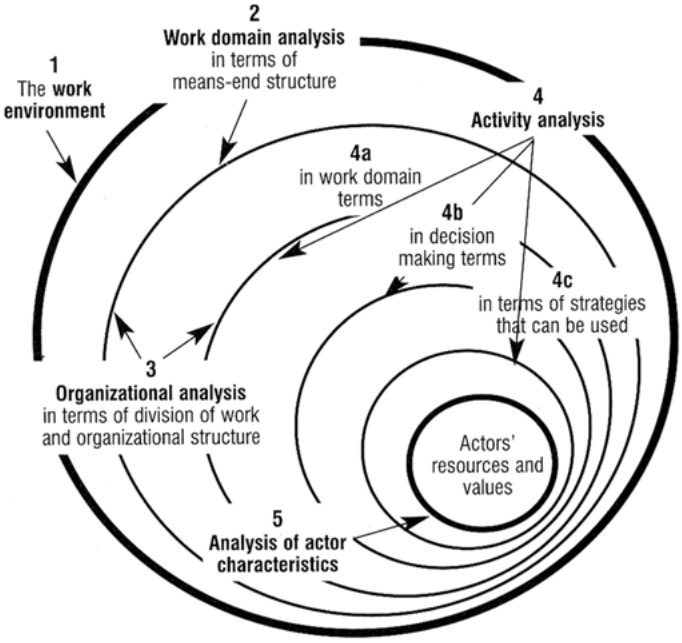


Figure A.3.2: Dimensions of the Analysis of Cognitive Work. Source: Pejtersen & Fidel, 1998.

Analyzing the work area generates the context in which the actor works. Actors may take part in the same work topic as others, while (due to their placement in a certain organization) setting different priorities and goals than others. Organizational analysis provides insight into the workplace, management or position of the actor in the organization. Activity analysis brings to light what actors do to fulfill their tasks. How, when and where does the actor require which type of information? Is the information present or not? Which information do decision makers require? Finally, the actor's strategies of publishing and searching information are observed. All of these empirical aspects of analysis, which serve to uncover the various interactions between people and information, first provide an insight into the actor's possible information need. After this, they are useful for the development of methods and tools of knowledge representation. The procedure following CWA is elaborate, but purposeful. It helps uncover the conditions for good knowledge representation and information retrieval to aid an actor's specific tasks (Mai, 2006, 19):

> Each dimension contributes to the designer's understanding of the domain, the work and activities in the domain and the actors' resources and values. The analyses ensure that designers bring the relevant attributes, factors and variables to design work. While analysis of each dimension

> does not directly result in design recommendations, these analyses rule out many design alternatives and offer a basis from which designers can create systems for particular domains. To complete the design, designers need expertise in the advantages and disadvantages of different types of indexing languages, the construction and evaluation of indexing languages and approaches to and methods of subject indexing.

Next up is evaluation, which, in analogy to the analysis of cognitive work, may provide insights into a potentially improved further processing (Pejtersen & Fidel, 1998). It is checked whether the implemented system corresponds to the previously defined goals and whether it serves the user's need. In the evaluation, too, the objective is not to focus on the system as a whole, but on the work-centric framework. The evaluation questions and requirements concern the individual dimensions, and a structure of analytical, respectively empirical, examination characteristics is compiled. Various facets and measurements are employed here, relative to the different dimensions of cognitive work. Pejtersen and Fidel demonstrate, via a study observing students and their working environment during Web searches, how the analysis and evaluation of cognitive work can uncover problem criteria for this user group.

Ingwersen and Järvelin (2005) create a model for information retrieval and knowledge representation that places the cognitive work of the actor at the foreground (Figure A.3.3). The actor (or a team of actors) is anchored in cultural, social and organizational horizons (context) (Relation 1). Actors are, for instance, authors, information architects, system designers, interface designers, creators of KOSs, indexers and users (seekers of information). Via a man-machine interface (2), the actor comes face to face with the information objects as well as with information technology (3). "Information objects" are the documentary units (surrogates) as well as the documents (where digitally available) that are accessible via methods and tools of knowledge representation and that have been indexed by form and content. "Information technology" summarizes the programming-related building blocks of an indexing and retrieval system (database, retrieval software, algorithms of Natural Language Processing) as well as retrieval models. Since the information objects can only be processed via information technology, there is another close link between these two aspects (4).

The actor always interacts with the system via the interface, but he requires additional knowledge about the characteristics of the information objects (Relation 5) as well as the information technology being used (7); additionally, information objects and information technology do not work independently of the social, cultural and organizational backgrounds (Relations 6 and 8). As opposed to the interactive relations 1 through 4, Relations 5 through 8 feature cognitive influences "in the background", which—as long as an information system is "running"—must always be taken into consideration, but cannot be interactively influenced.

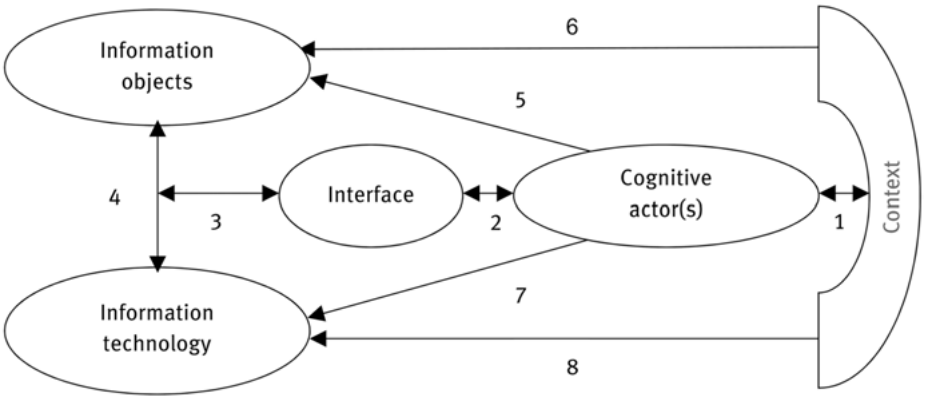


Figure A.3.3: The "Cognitive Actor" in Knowledge Representation and Information Retrieval. Source: Ingwersen & Järvelin, 2005, 261. Arrows: Cognitive Transformation and Influence; Double Arrows: Interactive Communication of Cognitive Structures.

Ingwersen and Järvelin (2005, 261-262) describe their model:

> First, processes of social interaction (1) are found between the actor(s) and their past and present socio-cultural or organizational context. ... Secondly, information interaction also takes place between the cognitive actor(s) and the cognitive manifestations embedded in the IT and the existing information objects via interfaces (2/3). The latter two components interact vertically (4) and constitute the core of an information system. This interaction only takes place at the lowest linguistic sign level. ... Third, cognitive and emotional transformations and generation of potential information may occur as required by the individual actor (5/7) as well as from the social, cultural or organizational context towards the IT and information object components (6/8) *over time*. This implies a steady influence on the information behaviour of other actors—and hence on the cognitive-emotional structures representing them.

