Does the Diversification of the Information Environment Turn Scientists into Information Illiterates?

Dr Rafael Ball
Forschungszentrum Jülich GmbH,
Central Library
Germany

1. Introduction
2. Factors influencing scientists' information behaviour
   2.1 The development of literature production and the presentation of library holdings
   2.2 Media diversity
   2.3 Information technology
3. Diversification of the information environment
4. Future trends in information retrieval behaviour and the influence of library services

1. Introduction

Information is the raw material of science and research. However, this is not just true of these fields. Access to information is not only the prerequisite for meaningful and successful scientific work but at the same time is also a basic democratic right for everybody. Those who are not granted access to information are robbed of their opportunities for personal and professional development and become second-class citizens.

Access and opportunities of access to information are multifarious, complex and dependent on a number of different factors. Access to information as a matter of principle does not by any means ensure that the required data are actually found. Today the available and usable search and retrieval opportunities decide on the success or failure of obtaining the required information. Information literacy is the most important prerequisite for dealing confidently and successfully with knowledge and information. The following key points already became apparent in the seventies in definitions of information literacy[1]:

- information is used in solving problems
- the application of information in the work
- special techniques and skills are needed for using the information tools and sources
- the utilization of information has to be efficient and effective.

The developments of the eighties with all their online databases, telecommunication services, electronic mail, abstracting and indexing services, library networks and alerting services extended the concept of information literacy to include the term 'computer literacy'. Although the meaning of the two terms is not identical, a command of the technical aids became more and more important for the skill of information literacy. At that time there were already fears that a divide would arise in society between information literates and information illiterates. Definitions of information literacy in the eighties emphasize: "An integrated set of skills is included as one of the characteristics of information literacy. These skills are identified as research strategy and evaluation. Information literacy extends beyond mere locating of information to include understanding and evaluating the information"[2].

This feared gulf between information literates and illiterates may also be postulated within an important group of information users, namely scientists. Whoever is not capable of keeping up with developments can rapidly degenerate into an information illiterate and thus become scientifically marginalized. In the following it will be explained why this can happen especially easily in the sciences and why it is so dangerous.

In science there are above all three determinants which define scientists' information behaviour and the success of searches:

1. the increase in the production of scientific (and general) literature
2. the change of media and the emergence of media diversity
3. the development of information technology.

The above-mentioned factors are basically responsible for the fact that the quantitative and qualitative efforts in searching for scientific literature and information have increased so significantly in the course of the last 50 years.

2. Factors influencing scientists' information behaviour

Access opportunities and adequate use of all types of literature and information are essential for scientific work. Information has long been a key resource for industry, science and research. More than ever before, above all scientific and technical research is becoming less and less

non-application-oriented and is already frequently encumbered by the necessity of practical application and the constraint of applying research results in industrially exploitable products or patents. The supply of literature and information must be correspondingly efficient, fast and selective.

This presentation proceeds from the hypothesis that although the increase in the production of literature, the emergence of media diversity and the development of information technology provide the scientist with an ever increasing volume of information with improved cataloguing of holdings in archives and libraries, the technical and intellectual efforts for access, searches and retrieval have already reached or exceeded the boundary of that which is reasonable. At the same time there are still many scientists who do not want to or cannot fully use the services offered by libraries. More than 50 years ago a physicist already regretted this at an ASLIB conference in London:"At the same time the research worker has never been able to make the fullest use of the (...) library, largely because of his ignorance of the possibilities of recent developments in library technique. Many research workers do their own library work, often in an antiquated way, or discover for themselves many of the techniques, already long known in the library world."[3]

2.1 The development of literature production and presentation of library holdings

Science must publish and publication decides on the success of research. With the increase in knowledge, literature production exploded in the course of the 19th century. In addition to the number of scientists, the number of scientific journals also increased continuously in the 19th century. Since the early 18th century, this figure has increased by a factor of 10 every 50 years. Today, the level of 150,000 different journal titles has already been exceeded. The flow of knowledge as well as the publishing and exchange of scientific results have become an indispensable part of the scientific process itself.

In the early fifties of the 20th century there were a number of publications estimating the literature development of past centuries. For example, it is assumed that in the past 500 years 12 million book titles have been published, 10 million of which appeared after 1850. In comparison with other growth rates, literature production is always above average. The population of the USA doubled in the first half of the 20th century, while the total book production rose by a factor of seven [4].

