ABSTRACT
        This paper presents the results of an empirical survey
conducted in the winter of 1995-96 among 413 Danish Internet users (business
people, higher education/research and students). Data from a recent Danish
Mediacenter Internet survey (n=2981) are also included. Findings indicate
that women tend to use the Internet to solve well-defined problems, while men
are inclined to serf the net (called "net watching" by serious users). In
terms of information retrieval, female Internet users seem to be more
goal-directed and convergent; in most cases they stop when the information
has been found, while male Internet users have adopted a more
divergent/explorative attitude. In this survey, the average user sent 13.3
messages per week (ranging from 0 to 170) and received 21.3 messages per week
(ranging from 0 to 1000). Survey respondents subscribed to 5.2 electronic
conferences on the average; 59%, however, did not subscribe at all and 24.3%
had limited themselves to fewer than six conferences. Most of the respondents
were using Internet resources for information retrieval, which, in their
opinion has improved the quality of their reference work to some extent; very
few respondents expressed negative expectations in terms of the quality of
Internet information sources. A recent survey among British business
librarians revealed some disadvantages, including a lack of: organized
information, information quality, relevant information, and reliable
information, as well as too much information. The inevitable "virtualization"
of libraries will necessarily include: the replacement of printed matter by
electronic information; increasing use of computer-mediated communications;
and universal access to information. (Contains 19 references.) (AEF)

Looking for the information needle in the Internet haystack

Helge Clausen

State and University Library, Denmark

Abstract: *In the present paper the most important findings from a comprehensive empirical survey among 413 Danish Internet users are presented. The survey was carried out in the winter of 1995-96. Special attention is paid to user patterns and information quality. Differences in expectations among various groups of users (business, higher education/research and students) are analysed and discussed. Comparisons with findings from other relevant surveys are made. On the basis of some of the empirical findings, two models concerning the placement of Internet information resources are discussed.*

Keywords: Internet, information quality, Denmark, research libraries, librarians, information professionals, controlled information network environment, virtuality, World Wide Web

1. Introduction

The information resources accessible through the Internet continue to grow in extent. An increasing number of users are trying to satisfy their information needs by using such resources. Are they getting stones for bread?

According to a recent survey from the Gartner Group (Ref 1) about half of the world's 280 million PCs will be connected to the Internet by the beginning of 1997. That is the optimistic forecast, while a more conservative one — also from the Gartner Group — says 10%. That means that between 28 million and 140 million PCs in a few month's time will become interconnected, and be potential sites for the input and retrieval of all kinds of information. It is thought-provoking, however, that for example 100 million people form less than 2% of the world's population. Even today, a century after the invention of the telephone, about half of mankind still does not have access to it.

The Matrix Information and Directory Services (MIDS) distinguishes between three concentric entities in calculating the number of Internet users (Ref 2): the *Matrix* includes the *Consumer Internet* which includes the *Core Internet*:

- *Core Internet*: 16.9 million users of 7.7 million computers that can *distribute* information by interactive TCP/IP services such as the World Wide Web (WWW) or FTP;

- *Consumer Internet*: 26.4 million users of 10.1 million computers that can *access* information by interactive TCP/IP services such as WWW or FTP;

- *The Matrix*: 39 million users of electronic mail.

The figures are taken from the latest MIDS Internet Demographic survey from October 1995 (issued in February 1996). This way of making up the number of Internet users gives a rather conservative estimate that lies between the optimistic and conservative forecasts of the Gartner Group mentioned above. Further, MIDS has calculated that the Internet has been growing at about 100% per year since 1988.

How can the Internet-based resources be seen? A Danish futurologist has pointed out that the Internet is not being steered because that is impossible. One can't steer or manage a jungle. The Internet is pure Darwinism, which means that everybody does as one likes and in the long run it is about 'survival of the fittest'. Now that the Internet has been commercialised, a few years ago a more exact phrase would be 'survival of the profittest'!

A balanced meta-point of view has been formulated by Floridi (Ref 3, p. 45):

'The global network is only a stage in the endless self-regulating process through which the human encyclopedia constantly strives to respond to its own growth.'

Let us try to be modern encyclopedists, who take nothing for granted but who are willing to see visions and dream dreams.

2. Empirical findings

2.1. The surveys

During the last two years the present author has carried out studies of the Internet as a tool for research libraries. The studies have been supported by the Danish Ministry of Culture from which the author received a £12,000 grant. Besides following the developments of the Internet as a user, the author has planned and carried out several studies of its use by Danish information professionals and common users. In 1995 a pilot study and a forecast (Delphi study) were completed. The most interesting of the findings have been published (Ref 4, pp. 65-67; Ref 5).

The general interest in figures regarding the Internet decreases in an age of an increasing number of users. Who is interested in the number of radio listeners or telephone users nowadays? In the Western World almost everyone has a radio and a telephone, among many other things. In five years or so from now a majority of the same people will also have access to the Internet as a result of computerisation and cheaper telecom networks.

Below, findings from two different Danish surveys will be presented.