Conclusion

- Knowledge representation and information retrieval are influenced by hermeneutical aspects (understanding, background knowledge, interpretation, tradition, language, hermeneutic circle and horizon) and cognitive processes.
- Every knowledge organization system and every subject description of documents cannot be stable over the course of time, but are subjects to possible changes.
- If a forwarded text, which becomes subject to interpretation, is to be understood, the underlying subjectivity must—according to Gadamer—be set into a relation with one's own thinking. The act of understanding effects a "fusion of horizons" between text and reader.
- Winograd and Flores establish a link between hermeneutics and information architecture. Tradition and language form the basis of human communication. In the same way that every man uses tools, the programmer uses his computer and the indexer his respective computer programs, methods of indexing and specific tools. "Breakdowns", in the sense of errors in the human flow

of activity, are regarded as positive occurrences, since in requiring solutions they encourage creative activity.
- Human work that requires decisions is called "cognitive work". Using an analysis of cognitive work, several dimensions that affect the entire working process within an environment can be uncovered. CWA (Cognitive Work Analysis) can contribute to the improvement of the further development of methods and tools of knowledge representation and information retrieval.
- When processing and searching information, the cognitive actor always stands in relation to information objects, information systems and interfaces as well as to the peripheral environment (organization, social and cultural context).

Bibliography

Brown, J.S., & Duguid, P. (2000). The Social Life of Information. Boston, MA: Harward Business School.

Buckland, M.K. (2012). Obsolescence in subject description. Journal of Documentation, 68(2), 154-161.

Day, R.E. (2010). Martin Heidegger's critique of informational modernity. In G.J. Leckie, L.M. Given, & J.E. Buschman (Eds.), Critical Theory for Library and Information Science (pp. 173-188). Santa Barbara, CA: Libraries Unlimited.

Fidel, R. (2006). An ecological approach to the design of information systems. Bulletin of the American Society for Information Science and Technology, 33(1), 6-8.

Gadamer, H.G. (1974). Hermeneutik. In J. Ritter (Ed.), Historisches Wörterbuch der Philosophie. Band 3 (pp. 1061-1074). Darmstadt: Wissenschaftliche Buchgesellschaft.

Gadamer, H.G. (1975). Truth and Method. London: Sheed & Ward.

Gust von Loh, S., Stock, M., & Stock, W.G. (2009). Knowledge organization systems and bibliographic records in the state of flux. Hermeneutical foundations of organizational information culture. In Thriving on Diversity – Information Opportunities in a Pluralistic World. Proceedings of the 72nd Annual Meeting of the American Society for Information Science and Technology, Vancouver, BC.

Heidegger, M. (1962[1927]). Being and Time. San Francisco, CA: Harper [Original: 1927].

Ingwersen, P., & Järvelin, K. (2005). The Turn. Integration of Information Seeking and Retrieval in Context. Dordrecht: Springer.

Mai, J.E. (2006). Contextual analysis for the design of controlled vocabularies. Bulletin of the American Society for Information Science and Technology, 33(1), 17-19.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, *Ba* and Leadership. A unified model of dynamic knowledge creation. Long Range Planning, 33(1), 5-34.

Pejtersen, A.M., & Fidel, R. (1998). A Framework for Work Centered Evaluation and Design. A Case Study of IR on the Web. Grenoble: Working Paper for MIRA Workshop.

Plato (1998). Cratylus. Indianapolis, IN: Hackett.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). Cognitive Systems Engineering. New York, NY: Wiley.

Simplicius (1954). Aristoteles Physicorum Libros Quattuor Posteriores Commentaria. Ed. by H.Diels. Berlin: de Gruyter.

Vicente, K. (1999). Cognitive Work Analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Winograd, T., & Flores, F. (1986). Understanding Computers and Cognition. A New Foundation for Design. Norwood, NJ: Ablex.

A.4 Documents

What is a Document?

The "document" is a crucially important concept for information science (Buckland, 1997; Frohmann, 2009; Lund, 2009). It is closely related to the concept of "resource", which is why both terms are used synonymously. In cases where only the intellectual content of a document is being considered, we speak of a "work". Floridi (2002, 46) even defines our entire scientific discipline via the concept of the document:

> Library and information science ... is the discipline concerned with documents, their life cycles and the procedures, techniques and devices by which these are implemented, managed and regulated.

For a long time, an area of information practice used to be described as "documentation" (Buckland & Liu, 1995)—some of the leading information science journals were called "American Documentation" (in the U.S.) or "Nachrichten für Dokumentation" (in Germany). Even today, the "Journal of Documentation" remains one of the top publications of information science (Hjørland, 2000, 28). Documentation deals with documents, both conceptually and factually. Like "documentation", "document" is etymologically made up of the Latin roots "doceo" and "mentum" (Lund, 2010, 743). "Doceo" refers to teaching and educating, pointing—like the etymology of "information"—to pedagogical contexts; "mentum" in Latin is a suffix used to turn verbs into nouns. Only from the 17th century onwards has "document" been understood as something written or otherwise fixed on a carrier (Lund, 2009, 400).

What is being stored in databases for the purposes of retrieval? We already know that distinctions must be made between documentary reference unit, documentary unit and original document—the documentary unit serving as the representation (surrogate) of the documentary reference unit. The latter is created through the database-appropriate analytical partition of documents into meaningful units. For instance: a book is a document, a single chapter within is the documentary reference unit and the data set containing the metadata (e.g. formal bibliographical description and content-depicting statements, perhaps complemented by the full text) is the documentary unit.

In a first approximation, we follow Briet in defining "document" as "any concrete or symbolic indexical sign [*indice*], preserved or recorded to the ends of representing, of reconstituting, or of providing a physical or intellectual phenomenon" (2006[1951], 10). It is intuitively clear that all manner of texts are documents. There is no restriction in terms of the documents' length. An extreme example of short texts is a correspondence between Victor Hugo and his publisher (Yeo, 2008, 139). After the publication of his novel "Les Misérables", Hugo sent a telegraph with only one character—"?"—apparently meaning to inquire about the sales numbers for his book. His publisher,

having understood the question, responded—as the novel had in fact been a great success—with "!" On the other hand, there are patent documents comprising more than 10,000 pages, and which still represent one single document.

Are there further document types besides texts? Buckland (1991, 586) claims that there is "information retrieval of more than text". The idea of infusing non-textual documents with informational added value goes back to Briet (2006[1951]). The antelope she described has become the paradigm for the (additional) processing of non-textual documents in documentation (Maack, 2004). Briet asks: "Can an antelope in the wilderness of Africa be a document?", to which the answer is a resounding "no". However, if the animal is captured and kept in a zoo, it becomes a "document", since it is now being consciously presented to the public with the clear intention of showing something. Buckland (1991, 587), discussing Briet's antelope, observes:

> Could a star in the sky be considered a "document"? Could a pebble in a babbling rock? Could an antelope in the wilds of Africa be a document? No, she (Briet, A/N) concludes. But a photograph of a star would be a document and a pebble, if removed from nature and mounted as a specimen in a mineralogical museum, would have become one. So also an antelope if it is captured, held in a zoo, and made an object of research and an educational exhibit would have become a "document". ... (A)n article written about the antelope would be merely a secondary, derived document. The primary document was the antelope itself.

Objects are "documents" if they meet the following four criteria:
- materiality (they are physically—including the digital form—present),
- intentionality (they carry meaning),
- development (they are created),
- perception (they are described as a document).

In Buckland's words (1997, 806), this reads as follows:

> Briet's rules for determining when an object has become a document are not made clear. We infer, however, from her discussion that:
> 1. there is materiality: Physical objects and physical signs only;
> 2. there is intentionality: It is intended that the object be treated as evidence;
> 3. the objects have to be processed: They have to be made into documents; and we think,
> 4. there is a phenomenological position: The object is perceived to be a document.

