UNIVERSITY OF COPENHAGEN

UNISIST model and knowledge domains

Hjørland, Birger; Andersen, Jack; Søndergaard, Trine Fjordback

Published in:
Encyclopedia of library and information science

Publication date:
2005

Document version
Publisher's PDF, also known as Version of record

UNIVERSITY OF COPENHAGEN

UNISIST Model and Knowledge Domains

Birger Hjørland
Trine Fjordback Søndergaard
Jack Andersen
Royal School of Library and Information Science, Copenhagen, Denmark

INTRODUCTION

Scientific and scholarly communication is an important area within information science (IS). The social system of communication consists of knowledge producers, intermediaries, and users. These groups of people (or *actors*) are different kinds of professionals. The social system also comprises *institutes* such as research institutes, publishers, and libraries. The actors and institutions perform information *services* such as writing, publishing, storing, and retrieving documents and information. The actors communicate in both formal and informal ways and produce different kinds of *documents* such as journal articles, books, book reviews, proceedings, bibliographies, catalogs, dictionaries, handbooks, encyclopedias, and review articles.

It is of obvious interest for IS to develop adequate models of this system of scholarly communication. It is of course a dynamic system, and new information technologies are one of the causes underlying the dynamics. Many different attempts have been made to model this system. The best one is probably that proposed by UNISIST in 1971. The United Nations Information System in Science and Technology (UNISIST) was a program developed by the United Nations Educational, Scientific and Cultural Organization (UNESCO) to improve scientific and technological communication. The UNISIST program has been terminated, but the model has been updated and modified and is an important analytical tool in IS. Whereas the original model only considered scientific and technological communication, the modified version also considers social sciences and humanities and is regarded an important analytical model in domain analysis.

OVERALL STRUCTURE AND ITS ELEMENTS[a]

The original UNISIST model of scientific communication was proposed in 1971 in a report.[1, p.26] Fig. 1 is a reproduction of this original model. This is but one of many communication models. In spite of its age, newer models have not been able to replace it. The model describes information communication between knowledge producer and knowledge user, as a socio-technical system consisting of diverse organizational and documentary units, each contributing to the division of labor in scholarly communication. A mapping of these agents, their information services, and document production should be an essential task of library and IS research.

This model was originally proposed as a generalized model of the information structures within science and technology but may be generalized to other fields as well.

The point of departure in the model is the knowledge *producers*. According to the model, there are three main categories of information distribution channels available for communicating research: *informal* and *formal communication channels*, and *tabular channels*.

Informal communication may be in oral or written form and takes place when the producer and the user know each other and exchange information via channels such as personal correspondence (i.e., letters), manuscripts and preprints, private exchange of bibliographical references, etc. It might also occur semiformally as in professional conferences, meetings, or lectures.

Two means of *formal ways of communicating* research are depicted in the UNISIST figure: *published documents* and *unpublished documents*. Published documents (books and journals) go from the knowledge producer through publishers or editors, from whom they reach the users, perhaps through abstracting and indexing services, and libraries and information centers. Bibliographic control of the documents contributes in ensuring the visibility and publicity of documents. A published document has an increased amount of visibility compared to unpublished ones.

Unpublished documents consist, among other things, of theses, and research and technical reports distributed in limited copies by, e.g., government agencies. These documents may reach the users through clearinghouses and information centers. Hence, they
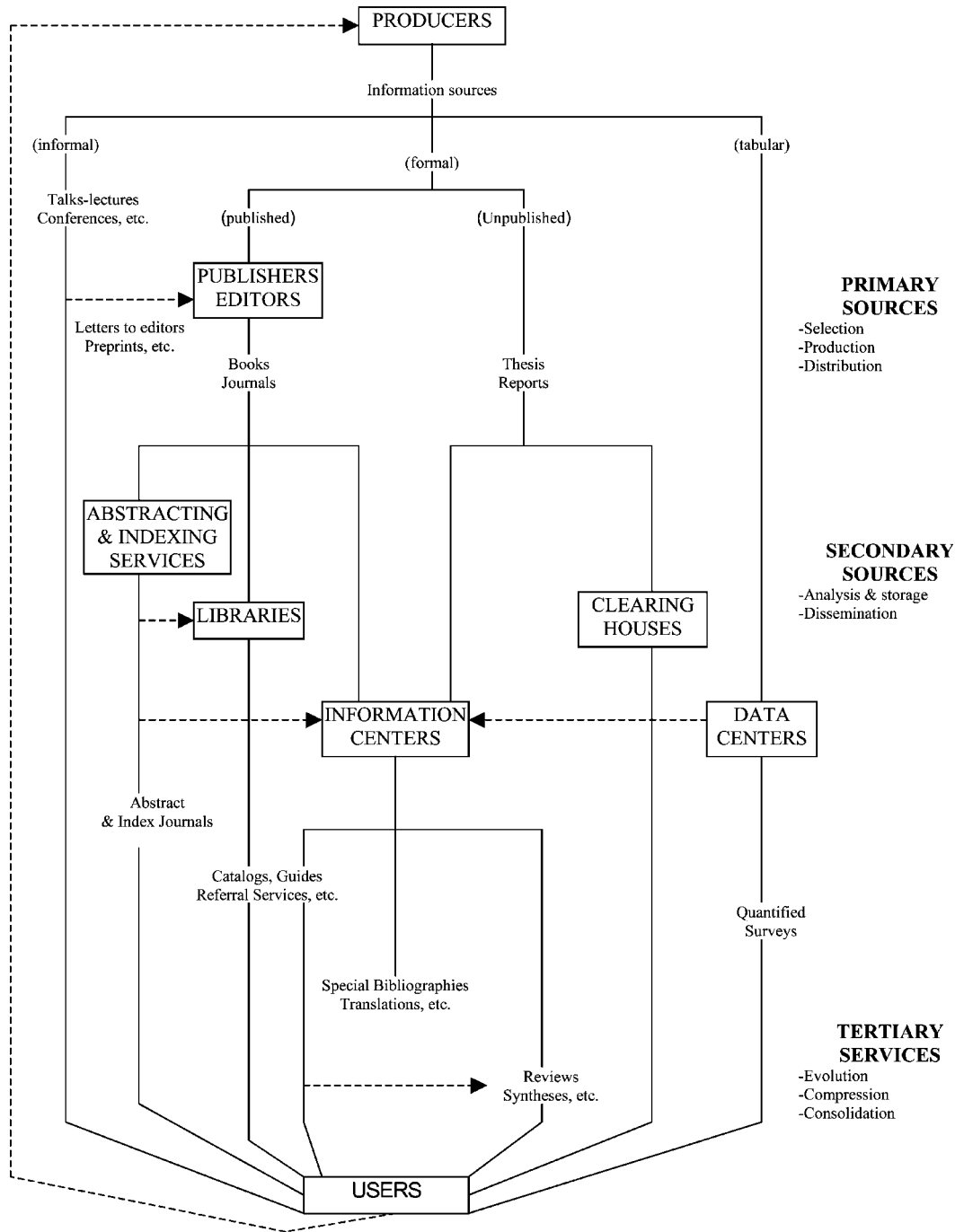
---