Book production in Germany, for example, rose from 30,000 titles in 1913 to more than 70,000 in 1994[5]. UNESCO reports that in 1980 more than 24,000 titles were published annually in 44 countries of the world[6]. In 1999 the number of scientific journal titles worldwide amounted to about 160,000, of which roughly 11,000 were already available as an online version of the print edition. Added to this are about 3500 journals on CD-ROM[7].

In keeping with the increase in knowledge, abstracting journals already became established in the 19th century in order to facilitate access to knowledge or indeed to make such access possible at all. The number of these journals also increased enormously[8]. Nevertheless, studies from the early 20th century are known indicating that the abstracting journals were even then no longer able to adequately screen and evaluate the increasing volume of knowledge so that a large proportion of knowledge remained hidden since it was no longer selectively searchable and retrievable: "We are forced to the conclusion that less than half the useful papers are noticed in the current abstracting and indexing periodicals"[9].

The volume of literature, scientific work with literature and information, as well as the presentation of library holdings are therefore greatly dependent on each other. Up to the time of the Enlightenment, with a manageable production of books, it was a matter of course that the literature should be systematically shelved in the libraries. Scientists therefore naturally had access according to their knowledge of the classification system of their own subject area. Special knowledge and skills in using a library were not necessary. The use of a library and the literature in it and scientific work itself were identical. Systematic shelving as an unquestioned presentation of holdings would only gradually disappear in Germany at the end of the 19th century as a consequence of the enormous growth in library holdings.

With Georg Leyh's innovative article[10] a mechanical sequential location of holdings in closed stacks no longer accessible to scientists became the standard presentation for holdings at large scientific general libraries. There may have been objective constraints for this, nevertheless this was no compensation for the fact that a majority of the scientific (and also the nonacademic) users felt hampered in their use of libraries and the acceptance of the central university libraries was therefore perceptively reduced by the introduction of closed-access collections. Many scientists regarded the establishment of institute or department libraries as a suitable replacement, with the manageable holdings of their own subject area in the familiar and congenial systematic shelving system.

It was, however, only possible to use the inaccessible holdings in the central libraries by developing suitable indexing instruments. This gave rise to catalogues which as meta-information had to identify all the literature in the stacks even if uncoupled from the actual location. A multitude of formal and subject classification features now had to be used to create more or less reliable access to the literature which was now no longer directly accessible. Eppelsheimer's catalogue system should be mentioned as a particularly successful example of the possible types of catalogue design. This is a subject-oriented catalogue system displaying subject, geographical and topographical, and biographical facets and enabling a variety of classifications from the catalogue. Nevertheless, in spite of all efforts to create an adequate cataloguing system it soon became clear that "shelving the books according to serial number ..." would lead "necessarily to widening the gap between library and user, above all the qualified user"[11]. And this gap was not only apparent in a shift away from central libraries but also primarily meant greater efforts for access and retrieval of the required literature.

The development of catalogues (which had become necessary for whatever reasons) represented a first and severe obstacle to access to the required literature and information for scientists. Direct usage and direct access to literature on the shelf was preceded by the step of searching in the catalogue. Although the further differentiation of cataloguing systems enabled selective and detailed access to information holdings, nevertheless this improvement had to be bought at the price of precise knowledge of the catalogue structure. Only those scientists familiar with the cataloguing system of "their" library and who understood the structures were able to use the tools efficiently.

Added to this was the wide range of catalogues used for indexing the materials in Germany alone. Far removed from a uniform classification and a uniform catalogue, users were forced to familiarize themselves with the respective cataloguing tools of the individual libraries in order to be able to make appropriate use of other libraries' holdings. But this was only to be the beginning.