This definition is exceedingly broad. Following Martinez-Comeche (2000), it can be stated that 1st anything can be a document, but that 2nd nothing is a document unless it is explicitly considered as such. "A document is", according to Lund (2010, 747), "any result of human documentation." Documents are "knowledge artefacts that reflect the cultural milieu in which they arose" (Smiraglia, 2008, 29) and each document "is a product of its time and circumstances," Smiraglia (2008, 35) adds. For Mooers (1952), a document is the channel between a creator (e.g., author) and a receiver (e.g., reader) which contains messages.

This comprehensive definition of "document" now includes textual and non-textual, digital and non-digital resources. Textual and factual indexing and retrieval are always oriented on documents, leading Hjørland (2000, 39) to suggest "documentation science" as an alternative name for the corresponding scientific discipline (instead of "information science").

The Documents' Audience

Documents are no ends in themselves, but act as means of asynchronous knowledge sharing for the benefit of an audience. This audience consists of the factual or—in future—the hypothetical users. Like the documents' creators, the documents' users have different intellectual backgrounds and speak different jargons. Where author and user share the same background, Østerlund and Crowston (2011) speak of "symmetric knowledge", where they don't, of "asymmetric knowledge". The asymmetric knowledge of heterogeneous communities leads to the conception of "boundary objects" by Star and Griesemer (1989). "We can assume that documents typically shared among groups with asymmetric access to knowledge may take the form of boundary objects" (Østerlund & Crowston, 2011, 2). Such boundary documents "seem to explicate their own use in more detail" (Østerlund & Crowston, 2011, 7). It is possible for the document's user to produce a new document, which is then addressed to the author of the first document. Now we see a document cycle (Østerlund & Boland, 2009)—e.g. in healthcare between medical records, request forms, lab work reports, orders, etc. Boundary documents call for polyrepresentation in the surrogates: the application of either different Knowledge Organization Systems (KOSs) for the heterogeneous user groups or of one KOS with terminological interfaces for the user groups.

Documents are produced in a certain place at a certain time. For Østerlund (2008, 201) a document can be a "portable place" that "helps the reader locate the meaning and the spatio-temporal order out of which it emerges." In Østerlund's healthcare case study doctors communicate via documents. Here is a typical example: "One doctor could be reporting on her relation to the patient to another physician who has no prior knowledge of that patient" (Østerlund, 2008, 202).

Documents have a time reference. They have been created at a certain time and are—at least sometimes—only valid for a certain duration. The date of creation can play a significant role under certain circumstances. E.g., patents, when they are granted, have a 20-year term, starting from the date of submission.

It is a truism, but all documents are created in a social context and against a cultural background. What has influenced and continues to influence every single human being has been learned via interaction with his environment, culture, and past. Hence, language and documents are not neutral but are always used from a certain socio-cultural viewpoint. In hermeneutics, this is called the document's "horizon" (see Ch. A.3).

The surrogates of documents in information systems are documents as well. For Briet (2006[1951], 13), the creation of surrogates (called "documentation" following Otlet, 1934) is a "cultural technique". So the indexer (or the librarian, the knowledge manager, the information systems engineer, etc.) is an active and important part in the process of maintaining the institutional memory. He decides whether a document should be stored or not, and which metadata are used to describe the document. Such decisions are of great importance for the preservation (Larsen, 1999) and retrievability of documents. For Briet, the indexer is embedded in the social context of an institution; she calls him a "team player" (2006[1951], 15). "Documentation is part of public spaces of production—social networks and cultural forms," Day (2006, 56) comments on Briet's thesis.

Digital Documents and Linked Data

Documents are either available in digital or non-digital form; indeed, it is possible for both forms to coexist side by side. One example for such a parallel form is an article in the printed edition of a scientific journal, whose counterpart is available digitally in the World Wide Web (as a facsimile). Texts can always be digitalized (via scanning), at least in principle, which is not always the case for other objects. A museum object, e.g. our pebble, can never be digitalized. Here, we are always forced to represent the document exclusively via the documentary unit. Other objects, such as images, films or sound recordings, are principally digitizable, like texts.

The complete versions of the documents can only be entered into information retrieval systems if they are digitally available. If digitization is impossible as a matter of principle or for practical reasons (such as Briet's antelope) documentary reference units must be defined and documentary units created. In the case of digital documents, the original may under certain circumstances be entered into the system without any further processing. It is doubtful, however, whether it makes sense to forego the informational added value of a documentary unit. If both the documents and their surrogates are unified in a single system, we speak of "digital libraries".

In contrast to typical documents written on paper, digital documents (Buckland, 1998; Liu, 2004) display two new characteristics: fluidity and intangibility (Liu, 2008, 1). They do not have a tangible carrier (such as paper), consisting merely of files on a computer; they can principally be altered at any point (one good example for this being Wikipedia entries). Phases of stability alternate with phases of change (Levy, 1994, 26), in which the very identity of a document may be lost. It is easily possible at any point to preserve a stable phase (and thus the identity of a digital document), at least from a technological standpoint (Levy, 1994, 30).

A French research group—RTP-DOC (Réseau Thématique Pluridisciplinaire—33: Documents and content: creating, indexing, browsing) at the Centre National de la Recherche Scientifique (CNRS)—has discussed the concept "document" under the

aspects of digitization and networking. In publications, "RTP-DOC" became the pseudonym "Roger T. Pédauque" (Pédauque, 2003; cf. also Francke, 2005; Gradmann & Meister, 2008; Lund, 2009, 420 et seq.). Analogously to syntax, semantics and pragmatics, R. T. Pédauque (2003, 3) studies documents from the points of view of form, sign and medium. On the level of form, digital documents are a whole made up of structure and data. The structure describes, for instance, fonts that are used (e.g. in Unicode) or markup languages for web pages (such as HTML or XML), whereas data refers to the character sequences occurring within the document. On the level of sign, digital documents present themselves as the connection between informed text and knowledge; insofar as the text bears knowledge that a competent reader is able to extrapolate. The object in discussing digital documents as a medium is the pairing of inscription and legitimacy. That which is digitally written down has different levels of legitimacy. Pédauque (2003, 18) emphasizes:

> A document is not necessarily published. Many documents, for instance ones on private matters (...), or ones containing undisclosable confidential information can only be viewed by a very limited number of people. However, they have a social character in that they are written according to established rules, making them legitimate, are used in official relations and can be submitted as references in case of a dysfunction (dispute). Conversely, publication, general or limited, is a simple means of legitimization since once a text has been made public ... it becomes part of the common heritage. It can no longer easily be amended, and its value is appreciated collectively.

In the sense of legitimacy, we will distinguish between formally published, informally published and unpublished documents from here on out.

The fact that we can distinguish between structure and data on the syntactical level of digital documents means—at least in theory—that we are always capable of automatically extracting the data contained in a document and to save or interconnect them as individual factual documents. "In principle, any quantity of data carrying information can be understood as a document," Voß (2009, 17) writes. In order for all documents—text documents, data documents or whatever—to be purposefully accessible (to men and machines), they must be clearly designated. This is accomplished by the Uniform Resource Identifiers (URI) or the Digital Object Identifiers (DOI; Linde & Stock, 2011, 235-236). Under these conditions, it stands to reason that sets of factual knowledge which complement each other should be brought together even if they originate in different documents. This is the basic idea of "Linked Data" (Bizer, Heath, & Berners-Lee, 2009, 2):

> Linked Data is simply about using the Web to create typed links between data from different sources. ... Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets.