[a]The description of this model is mainly based on Fjordback Søndergaard, Andersen, and Hjørland.[3] Readers are referred to this paper for more details.

Fig. 1 The flow of scientific and technical information. (From Ref. [1]; p. 26.)

are not subject to the same selection, production, and distribution procedures of publishers and editors that published documents go through. The notion of *gray literature* is somewhere in between published and unpublished documents. A private letter is unpublished, while a report registered by a clearinghouse is semipublished or gray literature. Ph.D. dissertations, published on demand at the University Microfilm

International (UMI) and indexed in *Dissertation Abstracts International*, are another example of semipublications.

The third category of communicating research is through *tabular channels*. This consists of numerical data. The original model chooses to regard this an independent channel, although it recognizes that much tabular information is distributed in normal

publications (and also in semipublications or in unpublished documents).[b]

So far, *the primary sources* of scientific and technical information have been described in terms of their selection, production, and distribution functions. Primary literature is the point of departure in the production of scientific and scholarly knowledge and for the communicative division of labor of the literatures involved. The task of this literature is to produce and present original knowledge claims; this is why primary literature places emphasis on methodological descriptions.

In the UNISIST model, we observe three levels of information source services between the knowledge producers and users: Besides the primary sources, there are also secondary and tertiary sources/services.[c] This is the basic taxonomy of information sources in the model (see the Appendix for the full model including the categories added in the revised version). *The secondary information services* register and describe primary documents for the purpose of retrieval and documentation. Secondary literature, such as subject bibliographies; citation indexes; library catalogs; and databases, analyses, describes, and registers primary literature[d] mainly in these bibliographical instruments. The central working processes of the secondary sources are analysis, storage, and dissemination. The model shows abstracting and indexing services, libraries, information centers, clearinghouses, and data centers, which are considered secondary information services, each with particular functions to perform. *Tertiary literature* is, for example, professional handbooks and encyclopedias and literature reviews, which consolidate, collect, and synthesize the primary literature.

Abstracting and Indexing Services

The UNISIST model distinguishes between two kinds of abstracting and indexing services. The first one consists of printed secondary journals prepared and distributed by scientific associations that operate on a profit basis. An example of this kind is Chemical Abstracts or the Citation Indexes produced by the Institute for Scientific Information (ISI); that is the content of these does not represent collection of a physical location such as a library. The other service consists of catalogs and index files compiled by the staff of libraries or information centers as a mean of accessing their own collections. Also, guides and referral services belong to the second service produced by libraries and information centers. Referral services provide an indication of sources (persons, institutions, publications, etc.), from which scientific information may be obtained on a given subject and are mechanisms for switching users to such sources.

Information Centers

The concept of an information center is ambiguous, and the model's relation to similar concepts such as libraries or documentation centers is not clarified. According to the model, the information center combines some of the functions of secondary journals and specialized libraries, to which are added specific duties such as the selective dissemination of information, or the preparation of state-of-art monographs, trend reports, reviews, etc. for the benefit of a specialized field or well-defined and more restricted user groups. "The role of such [information] centers is sometimes spoken of as that of 'repackaging' the information provided by abstracting and indexing services, according to the requirements of specific users; they operate then as 'tertiary' services, with a synthesis function added to those of indexing and classification: reorganization, quality control, compression, synthesis, evaluation etc."[1, p. 30] Typical examples of tertiary documents are as previously mentioned literature reviews and syntheses.

Because information centers perform both secondary and tertiary functions, they actually differ from libraries. It may be stated that whereas libraries are a kind of information center, the latter is not a kind of library. Libraries do not usually produce reviews, syntheses, or other kinds of tertiary documents. Moreover, information centers are normally not in the possession of a physical collection of documents and are not primarily concerned with giving access to these collections, as are libraries. When the UNISIST (1971) model was proposed, information centers were often conceived of as centers that provided bibliographic references or information from documents "information" but not documents. See also Ref.[2]). It has often been seen as an ideal means to present pure information, whereas documents have been viewed as obsolete medium of information. Centers that delivered bibliographical records (usually for a fee) from

---

[b]The UNISIST model privileged "tabular data" that presumably meant socio-economic numeric data series, tables of constants, observational data in numeric form, and the like; but it was an awkward phrase even then. That view (or at least wording) is obsolete in the sense that the key characteristic now is "digital" not "tabular" and includes vast databases of images and other resources not usually considered "tabular data."

[c]The terms primary (or secondary or tertiary) information sources may be used interchangeably with primary (or secondary or tertiary) information services: The actors or institutions that produce the sources represent the services. Also, the term literature(s) might be used interchangeably with information sources.

[d]Although emphasis is on primary sources, secondary sources may also include tertiary sources.

electronic databases grew rapidly developed from about 1965 until 1990 and were usually part of libraries, particularly in the scientific and technological fields. Such departments were important, for example, those in the Danish Technological Library and the Main Medical Library in Copenhagen (and correspondingly, for example, centers in Stockholm as well as in other cities in the world). They were usually referred to as *Documentation Departments*. With the introduction of CD-ROM databases and "end-user" oriented search interfaces around 1990, these departments have by and large disappeared or have been downsized.

The UNISIST model operates only for a basic concept of information center. In the following, we use information center as an umbrella term for libraries, documentation centers, and other similar activities concerning the collection, dissemination, storage, retrieval, and organization of documents (or knowledge).