2.2 Media diversity

The development of the media proceeded in a manner comparable to the increase of knowledge and literature production. For centuries (indeed for almost two millennia) the medium for disseminating written information had remained unchanged. Only in the past fifty years (and particularly in the past twenty years) has an acceleration of media development taken place which is no longer comparable with earlier developments. Ever since Johannes Gutenberg invented printing by movable type in 1455 (and many centuries before that in manuscripts) the book in codex form had been the exclusive medium for disseminating written information. The fixation of knowledge as the "writing" of books, the production and dissemination of books and library handling procedures had been oriented to printed paper for centuries (in the form of books and only later as journals). Scientific literature research was therefore directed towards printed works. Access to stored knowledge was gained by using library holdings and their cataloguing tools. This meta-information was also initially exclusively paper-based. Holdings and catalogues, whether handwritten or printed, were decisive for scientists in this form for centuries. The demands made on the users of this paper-based information were simple and clear. They had to be able to read and have some knowledge of the structure of the catalogues used and their classification system. For nearly two millennia this was sufficient to use information in a library.

Dramatically increasing literature holdings, on the one hand, and concern about the decay of paper material, initially for newspapers and journals, on the other hand, confronted librarians and scientists with the new medium of "microfiche" and "microcard". These media represent a photographic equivalent of the literature printed on paper and its records (catalogues) and are thus not a qualitatively different method from printed paper. The advantages of microfiche, microfilm and microcard are their modest space requirements and the good durability of the materials. Microfilms have been used in libraries since about 1920, but only in 1939 was the flat microform technique (now known as microfiche) taken up by libraries[12].

Even in the middle of the 20th century, the solution to the flood of information (which had already been noticed at that time) and the problem of mass was tackled by reducing the primary, secondary and meta-information (catalogues) to microfiche or microfilm[13]. Indeed, micromaterials were and are space-saving materials and help to deal economically with the limited stack capacity in libraries. It is fairly inexpensive to dispatch literature on micromaterials and in comparison to digital storage media the durability of the plastic material is also exemplary. However, micromaterials have not contributed to coping with the flood of information. They are difficult to handle and not very user-friendly. A technical device is needed as a reading aid and readers must at least have a rough idea of how micromaterials are structured so that they do not become submerged in the apparently endless jumble of data on the microfiches in the viewer. However, in principle, micromaterials do not represent any increased effort for literature searches on the part of scientists so that until well into the eighties they were a popular medium for storing metadata.

A real paradigm shift in the media was caused by the digitization of all types of information. The revolutionary aspect was not only the way in which information was stored and made available. For the scientific user, the opportunities of literature and information searches and utilization changed radically. In the library sector, the digital world was initially not designed for the use of full texts. In the early sixties, the first online information was available in libraries and comprised online databases with metadata or special fact databases accessed from a central host via a dedicated line. At the beginning of the digital era, the demands made on the software and hardware, on the one hand, and on knowledge of the retrieval languages and the thesauruses used, on the other hand, were very high. For these reasons, use of and access to these databases were exclusively reserved for information professionals in the libraries and documentation centres. At that time, scientists still did not have direct access to these systems. They could only access the contents with the assistance of specialists.

From then on there was no stopping the triumphant progress of electronic and digital information. This headlong development had nothing to do with the now apparently undesirable medium of paper but rather with the very limited search and retrieval possibilities in printed books and journals. As long ago as 1965, Licklider pointed out in his widely noted study on the future of libraries that the problem was not replacing paper as such but rather its limited retrievability: "As a medium for the display of information the printed page is superb ... When printed pages were bound together to make books or journals, many of the display features of the individual pages are diminished or destroyed ... In fulfilling the storage function, they are only fair. With respect to retrievability they are poor. And when it comes to organizing the body of knowledge, or even to indexing and abstracting it, books by themselves make no active contribution at all."[14]

The availability of digital data opens up a whole new range of search and retrieval possibilities, whose new quality has nothing in common with previous cataloguing instruments and their mechanisms. Printed library catalogues are, for example, always only one-dimensional and cannot be logically linked. This is information of a formal type. In contrast, although the almost unlimited retrieval possibilities of electronic catalogues permit qualitatively new access to library holdings, on the other hand, handling has become considerably more complicated for scientists as has the updating work for library staff. Searches in an electronic database system (search and retrieval) is a complex process requiring, in addition to knowledge of the retrieval software, command language and the thesauruses used, the ability to construct a meaningful search strategy[15]. A database can in fact only be used optimally if indexing and searching is kept in the same hands (which is seldom the case). For example, in addition to knowledge of all system functions, scientists would also require creativity and subject-related imagination for a meaningful and successful search in the database of a library catalogue in order to achieve relevance (precision) and pertinence (recall) in their search results.