In cases of automatic fact extraction in the context of Linked Data, the original text document fades into the background, in favor of the received data and their factual documents (Dudek, 2011, 28). (Of course it remains extant as a document.) It must be noted that the data must be defined in a machine-readable manner (with the help of standards like RDF, the Resource Description Framework). Additionally, we absolutely require a knowledge organization system (KOS) that contains the needed terms and sets them in relation to one another (Dudek, 2011, 27).

In addition to extracted facts there are data files which cover certain topical areas, e.g. economic data (say, employment figures) and other statistical materials (cf. Linde & Stock, 2011, 192-201), geographic data (used for example by Google Maps) as well as real-time data (such as current weather data, data on delays in mass traffic or tracking data of current flights). All these data can—inasmuch as it makes sense—be combined in so-called "mash-ups". So it is possible to merge employment figures with geographical data resulting in a map which shows employment rates in different regions. Dadzie and Rowe (2011, 89) sum up:

> The Web of Linked Data provides a large, distributed and interlinked network of information fragments contained within disparate datasets and provided by unique data publishers.

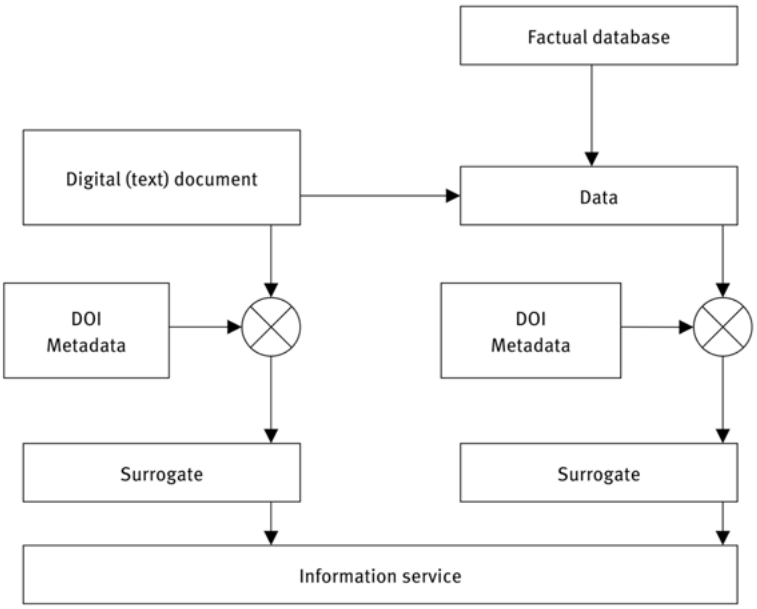


Figure A.4.1: Digital Text Documents, Data Documents and Their Surrogates in Information Services.

In Figure A.4.1, we attempt to graphically summarize the relations between digital text documents, the data contained therein and their representations as surrogates

in information services (such as the WWW, specialist databases or the so-called "Semantic Web"). In the information service, the linking of text documents (among one another and to the data) and data documents (also to one another and to the text documents) is facilitated via standards (like RDF) and the mediation of the KOS.

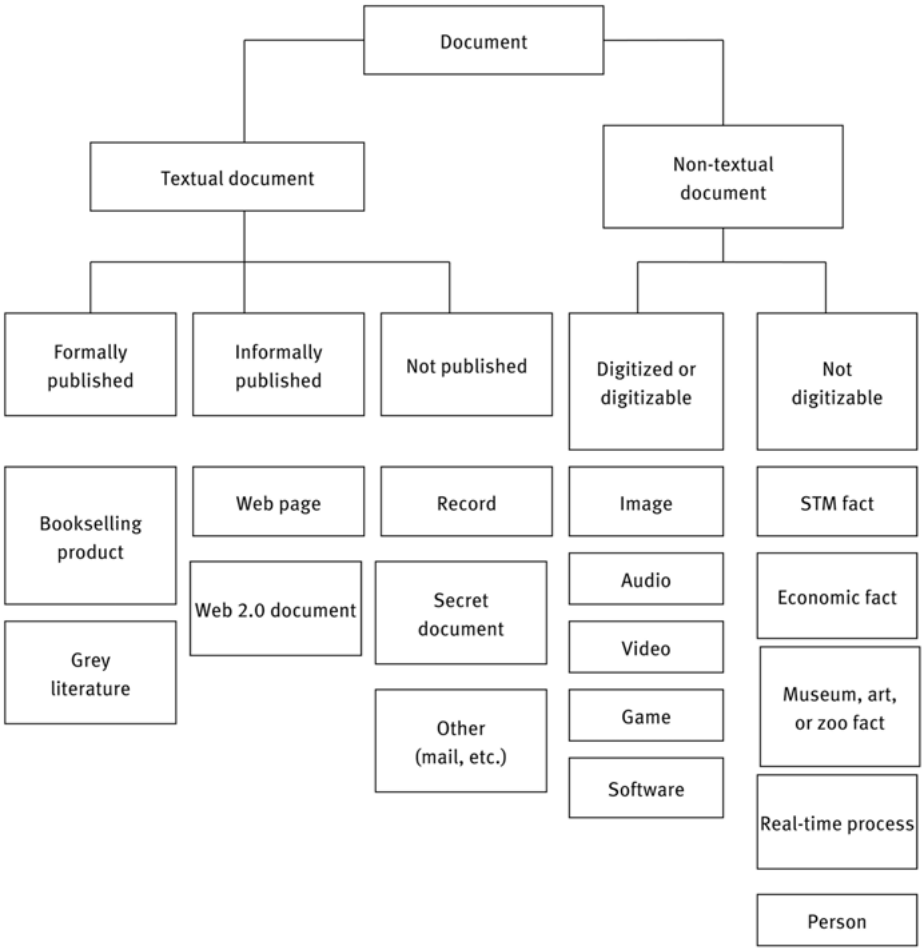


Figure A.4.2: A Rough Classification of Document Types.

An Overview of Document Types

In a first rough classification (Figure A.4.2), we distinguish between textual and non-textual documents. Text documents are either formally published (as a book, for instance), informally published (as a weblog) or not published (such as busi-

ness records). In the case of non-textual documents, we must draw lines between those that are available digitally or that can at least in principle be digitized (images, music, speech, video) and undigitizable documents. The latter include factual documents that have been extracted either intellectually, via factual indexing (resulting for example in a company dossier), or automatically from digital documents. In the context of information science, all documents are initially partitioned into units (dividing an anthology into its individual articles, for instance) and enriched via informational added values. In other words, they are formally described (e.g. by stating the author's name) and indexed for content (via methods of knowledge representation) (see Figure A.4.3). The resulting surrogates, or documentary units, carry metadata: text documents and digital (digitizable) non-textual documents contain bibliographical metadata (Ch. J.1), whereas undigitizable documents carry metadata about objects (Ch. J.2).

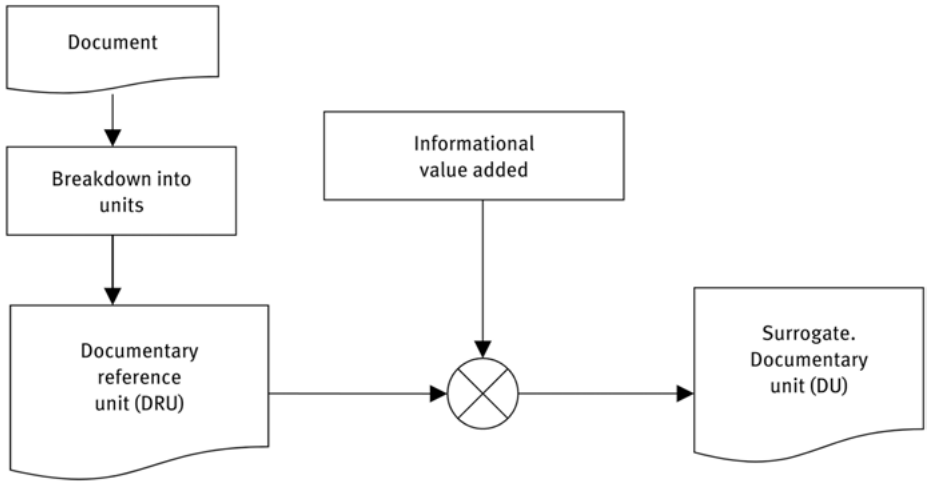


Figure A.4.3: Documents and Surrogates.