Clearinghouses

Clearinghouses are defined by the UNISIST as "...institutions entrusted with the procurement and dissemination of special categories of documents, such as technical reports, dissertation theses, thesauri, etc."[1, p. 147] According to the UNISIST report, the modes of analysis, storage, and dissemination here are the same as those of libraries or information centers. However, what differentiates clearinghouses from libraries or information centers is the attention exclusively paid to unpublished documents.[e]

Data Centers

The UNISIST model conceives of data centers as being different from the abovementioned secondary services because "...they [data centers] deal exclusively with the raw material of science, parallel or even prior to publication...,"[1, p. 30] and because functions (cataloging, abstracting, indexing, synthesizing, etc.) of the other kinds of secondary services "...are normally defined with respect to written documents."[1, p. 31] That is, the UNISIST model conceives of data centers as being concerned with "raw" data and nonwritten documents such as quantified surveys.

Fjordback Søndergaard, Andersen, and Hjørland[3] in their revision decided, however, to consider data

centers as part of the other units in the model as data that are published through formal communication channels such as publishers. This implies that we will exclude data centers as an independent form of information service and information source. This decision to not regard data centers as a specific communication channel is also motivated by the facts that other kinds of messages such as computer programs, pictures, and sounds are not represented by separate channels.

Special Bibliographies, Translations, etc.

As seen in the UNISIST model, organizations for information and documentation may produce special bibliographies, whether current or those of earlier times. *Translation services* may be commercial, or as in government centers they may be special translation journals (e.g., "cover to cover" translation of Russian scientific journals to English). They may also be bibliographies covering translations (such as *Index translationum. Répertoire International des Traductions/ International Bibliography of Translations*—published by UNESCO).

Review, Syntheses, etc.

Here, reviews should not be confused with book reviews. Reviews are syntheses of the primary literature, e.g., in the form of handbooks, review articles, scientific and professional encyclopedias, and the like. (This is unlike works like general encyclopedias that are not primarily written for subject specialists nor are a part of the scientific literature described in the original UNISIST model. They may be seen as a special kind of literature, popular literature.) One important kind of tertiary literature is the series titled "Annual Reviews of X," e.g., the *Annual Review of Information Science and Technology*. Many of such annual reviews have high impact factors in their respective fields.

Users

The final unit in the model is the users. Users of scientific and technical information are in most cases also identical to the producers, or they may also be practitioners such as physicians. The different roles of users determine their information needs and make specific constraints on their use of information. Researchers are generally intensive users of informal and primary sources within a narrow specialty, whereas practitioners are generally users of a broader range of sources and may rely more on tertiary sources.

---

[e]When broadening the UNISIST model to other fields such as the humanities, there is an obvious need for comparing the functions of clearinghouses and archives.

THEORETICAL REVISION AND TECHNOLOGICAL UPDATION OF THE MODEL

There are two main reasons why Fjordback Søndergaard, Andersen, and Hjørland[3] found it necessary to update and revise the original model:

1. The need to consider differences between different disciplines and domains and to generalize the model and its *document typology* (or the system of kinds of document) from science and technology to other domains.
2. The growing impact of the Internet on scientific and scholarly communication.

These considerations have been incorporated in the modified model shown in Figs. 2 and 3 and are introduced in the rest of this article. In Fig. 2, the Internet services have been displayed as a separate entity, whereas print and electronic media are integrated as seen in Fig. 3

Concern 1: Considering Different Domains

Fig. 2 has a punctured ellipse symbolizing a scientific discipline or a knowledge domain. The ellipse can symbolize, for example, the biological, the medical, or the legal discourse community. Such domains are typically overlapping, open structures, as is the case in the biomedical domain, where the borders of biology and medicine are difficult to identify. The ellipse is (more or less) punctured in order to illustrate this open nature of domains. The knowledge producers, the users, and the intermediaries are thus all considered members of a domain (or discourse community). Different services and institutions may, however, be shared between different domains. This is especially the case with interdisciplinary institutions and tools such as, for example, the *Science Citation Index* and national libraries. A given domain may, in other words, exercise more or less control with regard to its own informational infrastructure, and the institutions (e.g., publishers and libraries) may or may not have adequate subject knowledge and engagements to fulfill the functions in optimal ways.

A new feature in the revised model is the arrows symbolizing the import and export of knowledge into the domain and out of the domain. The revised model recognizes that the information producers also use information sources. They may get information from sources in their own domain, other domains, direct observation of natural information sources, or from the general public. Some domains get most of their knowledge from their own domain. This is the case with mathematics, economics, and psychology. Other domains such as agriculture get most of their input from other domains, e.g., chemistry. Different epistemologies in a given domain will emphasize different knowledge sources. Social constructivism is, for example, an epistemological position that claims that researchers' direct observations of nature are mediated by information sources of a social nature, which is a contrast to an empiricist or positivist epistemology.

Export may be directed toward other disciplines or the general public. In engineering, physical products are, e.g., speakers or cars, the normal products exported from the knowledge production in the domain. In this domain, researchers seem to prefer making commercial products rather than writing research papers. Such products may however, be documented in patents. In science, publication of articles in journals is the norm, whereas journals compete with books in humanities. The articles may be printed in journals of the researcher's own discipline, in those related to other disciplines, or in general scientific journals. The category of popular literature in the document typology is especially designed to serve dissemination of knowledge to the public domain or to beginners (See Appendix, "Popular literature"). Domains vary much in their exports of knowledge to the general public. Humanities typically have a much closer relationship to the mass media and the general public compared to scientific publications.