It makes no difference whether the search is performed in a CD-ROM database, a medium that began its inexorable march through the libraries in 1985, or in an online database. Open and closed discussion groups, scientific forums and other medial forms on the Internet complement the extensive range of information media available to scientists today. The multiplication factor for generally accessible information databases is now estimated to be 20 % per year[16].

The final step to date in the medial development of information tools for science is multimedia technology. Completely new access to information is created by the interactive utilization of image, sound and text documents. The interactions arising here and the influence of these new medial phenomena on science and scientific research, as well as on the process of acquiring knowledge as such, has not yet been systematically studied at all. The realization is just beginning to dawn that the existence of dynamic documents will basically revolutionize scientific results and their output in the form of scientific publications. The acquisition and processing of knowledge as well as the dissemination and discussion of ideas will, so to speak, enter into a "real-time relationship". Stationariness and immediacy will disintegrate into pure dynamics. It remains to be seen whether against this background an updating of data at fixed points in time will still make any sense.

2.3 Information technology

The development of information technology and medial changes are closely correlated. This brief outline therefore draws strongly on the previous chapter and reflects the technical development in the epochs of the media.

The use of modern communication technology for the primary library sector dates back to the early sixties. Data processing by means of automated mechanical methods such as punched card or punched tape processing will not be discussed here. Digital information has been available in the library sector in an online form since the early sixties. The up-to-dateness of the data, immediate access even from decentralized locations and improved search strategies made this technology interesting for libraries. Online information was, however, not initially designed for fulltext; only fact databases or metadata were available. The utilization of this digital information made high demands on hardware and software equipment. The special software was initially run on mainframe computers. Elaborate log-on procedures were a matter of course, as well as the requirement for detailed knowledge of the very heterogeneous retrieval languages. At that time responsibility for online searches was solely in the hands of central specialists.

The development of the personal computer was to represent a turning point. Microcomputers have been available since 1971, that is to say computers with a single chip as the processor. In 1979 it was possible to store about one million bits, about 40 book pages, on one memory chip. This did not initially have any impact on scientific literature searches, but it was the basis for the application of mass memories in every single office and at every workplace.

The real breakthrough for the universal application of digital data in science and libraries, however, was first brought about by the introduction of CD-ROM technology in 1985. Similarly to online databases, this medium was also not originally intended for fulltext. CD-ROMs in libraries primarily served to store and utilize metadata, for example, bibliographic references, to accommodate digitized catalogues and other secondary information. At the beginning, difficulties in operating CD-ROMs were just as great as in using online databases so that they were also initially only used and operated by library staff. However, the responsibility for searches was soon delegated to the users, who then had to learn the retrieval languages and technical handling themselves. In contrast to online searches, the retrieval languages usually had a simpler structure and, above all, could be reproduced as often as required without any additional costs for connection times and downloads.

A revolutionary innovation in communication technology was the establishment of network technology. The mere presence of digitized information on new media with corresponding retrieval techniques alone did not amount to an innovative technology. The transition from the purely stand-alone solution of CD-ROMs in the early years to network technology permitted the linked collaboration of the most varied fields of libraries and science. CD-ROM networks therefore led to a "centralization of decentralized access" to the libraries' information products and by means of a central menu made all CD-ROM databases available for the members of a scientific establishment in a decentralized manner. The implementation of this end-user strategy meant that searches were now finally and completely the responsibility of library users, i.e. scientists. Searches are performed in a decentralized manner at the workplace by the scientists themselves. This autonomy was, however, only bought at the price of knowledge of another, even if usually simple, retrieval language (operating software of the CD-ROM network).

Associative hypertext technology, the possibility of structuring and linking all forms of digitized materials such as text, image, sound and film for multimedial applications implemented in the hypertext editor of the World Wide Web can be regarded as the real breakthrough for the Internet providing worldwide, permanent access to continuously updateable and updated information. Simplified search strategies by means of graphical user interfaces give access to worldwide amounts of data, whose extent can neither be reasonably calculated nor qualitatively assessed. Internet technology, however, finally transfers information searches and strategies to the scientists themselves.

The linking of metadata with associated texts and fulltext currently marks the most recent trends in information technology.