Formally Published Documents

There are textual documents that are published and there are those that are not. The former are divided into the two classes of (verified) formal and (generally unverified) informal texts. We define "text" in a way that includes documents that may partly contain elements other than writing (e.g. images).

Formal publications include all texts that have passed through a formal publishing process. This goes for all bookselling products: books, newspapers (including press reports from news agencies) and journals (including their articles), sheet music, maps etc. They also include all legal norms, such as laws or edicts as well as court

rulings, if they are published by a court. Furthermore, they include those texts that are the result of intellectual property rights processes, i.e. patents (applications and granted patents), utility models, designs and trademarks. A characteristic of formally published texts is that they have been verified prior to being published (by the editor of a journal, newspaper or publishing house, an employee of a patent office etc.). This "gatekeeper" function of the verifiers leads us to assume a degree of selection, and thus to ascribe a certain inherent quality to formally published documents.

From the perspective of content, we will divide formally verified documents into the following classes:

- Business, market and press documents (Linde & Stock, 2011, Ch. 7),
- Legal documents (Linde & Stock, 2011, Ch. 8),
- STM documents (science, technology and medicine) (Linde & Stock, 2011, Ch. 9),
- Fiction and art.

So-called "grey literature" (Luzi, 2000) is a somewhat problematic area, since there is no formal act of publication. It includes texts that are distributed outside the sphere of publishing and the book trade, e.g. university theses, series of "working papers" by scientific institutions or business documents by enterprises. In the case of university theses, the formal act of verification is self-evident; the situation is similar in the case of working papers, which, in general, present research results a lot more quickly than journals do. Business documents, such as company magazines or press reports, may contain texts about inventions that are not submitted for patents but whose priority must still be safeguarded. This alone is a reason for us to pay attention to these texts. For Artus (1984, 145),

> there can be no doubt ... that grey literature is just as legitimate an object of documentation as formally published literature.

Where can we locate the indivisible entity of a formally published document (IFLA, 1998)? Bawden (2004, 243) asks:

> (W)e ... have a problem deciding when two documents are "the same". If I have a copy of a textbook, and you have a copy of the same edition of the same book, then these are clearly different objects. But in a library catalogue they will be treated as two examples of the same document. What if the book is translated, word for word, as precisely as possible, into another language: is it the same document, or different? At what point do an author's ideas, on their way to becoming a published book, become a "document"?

We will answer Bawden's question in Chapter J.1.

Informally Published Texts

We now come to informally published texts. These mainly include those texts that are published on the Internet, respectively on one of its services. Note: we say published, but not verified.

Documents on the World Wide Web are marked via an individual address (URL: Uniform Resource Locator). The basis of the collaboration between servers on the WWW is the Hypertext Transfer Protocol (HTTP). The most frequent data type is HTML (Hypertext Markup Language); it allows users to directly navigate to other documents via links. A web page is a singular document with precisely one URL. A connected area of web pages is called a website (in analogy to a camping site, on which the individual tents, i.e. web pages, are arranged in a certain fashion), with the entry page of a website being called the homepage. Websites are run by private individuals, companies, scientific institutions, other types of institutions as well as specific online service providers (e.g. search engines or content aggregators). Certain web pages contain hardly any content or even none at all, only serving purposes of navigation. Besides HTML files, websites can contain other document types (such as PDF or WORD, but also graphic or sound files).

In Web 2.0 services in particular, many people collaborate in the publication of digital documents (Linde & Stock, 2011, Ch. 11). Texts on message boards are of a rather private or semi-private character. It is here that people who are interested in a given subject meet online in order to "chat". In contrast to chatrooms, the entries on message boards are saved.

Some sites on the WWW are continuously updated. They can be written in the form of a diary (like weblogs, or blogs in short) and contain either private or professional content, or they can filter information from other pages and incorporate them on their own URL. Short statements are called microblogs (like the "tweets" on Twitter or Weibo). Status updates in social networks (like Facebook) are posted frequently—depending on the user's activity—while personal data remain rather static.

A position in between formally and informally published texts is occupied by documents in Wikis. Since anyone can create and correct entries at any point (as well as retract erroneous corrections), we can describe such Web pages as cooperatively verified publications. The criteria for such verification, however, are entirely undefined. Likewise, the question as to which documents should be accepted into a Wiki in the first place, and to what level of detail, is probably not formally defined to any degree of exhaustiveness (Linde & Stock, 2011, 271-273).

All of these Web 2.0 services confront information retrieval systems with the task of locating the documents, indexing them and making them available to their users incredibly quickly—as soon after their time of publication as possible.

Unpublished Texts

The last group of text documents are texts that are not published. These include all texts that accrue for private individuals, companies, administrations, etc., i.e. letters, bills, files, memos, internal reports, presentations etc. and additionally secret documents. This area also includes those documents from institutions which are released on their Intranet or Extranet (part of the Internet that is protected and can only be accessed by authorized personnel, e.g. clients and suppliers). Such texts can be of great importance for the institutions concerned; their objective is to store them and make them available via information retrieval. This opens up the large area of corporate knowledge management.

Equally unpublished are texts in certain areas of the Internet, such as e-mails and chats. Here, too, information retrieval is purposeful, e.g. to aid the individual user in searching his sent and received mails or—with great ethical reservations—for intelligence services to detect "suspicious" content.

Documents and Records

In a specification for records management, the European Commission defines "record" via the concept of "document" (European Commission, 2002, 11):

> (Records are) (d)ocument(s) produced or received by a person or organisation in the course of business, and retained by that person or organisation. ... (A) record may incorporate one or several *documents* (e.g. when one document has attachments), and may be on any medium in any format. In addition to the content of the document(s), it should include contextual information and, if applicable, structural information (i.e. information which describes the components of the record). A key feature of a record is that it cannot be changed.

Are all files really documents (Yeo, 2011, 9)? Yeo (2008, 136) defines "records" with regard to activities:

> Records (are) persistent representations of activities or other occurrents, created by participants or observers of those occurrent by their proxies; or sets of such representations representing particular occurrents.

According to Yeo (2007; 2008), this relation to activities, or occurrences, is missing in the case of documents, so perhaps records cannot be defined as documents. However, we will construe records as a specific kind of documents, since, in the information-science sense, the same methods are used to represent them as for other documents. It is important to note, though, that a record (say, a personnel file) may contain various individual records (school reports, furloughs, sick leaves etc.). Depending on the choice of documentary reference unit, one can either display the personnel file as

a whole or each of the documents contained therein individually. When selecting the latter option, one must add a reference to the file as a whole. Otherwise, the connection is lost.

A similar division between superordinate document and its component parts (which are also documents) exists in archiving, specifically in the case of dossiers. In a press archive, for instance, there are ordered dossiers about people—generally in so-called press packs. These dossiers contain collections of press clippings with reports on the respective individual.