The domain itself reflects the division of labor in society (e.g., the division between people working with health problems in the health domain and those working with legal problems in the legal domain). Inside each domain, there is a more or less developed internal division of labor, e.g., between theoretical and empirical researchers, assistants, administrators, computer specialists, librarians/information specialists, translators, publishers, practitioners, etc. Often, the practitioners are the end users of the knowledge produced by the researchers. This can be, for example, the doctor curing a patient by applying new research results. The most important thing to realize is that each domain has a unique structure that should be described empirically and explained theoretically. A central point in the domain analytic approach to IS is the claim that tools, languages for special purposes (LSP), concepts, meaning, information structures, information needs, and relevance criteria are shaped in discourse communities. Different scientific, scholarly, or professional domains have *unique* structures of documents, which reflect an adaptation to the special needs in the domain. Some examples of unique kinds of documents are:

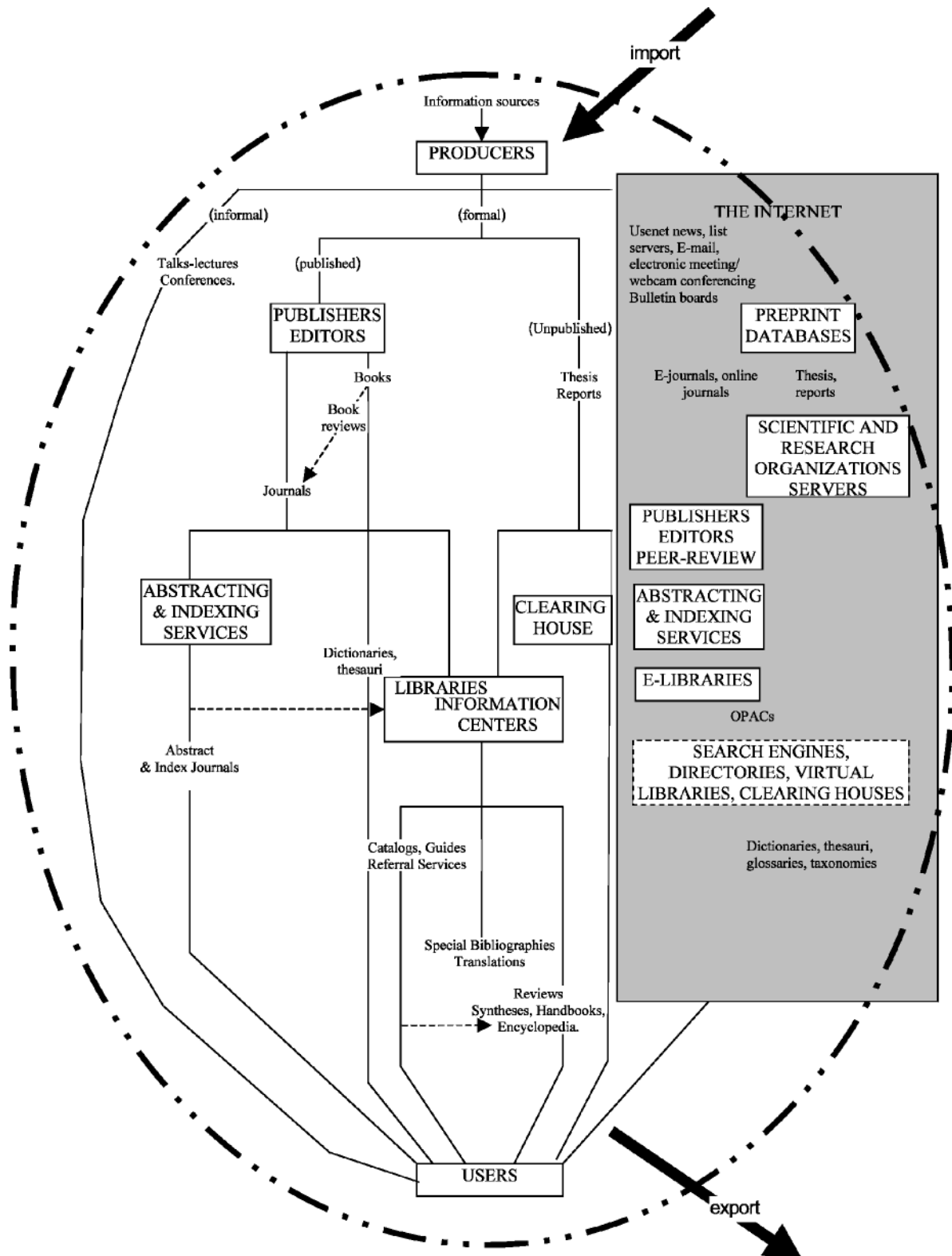- Almanacs in astronomy
- Patents in engineering

Fig. 2 The revised UNISIST model modified for the domain analytic approach. © Emerald. (From Ref.[3]; p. 303.)
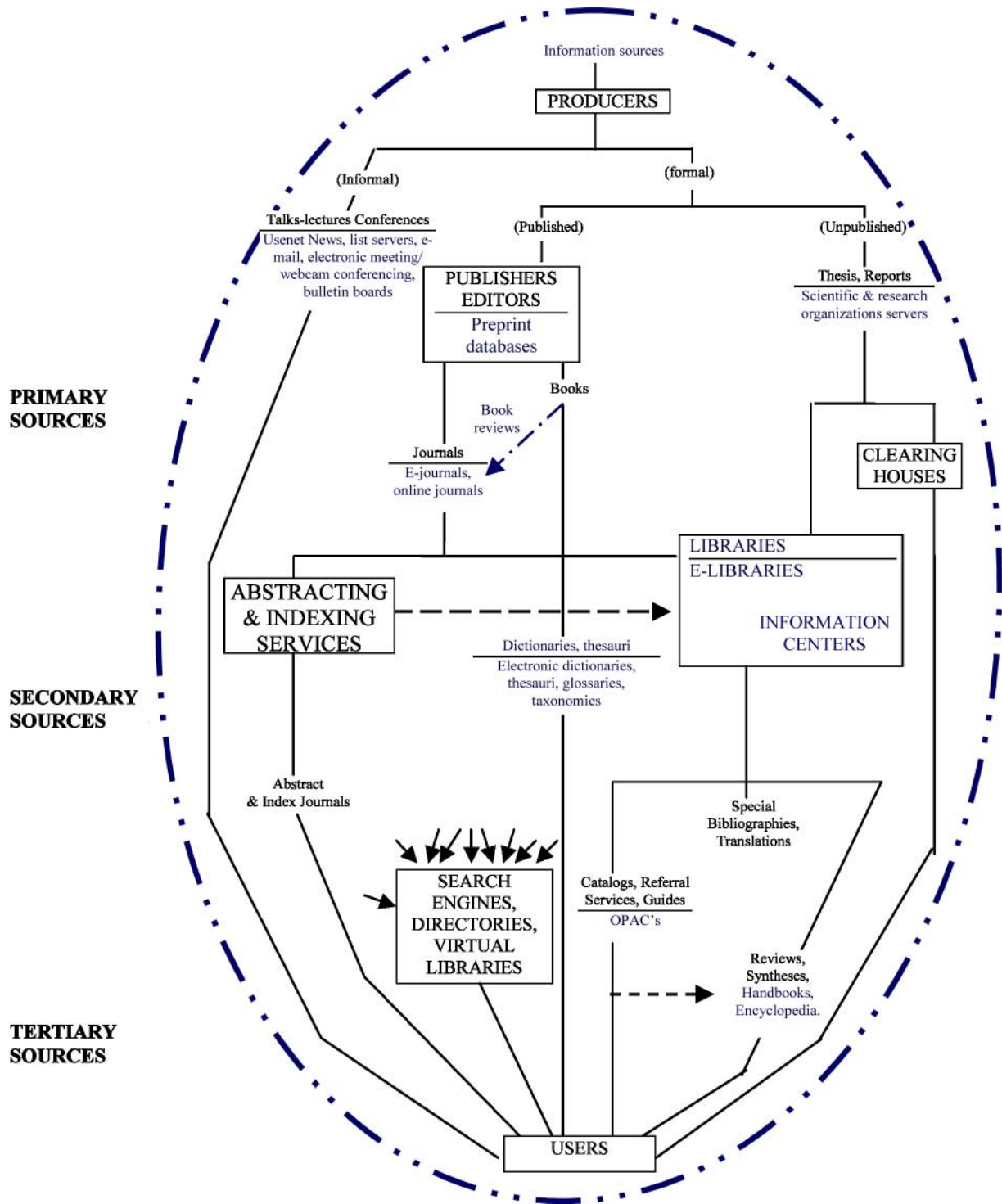
Fig. 3 The revised UNISIST model integrating printed and Internet resources and modified according to the domain analytic approach. (From Ref.[4]; p. 96.) *(View this art in color at www.dekker.com.)*