3. Diversification of the information environment

Chapter 2 illustrated the factors influencing scientists' information behaviour. The increase in knowledge (and as a consequence the development of literature production), media diversity and the development of information and communications technology are therefore decisive for shaping the information environment.

It was shown that the effort expended by a scientist for literature studies must have increased in the course of the past centuries due to the changed presentation of literature and information and the corresponding reference tools. Starting from systematic shelving which integrated access to information as part of scientific work, through the development of the different cataloguing systems which became necessary when books and journals were stored in stacks, up to the apparently unlimited availability of digital information for some years now in real time via worldwide data networks, access to information has become better and more comprehensive, but also more complex, difficult and time-consuming.

Above all the diversity of the information itself and access to it, as well as that of the corresponding tools, has increased greatly. The complexity and diversity of information types, structures and tools have become continuously greater. The variety of the old catalogues, for example, has not been reduced by the introduction of electronic metadata. In a continuous process of expanding the electronic DP systems in libraries (e.g. OPAC, CD-ROM network, Internet), a number of complex information supply services have emerged, usually not coupled to each other, which when consulted provide the scientist with an immense number of hits and give the impression of a successful and complete check of all the holdings, without really having filtered the information and verified it for relevance.

Whoever enters a university library with old stock today will find, in addition to the open-access shelving of some of the literature, a complex stack collection, access to which must be obtained via hand-written or printed card catalogues of the most various kinds and via an OPAC, with possibly different regulations for borrowing and use, which are not seldom reflected in different catalogues[17]. Access is also available to an extensive range of electronic information consisting of fulltext, fact databases and an extensive range of metadata (for example bibliographic databases and catalogues). All these electronic services are usually available via CD-ROM, and the majority of databases have their own retrieval software.

Moreover, for several years now many libraries have been offering extensive information services on the Internet ranging from bibliographic databases to electronic fulltext. Services involving costs are frequently combined here with services free of charge or already licensed by the library, and often involve a rather cumbersome password procedure. The information environment has now become so complex that the majority of users do not understand it in its entirety. The necessary traditional **"differentiations between subject indexing and the author/title approach"** in the catalogue are complemented by the necessity for a **"media decision"** (on what medium is the information available or desired?), as well as a **"technology decision"** (by means of what technique can the desired information be obtained?) up to **"decisions on formats"** in which digital information is desired or available. If, for example, electronic journals are available, a **"publisher decision"** may be necessary since many electronic information services are made available according to this criterion.

In a letter to the Library Commission, a scientific user of the library at Munich University of Technology asked "Shall we have to be trained to use 8000 databases in the future?" and argued in favour of a uniform, simple and transparent information management system.

The number of "decisions" to be made by users of library services is much too high. In this context, decisions do not mean freedom but rather additional work for users forced on them by the infrastructure and which should not be an end in itself but rather a preparation and complement to research work.

Recently, many of those affected in the fields of publishing, agencies and libraries have recognized the problem of the complexity of the information environment and attempted to counteract this development in the infrastructure fields so detrimental for the efficiency and effectiveness of science. One result is the publishers' or agencies' creation of their own "information environments". Complex electronic services will make comprehensive access to all necessary information and corresponding literature available to the user by one click on the server of the "right" publisher or the "right" agency.

However, these services only apparently package information for the scientist. In fact, these private packages from the respective publishers, agencies and academic societies merely increase the diversity of information services and thus through the introduction of a new "differentiation category" (information producer) (unintentionally) further encourage the compartmentalization of the information supply. The victims of this development are libraries and scientific users. The libraries must now integrate these (generally expensive) "information packages" into the already diverse information supply (and thus frequently create redundancy in some parts). In most cases, access to a single one of these packages is not sufficient for scientists since they have to make use of information from a great many publishers.

4 Future trends in information retrieval behaviour and the influence of library services

The mass and variety of data as well as the quantity of available information is growing, and the complexity and diversity of the reference tools has reached a high level. Against this background, scientists are often occupied too long in selecting the relevant information (and only this is required) and are frequently overstretched.

Information illiteracy is not a pseudoproblem. Access to relevant and precise information involves great effort for the scientist of today and at the same time is associated with uncertainty about its completeness. Information professionals in libraries have the task of easing the burden for scientists.