Non-Textual Image, Audio and Video Documents

In the case of non-textual documents, we distinguish between digital (or at least digitizable) and principally non-digital forms. Digitizable documents include visual media such as images (photographs, schematic representations etc.) and films (moving images), including individual film sequences, as well as audio media such as music, spoken word and sounds (Chu, 2010, 48 et seq.). Digital games (Linde & Stock, 2011, Ch. 13) as well as software applications (Linde & Stock, 2011, Ch. 14) also belong to this group. Such documents can be searched and retrieved via special systems of information retrieval—even without the crutch of descriptive texts—as in the example of searching images for their shapes and distributions of light and color (Ch. E.4).

Images, videos and pieces of music from Web 2.0 services (such as Flickr, YouTube and Last.fm) fall under the category of digital, informally published non-textual documents; they all have in common the fact that they can be intellectually indexed (via folksonomy tags).

Undigitizable Documents: Factual Documents

Undigitizable documents are represented in information retrieval systems via documentary units—the surrogates—only. We distinguish between five large groups:

- STM facts,
- Business and economic facts,
- Facts about museum artifacts and works of art,
- Persons,
- Real-time facts.

Of central importance for science, technology and medicine (STM) are facts, e.g. chemical compounds and their characteristics, chemical reactions, diseases and their symptoms, patients and their medical records as well as statistical and demographic data.

Non-textual business documents regard entire industries or single markets (i.e. with statements about products sold, revenue, foreign trade), companies (ownership, balance sheet ratio and solvency) up to individual products. The latter are important for information retrieval systems in e-commerce; the products must be retrievable via a search for their characteristics.

The third group of non-textual documents contains museum artifacts (Latham, 2012) and works of art. These include, among others, archaeological finds, historico-cultural objects, works of art in museums, galleries or private ownership as well as objects in zoos, collections etc. This is the documentary home of Briet's famous antelope.

A fourth group of undigitizable documents is formed by persons. They normally appear in the context of other undigitizable documents, e.g. as patients (STM), employees of a company (business) or artists (museum artifacts).

The last group of factual documents consists of data which are time-dependent and are collected real-time. There is a broad range of real-time data; examples are weather data or tracking data for ongoing flights.

In the intellectual compilation of metadata, the documentary reference units of undigitizable non-textual documents are described textually and, in addition, represented via audio files (e.g. an antelope's alarm call), photos (e.g. of a painting) or video clips (e.g. of a sculpture seen from different perspectives or in different lighting modes), where possible. In automatic fact extraction from digital documents, the respective data are either explicitly marked in the document or—with the caveat of great uncertainty—gleaned via special methods of information retrieval (Ch. G.6).

Conclusion

- Apart from texts, there are non-textual (factual) documents. According to the general definition, objects are documents if and only if they meet the criteria of materiality (including digitality), intentionality, development, and perception as a document.
- Digital documents display characteristics of fluidity and intangibility. Stable digital documents can be created and stored at any time.
- The data contained in digital documents can be extracted and transformed into individual data documents. All digital documents (texts as well as data) require a clear designation (e.g., a DOI) and must be indexed via metadata. It is subsequently possible to link texts and data with one another ("linked data").
- When discussing the legitimacy of documents, one must distinguish between formally published, informally published and unpublished documents. In the former, we can assume that the documents have been subject to a process of verification. This allows us to implicitly ascribe a certain measure of quality to these formal publications.
- Formally published text documents include bookselling media, legal norms and rulings, texts of intellectual property rights as well as grey literature.
- Informally published texts are available in large quantities on the World Wide Web. On the Web, we find interconnected individual web pages within a website, which has a firm entry page called

the homepage. Especially in the case of Wikis, there are web pages whose content is subject to frequent changes.
- Unpublished texts are all documents that accrue for companies, administrations, private individuals etc., such as letters, bills, memos etc. Important special forms are records, which refer to certain activities or other occurrences. These can be available non-digitally, but they can just as well be stored in an in-house Intranet or Extranet. These documents are objects of both records management and corporate knowledge management.
- Non-textual documents are either digital (or at least digitizable), such as visual (images, films) and audio media (music, spoken word, sounds) or they are principally non-digital, such as facts from science, technology and medicine, economic objects (industries, companies, products), museum artifacts, persons as well as facts from ongoing processes (weather, delays, flights, etc.).
- All documents without exception are, at least in principle, material for retrieval systems. Where possible, the systems should dispose of both the documentary units (with metadata as informational added value) and the complete digital forms of the documents (always including their DOI).

Bibliography

Artus, H.M. (1984). Graue Literatur als Medium wissenschaftlicher Kommunikation. Nachrichten für Dokumentation, 35(3), 139-147.

Bawden, D. (2004). Understanding documents and documentation. Journal of Documentation, 60(3), 243-244.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data. The story so far. International Journal of Semantic Web and Information Systems, 5(3), 1-22.

Briet, S. (2006[1951]). What is Documentation? Transl. by R.E. Day, L. Martinet, & H.G.B. Anghelescu. Lanham, MD: Scarecrow. [Original: 1951].

Buckland, M.K. (1991). Information retrieval of more than text. Journal of the American Society for Information Science, 42(8), 586-588.

Buckland, M.K. (1997). What is a "document"? Journal of the American Society for Information Science, 48(9), 804-809.

Buckland, M.K. (1998). What is a "digital document"? Document Numérique, 2(2), 221-230.

Buckland, M.K., & Liu, Z. (1995). History of information science. Annual Review of Information Science and Technology, 30, 385-416.

Chu, H. (2010). Information Representation and Retrieval in the Digital Age. 2nd Ed. Medford, NJ: Information Today.

Dadzie, A.S., & Rowe, M. (2011). Approaches to visualising Linked Data. A survey. Semantic Web, 2(2), 89-124.

Day, R.E. (2006). 'A necessity of our time'. Documentation as 'cultural technique' in *What Is Documentation.* In S. Briet, What is Documentation? (pp. 47-63). Lanham, MD: Scarecrow.

Dudek, S. (2011). Schöne Literatur binär kodiert. Die Veränderung des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken. Berlin: Humboldt-Universität zu Berlin / Institut für Bibliotheks- und Informationswissenschaft. (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft; 290).

European Commission (2002). Model Requirements for the Management of Electronic Records. MoReq Specification. Luxembourg: Office for Official Publications of the European Community.

Floridi, L. (2002). On defining library and information science as applied philosophy of information. Social Epistemology, 16(1), 37-49.

Francke, H. (2005). What's in a name? Contextualizing the document concept. Literary and Linguistic Computing, 20(1), 61-69.

Frohmann, B. (2009). Revisiting "what is a document?" Journal of Documentation, 65(2), 291-303.

Gradmann, S., & Meister, J.C. (2008). Digital document and interpretation. Re-thinking "text" and scholarship in electronic settings. Poiesis & Praxis, 5(2), 139-153.

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

IFLA (1998). Functional Requirements for Bibliographic Records. Final Report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München: Saur.

Larsen, P.S. (1999). Books and bytes. Preserving documents for posterity. Journal of the American Society for Information Science, 50(11), 1020-1027.

Latham, K.F. (2012). Museum object as document. Using Buckland's information concepts to understanding museum experiences. Journal of Documentation, 68(1), 45-71.

Levy, D.M. (1994). Fixed or fluid? Document stability and new media. In European Conference on Hypertext Technology (ECHT) (pp. 24-31). New York, NY: ACM.