- Maps and atlases in geography
- Pedigrees and genealogical trees in genealogy
- Codes, bodies of law in law
- Sheets of music in music
- Tests in psychology

Such specific elements are, however, not contained in general versions of the UNISIST model; one must but await the development of domain-specific models.

Some scholarly fields have special information services providing *prepared materials for research*.

While all documents may serve research, some do this in a special way in the form of providing access to reliable source documents. This is especially the case in historical research. Archives and museums are important institutions missing in the original UNISIST model. They are very important in humanities and may also be important in some scientific fields. The study of archives and museums may thus be seen as parts of IS. Often, archives and other institutions reproduce and publish important historical documents in order to make such unique documents much more visible and available to scholars. This is a unique kind of work and should not be overlooked. In the Appendix, we have added "Source Literature" as a new category beyond the primary, secondary, tertiary forms known from the original UNISIST model.

Among the other changes in the revised model that can be mentioned are integration of libraries and information centers into one box and addition of book reviews. The fundamental structure of the original model remains, i.e., the differentiation between primary, secondary, and tertiary sources and services. As mentioned earlier, we have added source literature as a new category. The figure (and the Appendix) shows the addition of thesauri and dictionaries also as a new category. Finally, the classification "Incidental Information" in the Appendix is a primary one for other sources and services that do not fit into other categories but nevertheless are important in scholarly communication.

National and regional substructures

The original UNISIST model was conceived of as one universal international structure. In reality, however, there exist more or less independent and elaborated national or regional information substructures in most domains. If we take psychology as an example, there exist regional journals of psychology, as well as internationally published journals from American and Germany. The German Information System for Psychology is the most elaborate information system outside the U.S.A. in this domain. In this national system, we find a complete system of primary, secondary, and tertiary information services. *The primary information system* in German psychology consists of journals covering all major subfields such as experimental psychology, social psychology, clinical psychology, developmental psychology, etc. proceedings of conferences, books in all fields of psychology, and scholarly treatises. It has under its fold about 150 publishers and producers of tests. *The secondary information system* contains a comprehensive bibliographical database, PSYNDEX, with abstracts and indexing of the German literature in both English and German.

Also, many kinds of dictionaries are being developed. *The tertiary information* system contains comprehensive handbooks, encyclopedias, review journals, etc. The German system also contains specialized information systems developed to communicate psychological knowledge to students and to the general public (e.g., textbooks and the magazine *Psychologie Heute*).

Today, there is increasing political pressure toward internationalization. Such a pressure may motivate German psychologists (among others) to publish in American (or international) journals, and it may leave the impression that other national or regional substructures are obsolete and inappropriate in scientific communication. This issue has been intensively debated and we shall not go deeper into this issue here.[f] As long as regional systems exist, they should be taken seriously by information scientists and should be reflected in our modeling and bibliometric studies of domains and the coverage of databases. The concept of regional subsystems is related to philosophical and sociological question about the nature of different traditions in science and scholarship.

Paradigmatic differences

In addition to disciplinary and regional/geographical differences, each domain will have—more or less noticeably—variations in its information system that are due to paradigmatic differences between the actors in the field. In psychology, for example, there is almost a complete information system for the psychoanalytic approach to psychology, consisting of primary journals, specific organizations, indexes and abstracts journals, encyclopedias and terminological works, etc. This system is not just relatively independent of the "general" psychological system; it also has attributes that reflect the special nature of that field. Thus, psychoanalytic literature is more related to the humanities compared to the literature of behaviorism, cognitivism, and neuroscience. This is reflected in the tendency to organize the literature on authorship basis (Sigmund Freud's works, C.G. Jung's works, Melanie Klein's works, etc). It is also reflected by the relative dominance of books compared to articles and by the rhetorical structure of the texts.

Although such paradigmatic tendencies may be weak, it is our opinion that they always tend to exist to some extent. One could say that in a given (sub)

---

[f]Michael Brittain[5] wrote "The claims of the social sciences to be of universal interest, value, and use are challenged. Citation data show that there is not a free flow of information across language and national boundaries," and he considers the implication for information services. This is still an important issue.

culture or domain, there will almost always exist different "paradigms" for how practices should be changed and how the discipline should be defined and further developed. Any change in practices implies a need to change the documents, the symbolic systems, and the concepts that support the existing practice. The given concepts and documents will always serve certain policies and practices better than other concepts and documents. This is the case whether or not people are aware of this relationship. Domains are dynamic, and an important factor is the theoretical and epistemological development. When paradigms change, the whole model of the domain (as reflected in the UNISIST model) has to change too.