As information professionals, on the basis of a comprehensive information concept (gathering and processing data) they will play an increasingly more important role in supporting scientists. In order to prevent information illiteracy from becoming a general problem in science, the task of the information specialist is to reduce this too greatly developed diversity of the information environment and to concentrate on central points.

These include:

- meaningful and informative presentation of all kinds of information
- user-oriented focusing of materials
- medium-independent presentation of relevant information
- combination of diverse electronic services such as fulltext, databases, catalogues under one single retrieval interface

- creation of transparency in the case of large numbers of hits and ending the illusion of any possible "exhaustiveness" of electronic search engines and meta-data
- the relativization of holdings lists and accessibility; creation of a realistic perspective for the possible use of media documented worldwide (what is the use of locating a book in the municipal library of Santiago de Chile?)
- elimination of all "fractions" (in media, technologies etc.) for the information materials and their reference tools
- integration of the various services offered by a library such as referencing, lending, acquisition in the form of "one-stop shopping".

The aim of all library efforts must be a simple and easy-to-use information search system which is in itself consistent and exhaustive. The user must be able to obtain an optimum, relevant and precise search result without any previous or detailed knowledge of retrieval software, formats, media or publishers. The introduction of a comprehensive electronic information mangement and information search system must mean the introduction of a program which is easy to use. Extensive training courses and introductions lasting several days are not appropriate in the contemporary world nor are they understood by the users.

This does not in itself reduce the complexity of the various information services but this complexity is then not apparent to the user. As a demanding task for interdisciplinary information professionals, the structuring of information must proceed in the background away from the users. Information structuring should not burden scientists in any way and certainly not be transferred to them in its entirety.

As library users, scientists must be supported in their work and not encumbered so that they will then be satisfied library patrons and not victims of information illiteracy.

References

1 Cf.: Zurkowski, P.G.: The Information Service Environment Relationships and Priorities. Washington D.C., National Commission on Libraries and Information Science. 1974

2 Behrens, S.J.: A conceptual and historical overview of information literacy. In: College and Research Libraries, 55(4), 1994, p. 312

3 Bernal, J.D.: Information Service as an Essential in the Progress of Science. In: Report of the Proceedings of the 20th Conference of Aslib, London, 1945, p. 20

4 Meadows, A.J. (ed.): The Origins of Information Science. London: Taylor Graham, 1987, pp. 12/13

5 Buch und Buchhandel in Zahlen 1995. In: Börsenblatt, No. 54, 1995

6 Lexikon des gesamten Buchwesens. Ed. Severin Corsten. Stuttgart: Hiersemann, Vol. 1, A-Buch, 1987

7 Ulrich's International Periodicals Dictionary 1999

8 In spite of the increase in scientific activities, the expansion of subject fields and the explosive increase in scientific results in the 20th century, this system stood the test of time unchanged up to the sixties. In order to cope with the flood of information, the first indexing and abstracting journals appeared in the early 19th century aiming to facilitate a structured information search for

scientific results. Until the emergence of digital information nothing changed in the basic understanding and the basic conventional publication technology.

9 Bradford, S.C.: Documentation. London: Crosby Lockwood, 1948, p. 121

10 Leyh, G.: Das Dogma von der systematischen Aufstellung. In: Zeitschrift für Bibliothekswesen, 29, 1912, pp. 241-259; 30, 1913, pp. 97-136

11 Braun, H.: Sacherschließung 1978. In: Alsheimer, R. (Ed.): Bestandserschließung und Bibliotheksstruktur. Wiesbaden: Harrosowitz, 1979, p. 108

12 Lexikon des gesamten Buchwesens. Ed. Severin Corsten. Stuttgart: Hiersemann, 1987

13 Cf. Bush, Vannevar: As we may think. In: Atlantic Monthly, 176 (1), 1945, pp. 101-108

14 Licklider, J.C.R.: Libraries of the future. Cambridge: The MIT Press, 1965, pp. 4/5.

15 Meiss, Brigitte: Information Retrieval und Dokumentenmanagement im Multimedia-Zeitalter. Deutsche Gesellschaft für Dokumentation, Frankfurt am Main, 1997, pp. 56-58

16 Meiss, Brigitte: loc. cit., p. 11

17 In large university libraries with old stock a usually only selective retrospective digitization of the catalogues leads to an inacceptable diversity of information structures within the scope of these reference tools alone.