Linde, F., & Stock, W.G. (2011). Information Markets. A Strategic Guideline for the I-Commerce. Berlin, New York, NY: De Gruyter Saur. (Knowledge & Information. Studies in Information Science.)

Liu, Z. (2004). The evolution of documents and its impacts. Journal of Documentation, 60(3), 279-288.

Liu, Z. (2008). Paper to Digital. Documents in the Information Age. Westport, CT: Libraries Unlimited.

Lund, N.W. (2009). Document theory. Annual Review of Information Science and Technology, 43, 399-432.

Lund, N.W. (2010). Document, text and medium. Concepts, theories and disciplines. Journal of Documentation, 66(5), 734-749.

Luzi, D. (2000). Trends and evolution in the development of grey literature. A review. International Journal on Grey Literature, 1(3), 106-117.

Maack, M.N. (2004). The lady and the antelope: Suzanne Briet's contribution to the French documentation movement. Library Trends, 52(4), 719-747.

Martinez-Comeche, J.A. (2000). The nature and qualities of the document in archives, libraries and information centres and museums. Journal of Spanish Research on Information Science, 1(1), 5-10.

Mooers, C.N. (1952). Information retrieval viewed as temporal signalling. In Proceedings of the International Congress of Mathematicians. Cambridge, Mass., August 30 – September 6, 1950. Vol. 1 (pp. 572-573). Providence, RI: American Mathematical Society.

Østerlund, C.S. (2008). Documents in place. Demarcating places for collaboration in healthcare settings. Computer Supported Cooperative Work, 17(2-3), 195-225.

Østerlund, C.S., & Boland, R.J. (2009). Document cycles: Knowledge flows in heterogeneous healthcare information system environments. In Proceedings of the 42nd Hawai'i International Conference on System Sciences. Washington, DC: IEEE Computer Science.

Østerlund, C.S., & Crowston, K. (2011). What characterize documents that bridge boundaries compared to documents that do not? An exploratory study of documentation in FLOSS teams. In Proceedings of the 44th Hawai'i International Conference on System Sciences, Washington, DC: IEEE Computer Science.

Otlet, P. (1934). Traité de Documentation. Bruxelles: Mundaneum.

Pédauque, R.T. (2003). Document. Form, Sign and Medium, as Reformulated for Electronic Documents. Version 3. CNRS-STIC. (Online).

Smiraglia, R.P. (2008). Rethinking what we catalog. Documents as cultural artifacts. Cataloging & Classification Quarterly, 45(3), 25-37.

Star, S.L., & Griesemer, J.R. (1989). 'Institutional ecology', 'translations' and boundary objects. Amateurs and professionals in Berkeley's Museums of Vertebrate Zoology 1907-39. Social Studies of Science, 19(3), 387-420.

Voß, J. (2009). Zur Neubestimmung des Dokumentbegriffs im rein Digitalen. Libreas. Library Ideas, No. 15, 13-18.

Yeo, G. (2007). Concepts of record (1). Evidence, information, and persistent representations. The American Archivist, 70(2), 315-343.

Yeo, G. (2008). Concepts of record (2). Prototypes and boundary objects. The American Archivist, 71(1), 118-143.

Yeo, G. (2011). Rising to the level of a record? Some thoughts on records and documents. Records Management Journal, 21(1), 8-27.

A.5 Information Literacy

Information Science in Everyday Life, in the Workplace, and at School

According to UNESCO and the IFLA, information literacy is a basic human right in the digital world. In the "Alexandria Proclamation on Information Literacy and Lifelong Learning" (UNESCO & IFLA, 2005), we read that

> Information Literacy ... empowers people in all walks of life to seek, evaluate, use and create information effectively to achieve their personal, social, occupational and educational goals.

If information literacy is understood as such a human right (Sturges & Gastinger, 2010), social inequalities in the knowledge society—the digital divide—can be counteracted and the individual's participation in the knowledge society strengthened (Linde & Stock, 2011, 93-96). Information literacy refers to the usage of information science knowledge by laymen in their professional or private lives. It refers to a level of competence in searching and finding as well as in creating and representing information—a competence that does not require the user to learn any information science details. In the same way that laypeople are able to operate light switches without knowing the physics behind alternating current, they can use Web search engines or upload resources to sharing services and tag them without having studied information retrieval and knowledge representation. To put it very simply, information literacy comprises those contents of this book that are needed by everyone wishing to deal with the knowledge society. Information literacy plays a role in the following three areas:

- information literacy in the Everyday,
- information literacy in the Workplace,
- information literacy at School (Training of information literacy skills).

Competences are arranged vertically in a layer model (Catts & Lau, 2008, 18) (Figure A.5.1). Their basis is literacy in reading, writing and numeracy. ICT and smartphone skills as well as Media Literacy are absolutely essential in an information society (where information and communication technology are prevalent; Linde & Stock, 2011, 82). Everybody should be able to operate a computer and a smartphone, be literate in basic office software (text processing, spreadsheet and presentation programs), know the Internet and its important services (such as the WWW, e-mail or chat) and use the different media (such as print, radio, television and Internet) appropriately (Bawden, 2001, 225). All these competences mentioned above are preconditions of information literacy. The knowledge society is an information society where, in addition to the use of ICT, the information content—i.e., the knowledge—is available everywhere and at any time. Lifelong learning is essential for every single member of such a society (Linde & Stock, 2011, 83). Accordingly, important roles are played by people's

information literacy and their ability to learn (Marcum, 2002, 21). Since a lot of information is processed digitally, a more accurate way of putting it would be to say that it is "digital information literacy" (Bawden, 2001, 246) which we are talking about.

In investigating information literacy, we pursue two threads. The first of these deals with practical competences for information retrieval. It starts with the recognition of an information need, proceeds via the search, retrieval and evaluation of information, and leads finally to the application of information deemed positive. The second thread summarizes practical competences for knowledge representation. Apart from the creation of information, it emphasizes their indexing and storage in digital information services as well as the ability to sufficiently heed any demands for privacy in one's own information and others'. For both of these threads, it is of great use to possess basic knowledge of information law (Linde & Stock, 2011, 119-157) and information ethics (Linde & Stock, 2011, 159-180).