Different concepts, documents, ways of cooperation, etc. are simply better suited for certain "paradigms" than for other paradigms. This is why there is always a more or less latent tendency to develop separate information systems for separate views in any field. Only some disciplines (e.g., psychology and economics) have, however, specific journals specially devoted to different views. The epistemological issues may, however, turn out to be the most important dynamics underlying any information system.

Concern 2: Internet-Based Communication

As mentioned previously, Internet services have been displayed as a separate entity at the right in Fig. 2. The dynamics in scientific and scholarly communication caused by the technological changes from print to electronic media are of course very important. As stated by Smith,[6] the flow of computer mediated communication may cause a modification of the traditional categorization of documentary units. Some of the units may broaden or even modify their categorization in document type (primary, secondary, and tertiary literature). Also, concepts such as documents and publications may have to be redefined. It is, however, with some modifications still possible to classify information sources according to the UNISIST model.

How the growing impact of Internet based communication channels has changed the flow of scientific communication since the creation of the UNISIST model has been subject to research. Among others, Kling et al.,[7–8] Hurd,[9–10] Meadows[11] and Russel[12] have proposed or discussed models for communication in the digital age. However, none of the suggested new models can replace the UNISIST model. Most descriptions of the documentary and organizational units on the Internet emerge from a classification of hardware or software technicalities, rather than that by functional communication parameters as used in the UNISIST model.

As early as 1978, Lancaster suggested that "the distinction between formal and informal communi-

cation will be much less clear in an all-electronic environment."[13, p. 113–114] Though this is true for many of the new hybrid forms, most people would agree that a distinction between formal and informal is possible in everyday Internet based communication.

Informal communication

Consistent with contemporary use of the informal communication channels, the most significant informal documentary units on the Internet are:

1. Email
2. List servers, which are discussion groups or interest groups that distribute messages via mailing lists. Electronic conferences or newsletters are usually listserv mediated.
3. Usenet News, which is a collective term for thousands of newsgroups or discussion groups. Usenet News is managed centrally without the use of email in contrary to List servers. The messages or articles are most often cumulated and archived at least for a while. In most cases, this group includes bulletin boards, which are now rarely used on their own but rather as a feature among others in newsgroups. Thus, Lancaster[13, p.130] defined bulletin boards as a "public space to permit messages to be entered and made accessible without restriction to all users of the system."
4. Electronic meeting or webcam conferencing.

Each of these informal communication channels on the Internet may be located by the user through either (1) various search engines, including meta search tools such as Metacrawler that allows one to access several search engines from one place, or (2) diverse types of virtual libraries. However, informal communication channels often become known to the users in the course of serendipity, general browsing, or interpersonal contacts.

Generally speaking, the Internet mediates a less selective number of informal communication channels than the preceding nonelectronic ones. Information once available only through the professional grapevine is now found on personal or institutional web pages,[14, p. 274] this is why the Internet is believed to have a positive effect on the development of invisible colleges in the otherwise stratified scientific community. Harnad[15] has argued that when (informal) manuscripts and feedback are exchanged through the network, scholarship can progress at a speed more similar to that of natural thought and speech. Because of the ease of using the informal Internet based communication channels, the path from the producer to the user and vice versa is more freely and quickly accessed and less troublesome.

Formal communication

The increasing number of computer literate scholars through the 1990s, among other things, brought about the transition of the Internet from a predominantly informal communication channel to a significant formal communication channel.

The most significant formal documentary units on the Internet are:

1. *E-journals and online journals*: The true concept of the "electronic journal" or simply e-journal, as opposed to the "online journal," requires that the material be produced and stored only in electronic form.[16] Online journals on the other hand are electronic spin-offs from paper journals.
2. *Preprints*: Though some redaction has often occurred, preprints are documents distributed before the actual publication and perhaps before the peer-reviewing process is completed. Preprints are often considered a part of gray literature, but in recent years the emergence of preprint databases on the Internet has offered the means to gain access to this document type. However, not all disciplines have preprint databases; therefore depending on the knowledge domain is considered, preprints may or may not be considered gray.[g]
3. Gray or unpublished literature such as theses, reports, etc. mostly found on scientific and research organizational servers.

Real, operational *e-journals* are essentially a phenomenon of the late 1990s. The first peer-reviewed electronic, full-text e-journal including graphics was *Online Journal of Clinical Trials*, which began publication only in 1992.[17] Research completed in 1996 showed that some e-journals actually have a high impact factor. However, as Harter points out, few articles are published. "Indeed, e-journals cannot have a major impact on the advancement of knowledge until they publish many more articles annually than they do [in 1996], while maintaining the apparent high quality of their articles."[18, p.155]

Weller[19] indicates that the peer-review process of e-journals is generally similar to the traditional process found in paper-based journals. However, new models of editorial peer review have been suggested, and some are being experimented with, which either alter or

eliminate the traditional model of peer review. It is most important that any new model maintains the integrity of science and scholarly communication and yet implements the new emerging electronic environment and the need for decreasing turn-around time. As an example, Weller mentions that in the electronic environment, there is a need to re-examine the anonymity of reviewers.

Regarding the development of *preprint databases*, the Internet as a new media has played a vital role. Due to the scholar community's discontent with publishing delays and distribution problems with paper journals, Paul Ginsparg created Los Alamos ePrint archive in 1991. Arthur Smith who sees journals merely as an overlay on preprint databases describes the powerful position of ePrint archives. "The tension concerning responsibility for public distribution and communication of new work has been resolved in favor of the electronic preprint databases. Traditional journals still have some role in communication, providing archival material and interlinking, but they no longer form the primary communication medium at either the formal or the public levels."[6] As can be seen, the Internet has changed the perception and use of this document type at least in some science domains where rapid dissemination is required. However, further research is needed to identify and explain differences between domains on this matter.

If the development continues in line with Arthur Smith's ideas for the future, it could be argued that journals (electronic or paper-based) should be positioned as a secondary source instead of its present location among the primary sources. Smith[6] argues that the main purpose of the typical journal will be storage, and as a sign of formal confirmation the preprints will adopt the traditional journals' communicative role. Whether or not this will come about depends on various conditions. The traditional position of journals is therefore maintained in the domain general model of Internet based scientific and technical information.

Gray literature

The Fourth International Conference on Grey Literature defined gray literature as "That which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers."[20] Gray literature on the Internet is, if possible to locate, very easily accessed compared with non-Internet based alternatives. The Internet has created an opportunity to make gray literature publicly available without the expenses of traditional publication.

Several organizations, associations, and information systems such as the *European Association for Grey Literature in Europe* (EAGLE), *Information*

---

[g]The widely adopted practice of authors making copies of their papers available on their personal/institutional web sites should also be mentioned. In some domains, notably computer science, this has become the norm. It may be seen as a modern form of offprint.

for Grey Literature in Europe (SIGLE), and the *British Library Document Supply Centre* (DSC) are making special efforts to raise awareness of and provide access to gray literature such as reports, theses, translations, noncommercial conferences, and official (government) material. Several bibliographies devoted to gray literature can be found on the Internet. The problems concerning gray literature are therefore being addressed on international level but may still demand the searcher to be rather persistent.

Formal communication channels

The formal documentary units (in particular journal articles) may be directly accessed if their address is known. Otherwise, the units on the Internet may reach the user through, for example:

1. Preprint databases.[h]
2. Bibliographic or full text databases.[i]
3. Scientific and research organizations servers.
4. Publisher web sites.
5. Virtual libraries.
6. Search engines and meta search tools.

As with the role played by preprints, Smith[6] also anticipates changes in the function of preprint databases. Preprint databases are expected to become responsible for public distribution and communication of new works. This means that the user needs to have great domain specific knowledge or expertise since the content of these databases is not necessarily finalized by the peer-review process. For this reason, Smith presumes review papers to take on a more prominent role in providing guidance to the literature for those not familiar enough with the domain to deal with the nonreviewed preprint literature.

The majority of bibliographic or full-text databases are available via the Internet (given a password), but many databases have yet to make value addition to their services further to this media. The terms electronic libraries (e-libraries), digital libraries, and virtual libraries are often used rather inconsistently and

several different definitions are seen in the literature and on the Internet.

The preservation of paper based scientific communication is a part of the secondary organizational units in the original UNISIST model (e.g., libraries' copyright deposits). Several archival initiatives are seen on the Internet. The Internet archive (http://www.archive.org) is as an example of the preservation of former versions of web sites. Some countries (e.g., Denmark since 1998) have copyright deposits for some web documents, but coverage is selective and probably unsatisfactory. Also more subject specific archives exist.

Publishers' web sites are increasingly providing access to publications in addition to more traditional information such as subscription prizes, contributor instructions, and review policies. The specific searching and browsing facilities on these sites vary. The users' opportunities to view, print, or request documents differ as do their requirement for payment. The publishers' traditional role connected to selection, production, and distribution of the primary sources is increasingly supplemented with new roles concerned with storage and dissemination. Some publishers are beginning to utilize cross-referencing or reference linking as a browsing option. *CrossRef* is a collaborative reference linking service that functions as a sort of digital switchboard. It holds no full text content, but rather effects linkages through *Digital Object Identifiers* (DOI) that are tagged to article metadata supplied by the participating publishers. The result is a linking system through which a researcher can click on a reference in a journal and access the cited article.

These facilities, however, are only available to the users who subscribe to the various publications. A researcher clicking on a CrossRef link will be automatically connected to a page on the publisher's web site showing a full bibliographical citation of the article, and, in most cases, the abstract as well. Subscribers are generally authenticated for full text access, and nonsubscribed users presented with other options for access (such as subscription, document delivery, or pay-per-view). Researchers in library environments may find that CrossRef links redirect to local holdings. This development suggests that the publishers are taking up the traditional secondary sources such as indexing and abstracting services, libraries, and information centers. If cross-referencing between the diverse publishers becomes standard, a new and potentially powerful information-searching tool may rise.

Aids such as domain specific dictionaries, glossaries, taxonomies, and thesauri of various quality and coverage can be found on the Internet, mostly for free or as value-added service connected to fee based databases, virtual libraries, or clearinghouses. These secondary sources are both effectively and efficiently utilized on

---

[h]In the literature, these databases are not referred to as clearinghouses as found in the original UNISIST model. On the Internet, the term Clearinghouse seems to denote some kind of annotated directory or resource guide (See for example the Argus Clearinghouse at http://www.clearinghouse.net/mission.html for further information). Preprint databases primarily flourish within the science domains such as Los Alamos ePrint, LANL preprint database, and SPIRES.

[i]Representing both commercial (First Search, DIALOG, STN, Lexis-Nexis) and noncommercial databases available on the Internet (OPACs also termed electronic libraries or e-libraries).

the Internet compared to traditional paper versions. However, the various sources must be individually assessed with regard to quality.

The documentary units of the Internet can also be reached by search engines or diverse meta search tools. Though very often helpful, these retrieval algorithms or search engines typically suffer from a lack of semantic control (e.g., with synonyms, homonyms, broader and narrower terms, etc. as known in traditional metadata systems). Although they almost seem like miracles, there are still problems that they cannot tackle and for which further research and competing alternatives are necessary.

CONCLUSIONS

The UNISIST model is a fruitful model of scientific communication that help conceptualize IS in a perspective that is of great heuristic value and also fruitful for further empirical investigations. The model is also useful for practical information work, e.g., the construction of information guides.

Today we do not even have a description or model of the communicative system according to the UNISIST model *for even one single discipline* based on empirical studies. There is a big need to study different domains according to the model (thus also contribute to the further development of the model).

We also need to consider some basic issues in the model further. For example, to what degree do the primary, secondary, tertiary, source producing and intermediating level (text books and mass media) function as relatively independent systems? Do they have relatively independent groups of professionals? Do they have specific guidelines and norms? Do they have specific channels for publication (output)? Do they have specific educational programs and information input channels? What internal and external factors determine the structure of scientific communication systems?

Each point in the model as well as a large number of relations are also in need of more research. It is our hope that this model may stimulate further interest in scholarly communication and in documents and thus provide library and IS a much-needed theoretical inspiration.

APPENDIX: TYPOLOGY OF DOCUMENTS

I. Primary Literature

(Primary literature is the researchers' and knowledge producers' primary medium for claiming original findings, theoretical analysis, empirical data, etc.)