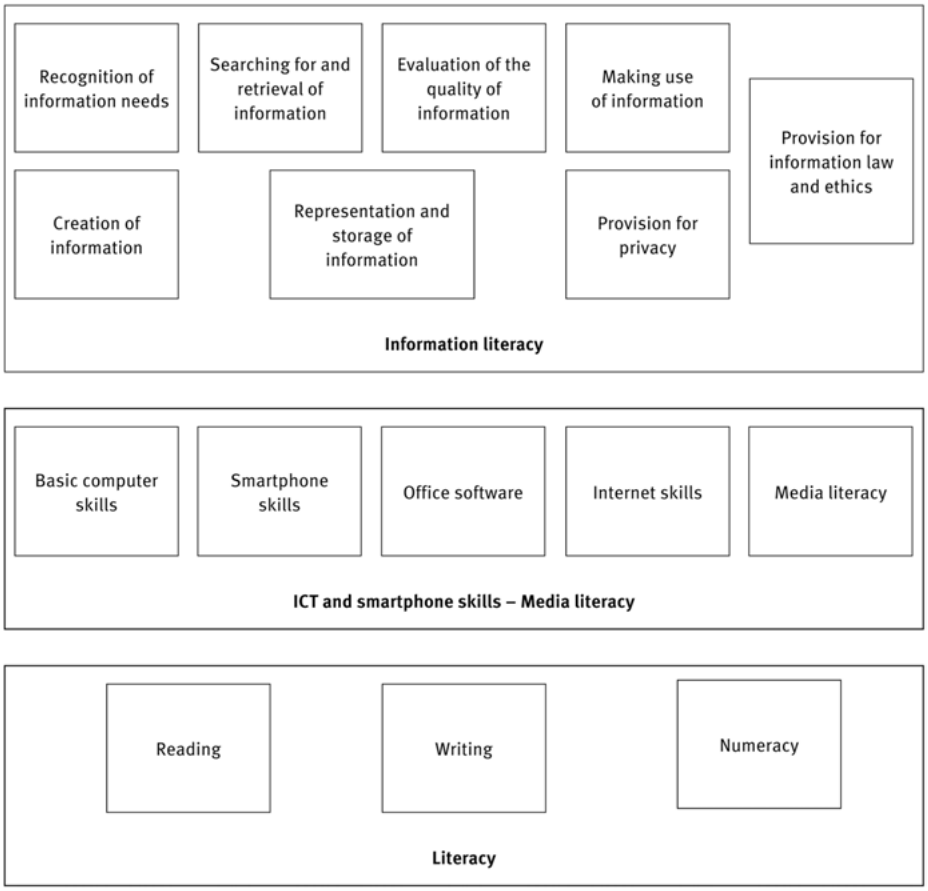


Figure A.5.1: Levels of Literacy.

Thread 1, which includes Information Retrieval Literacy, has been pursued for more than thirty years. It is rooted in bibliographic and library instruction and is practiced mainly by university libraries. A closely related approach is information science user research, e.g. the modelling of research steps in Kuhlthau (2004) (see Ch. H.3). Eisenberg and Berkowitz (1990) list "Six Big Skills" that make up information literacy: (1) task definition, (2) information seeking strategies, (3) location and access, (4) use of information, (5) synthesis, and (6) evaluation. The Association for College and Research Libraries (ACRL) of the American Library Association provides what has become a standard definition (Presidential Committee on Information Literacy, 1989):

> To be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information.

The British Standing Conference of National and University Libraries (SCONUL) summarizes Retrieval Literacy into seven skills (SCONUL Advisory Committee on Information Literacy, 1999, 6):

- The ability to recognise a need for information
- The ability to distinguish ways in which the information 'gap' may be addressed
- The ability to construct strategies for locating information
- The ability to locate and access information
- The ability to compare and evaluate information obtained from different sources
- The ability to organise, apply and communicate information to others in ways appropriate to the situation
- The ability to synthesise and build upon existing information, contributing to the creation of new knowledge.

When confronted with technically demanding information needs, Grafstein (2002) emphasizes, the general retrieval skills must be complemented by a specialist component capable of dealing with the contents in question. Thus it would hardly be possible to perform and correctly evaluate an adequate research for, let us say, Arsenicin A (perhaps regarding its structural formula) without any specialist knowledge of chemistry. "(B)eing information literate crucially involves being literate *about something*" (Grafstein, 2002, 202).

In the year 2000, the ACRL added another building block to their information literacies (ACRL, 2000, 14):

> The information literate student understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.

Important subsections here are copyright as well as intellectual property rights.

The second thread, including Knowledge Representation Literacy, has become increasingly important with the advent of Web 2.0 (Huvila, 2011). Web users, who used to be able only to request information in a passive role, now become producers

of information. These roles—information consumers and information producers—are conflated in the terms "prosumer" (Toffler, 1980) and "produsage" (Bruns, 2008), respectively. Users create information, such as blog posts, wiki articles, images, videos, or personal status posts, and publish them digitally via WordPress, Wikipedia, Flickr, YouTube and Facebook. It is, of course, important that these resources be retrievable, and so their creators give them informative titles and index them with relevant tags in the context of folksonomies (Peters, 2009). Here, too, aspects of information law and ethics must be regarded (e.g. may I use someone else's image on my Facebook page?). Additionally, the user is expected to have a keen sense for the level of privacy they are willing to surrender and the risks they will thus incur (Grimmelmann, 2009).

Information Literacy in the Workplace

In corporate knowledge management, an employee is expected to be able to acquire and gainfully apply the information needed to perform his tasks. Knowledge thus acquired in the workplace must be retrievable at any time. In knowledge-intensive companies in particular (e.g. consultants or high-tech companies), competences in information retrieval and knowledge representation are of central importance for almost all positions. With regard to Web 2.0 services, employees are expected not to reveal any secrets or negative opinions about their employer—be it via the company's official website or via private channels. For Lloyd (2003) and Ferguson (2009), knowledge management and information literacy are closely related to one another. Bruce (1999, 46) describes information competences in the workplace:

> Information literacy is about peoples' ability to operate effectively in an information society. This involves critical thinking, an awareness of personal and professional ethics, information evaluation, conceptualising information needs, organising information, interacting with information professionals and making effective use of information in problem-solving, decision-making and research. It is these information based processes which are crucial to the character of learning organisations and which need to be supported by the organisation's technology infrastructure.

Bruce (1997) lists seven aspects ("faces") of information literacy, each communicating with organizational processes in the enterprise (Bruce, 1999, 43):

- Information literacy is experienced as using information technology for information awareness and communication (workplace process: environmental scanning).
- Information literacy is experienced as finding information from appropriate sources (workplace process: provision of inhouse and external information resources and services).

- Information literacy is experienced as executing a process (workplace process: information processing; packaging for internal/external consumption).
- Information literacy is experienced as controlling information (workplace process: information / records management, archiving).
- Information literacy is experienced as building up a personal knowledge base in a new area of interest (workplace process: corporate memory).
- Information literacy is experienced as working with knowledge and personal perspectives adopted in such a way that novel insights are gained (workplace process: research and development).
- Information literacy is experienced as using information wisely for the benefit of others (workplace process: professional ethics / codes of conduct).

The task of organizing the communication of information literacy in enterprises falls to the department of Knowledge Management or to the Company Library (Kirton & Barham, 2005, 368). Learning in the workplace, and thus learning information literacy, "is a form of social interaction in which people learn together" (Crawford & Irving, 2009, 34). In-company sources of information gathering also include digital and printed resources, but a role of ever-increasing importance is being played by one's colleagues. Crawford and Irving (2009, 36) emphasize the significance of face-to-face information streams:

> The traditional view of information as deriving from electronic and printed sources only is invalid in the workplace and must include people as a source of information. It is essential to recognize the key role of human relationships in the development of information literacy in the workplace.

Thus it turns out to be a task for corporate knowledge management to create and maintain spaces for information exchange between employees. Information literacy in the workplace requires company libraries to install information competence in employees, create digital libraries (to import relevant knowledge into the company and to preserve internal knowledge), as well as to operate a library as a physical space for employees' face-to-face communication. Methods of knowledge management "ask, e.g., for knowledge cafés (...), space for story telling (...) and for corporate libraries as places" (Stock, Peters, & Weller, 2010, 143).

Information Literacy in Schools and Universities

According to Catts and Lau (2008, 17), the communication of information literacy begins at kindergarten, continues on through the various stages of primary education and leads up to university. The teaching and learning of information literacy is fundamentally "resource-based learning" (Breivik, 1998, 25), or—in our terminology (see Ch. A.4)—"document-based learning". Retrieval literacy allows people to retrieve documents that satisfy their information needs. Knowledge representation grants people