- Monographs 1 (as long as they communicate original findings)
- Journal articles and articles in edited books (as long as they communicate original findings)
- Critical analyses
- Conference presentations
- Gray literature including dissertations, treatises, master theses,
- Reports, various kinds of official publications and governmental publications
- Patents
- Standards

Ia. Source Literature

(Source literature is either literature produced in order to supply researchers with information (e.g., translation journals) or information produced to other purposes than research, but used as information by researchers (e.g., music and fiction). Primary literature (and anything else) serves of course as information sources, while source literature is negatively defined as not being primary, secondary, tertiary, accidental, or popular literature)

- Facsimiles
- Transcriptions
- Source editions, scientific editions, and standard editions. [Model: The works of Søren Kierkegaard . . .].
- Laws, court findings
- Music
- Data archives
- Statistical documents, tabular documents 1 (reporting original data)
- Translations (only qua translations; the translated work is, for example, primary literature)
- Product information/"trade literature"
- (Not applicable: Sourcebooks)

II. Secondary Literature / Bibliographical Literature

(Literature that registers, describes, and organizes the primary literature as well as the other categories (including the secondary literature itself). Secondary information systems are the core focus of the library, documentation, and IS profession. Bibliography is a discipline that studies this area).

- Subject bibliographies and bibliographical databases
- Abstract journals
- Indexes
- Citation indexes

- Current Contents
- Bibliographical guides, metabibliographies
- Bio-bibliographies/author-encyclopedias (including auto-bibliographies on personal web pages)
- Source inventories
- Catalogs

IIa. Dictionaries and Thesauri

(Dictionaries are the focus of the linguistic subdiscipline lexicography. Thesauri are a kind of dictionaries that has mostly been studied and developed in relation to bibliographical databases)

- Historical/etymological dictionaries
- Translation dictionaries
- Conceptual dictionaries and thesauri

III. Tertiary Literature/Review Literature/"Outlines"

(Literature summarizing and synthesizing knowledge in the primary literature)

- Handbooks
- (Textbooks)
- Monographs 2 (synthesizing existing literature without providing new, independent knowledge)
- Review articles (do)
- Scientific encyclopedias (general encyclopedias are normally considered as popular literature)
- Short, indicative reviews
- Chronological surveys
- Data handbooks, tabular documents 2 (synthesizing original statistical sources)

IV. Incidental Information

(Information about tools (including computers and software), developments in the job market, in the discipline/domain, etc. as long as such information cannot be seen as part of the domains' regular knowledge production.)

- Biographical documents
- Directories
- Conference calendars
- Lists of archives
- Directory to grants, scholarships, etc.
- Yearbooks (annual reports)
- Newsletters
- Personal homepages

V. Popular Literature

(Export of knowledge produced in a domain to the general public, other domains, or students.)

- Textbooks
- Magazines
- Newspapers (e.g., science journalism)
- Popular books (including general encyclopedias)
- Faction, science fiction
- Mass media, multimedia presentations, etc.

REFERENCES

1. UNISIST. *Study Report on the Feasibility of a World Science Information System*, By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions, Paris, UNESCO, 1971.
2. Hjørland, B. Documents, memory institutions, and information science. J. Doc. **2000**, *56* (1), 27–41.
3. Fjordback Søndergaard, T.; Andersen, J.; Hjørland, B. Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. J. Doc. **2003**, *59* (3), 278–320. Available at: http://www.db.dk/bh/UNISIST.pdf.
4. Hjørland, B. Fundamentals of knowledge organization. Knowl. Org. **2003**, *30* (2), 87–111.
5. Brittain, J.M. Internationality of the social sciences: implications for information transfer. J. Am. Soc. Inf. Sci. **1984**, *35* (1), 11–18.
6. Smith, A. The journal as an overlay on preprint databases. Talk prepared for ALPSP, 9 April 1999. Available at:http://ridge.aps.org/APSMITH/ALPSP/talk.html (accessed July 24, 2004).
7. Kling, R.; Callahan, E. Electronic journals, the internet, and scholarly communication. Annu. Rev. Inf. Sci. Technol. **2003**, *37*, 127–177.
8. Kling, R.; McKim, G.; King, A. A bit more to it: scholarly communication forums as socio-technical interaction networks. J. Am. Soc. Inf. Sci. Technol. **2003**, *54* (1), 47–67.
9. Hurd, J.M. The transformation of scientific communication: a model for 2020. J. Am. Soc. Inf. Sci. **2000**, *51* (14), 1279–1283.
10. Hurd, J.M. Models of scientific communication system. In *From Print to Electronic: The Transformation of Scientific Communication;* Crawford, S.Y., Hurd, J.M., Weller, A.C.; Medford: ASIS, 1996.
11. Meadows, A.J. *Communicating Research*; Academic Press, 1998.

12. Russel, J.M. Scientific Communication at the Beginning of the Twenty-First Century, ISSJ 168, UNESCO, 2001; 271–282.
13. Lancaster, F.W. *Indexing and Abstracting in Theory and Practice*, 2nd Ed.; Library Association: London, 1998.
14. Russel, J.M. *Scientific Communication at the Beginning of the Twenty-first Century*. UNESCO, Paris, 2001; 271–282.
15. Harnad, S. Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge. Public Access Comput. Syst. Rev. **1991**, *2* (1), 25–38.
16. Piternick, A.B. Attempts to find alternatives to the scientific journal: a brief review. J. Acad. Librariansh. **1989**, *15* (5), 260–266.
17. Keyhani, A. The online journal of current clinical trials: an innovation in electronic journal publishing. Database **1993**, *16* (1), 14–23.
18. Harter, S.P. Scholarly communication and electronic journals: an impact study. J. Am. Soc. Inf. Sci. **1998**, *49* (6), 507–516.
19. Weller, A.C. Editorial peer review for electronic journals: current issues and emerging models. J. Am. Soc. Inf. Sci. **2000**, *51* (14), 1328–1333.
20. Grey Literature Network Service. GL 1999 Conference Program. Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, GreyNet, Grey Literature Network Service, Washington, DC, 1999, Oct 4–5. Available at: http://www.nyam.org/library/greywhat.shtml (accessed July 24, 2004).