2.1.1. Methodological problems — for safety's sake.

Many Internet surveys are being carried out by asking users to complete an electronic questionnaire and return it by e-mail. In the nature of the case such surveys cannot be claimed to be representative of the population that is being investigated because respondents only make up a random sample. A strictly representative survey would be almost impossible to carry out because the whole group of Internet users in a country or of a certain category is not known, or is described by means of quantitative terms. These circumstances may call the reliability of the survey into question. As matters stand, one can only lean against the findings from other surveys which are carried out in different ways, e.g. of the Gallup poll type. It should however be kept in mind that many Internet surveys are being carried out by asking random users to complete an electronic questionnaire, e.g. GVU's WWW User Surveys (see below). Finally, according to the Law of Large Numbers, when the number of respondents increases, so too does the probability that they make a representative sample.

2.1.2. The author's survey.

In the winter of 1995-96 the present author performed a comprehensive empirical survey in order to throw light on Internet user behaviour. Three special groups of users (business people, higher education/research and students) were selected for further analysis. Focus has been on the use of the Internet as a communication tool and as a means of utilising information sources. A questionnaire was sent to all customers of four of the largest Internet distributors in Denmark (UNI-C, IBM Denmark, Cybernet, and Jyllands-Posten). The distributors agreed to support the survey by distributing the questionnaire via e-mail or traditional mail to their Internet subscribers, the number of which was confidential. A total of 413 filled-in questionnaires were returned, 97% of them via e-mail. The respondents had been Internet users for 18.4 months on the average. Two thirds of the respondents came under the three special groups selected for further analysis: business people 35.4%; higher education/research 19.2%; students 11.6%. Forty-seven percent of the respondents had complied with the author's request to add comments.

2.1.3. The Mediacenter Danmark Internet survey.

Data from a recent Danish Internet survey are also included. The survey was performed in March 1996 for Mediacenter Danmark by Gallup Denmark, and builds on telephone interviews with a sample of 2981 persons (Ref 6). The sample was made by a random selection of telephone numbers. Only persons over 13 years of age were interviewed.

2.2. The users — who are they?

According to the Mediacenter survey (Ref 6, pp. 4, 17), 11% of the adult Danish population has access to the Internet but only 7% are actually using it (10% of all men and 4% of all women).

In the author's survey the following distribution of age for respondents was found (Figure 1).
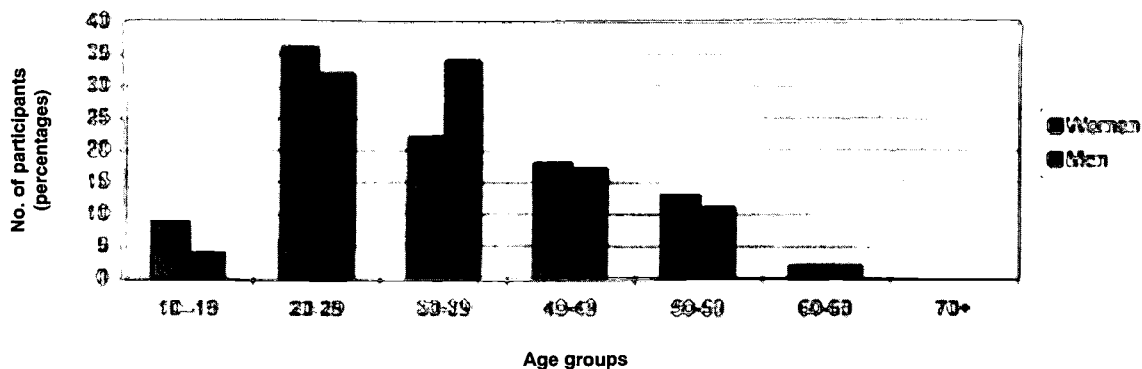
3

Distribution of age

Figure 1: The author's survey: Internet user survey. Distribution of age for women and men (percentages).

There is no doubt that people between 20 and 39 years of age dominate the group of respondents. The average age is 35 (ranging from 15 to 71). Female respondents are in the minority but from the relative figures they seem to be in the majority in all age groups, except for 30-39. In the Mediacenter survey the age group for 25-39 makes up 37% of the respondents and the group for 40-59 another 32% (Ref 6, p. 5).

From a comparison with six similar findings from 1995 it must be concluded that the percentage of female Internet users is 30-40% and is increasing (Ref 7):

'The Internet has traditionally been dominated by males, but this is changing. As more studies of the Net are conducted and research methodologies improve, female users continue to account for greater percentages of the online community. These results should provide valuable fodder for online commerce.'

At the Graphic, Visualization & Usability Center (GVU) at Georgia Technical Research Corporation (Atlanta, Georgia) WWW user surveys are being carried out regularly (Ref 8). The fifth and latest survey (April-May 1996) includes 11,700 respondents, of whom 31.5% are reported to be female, compared with 29.3% for the fourth survey half a year earlier and with 15.5% for the third survey (April 1995). A comprehensive survey of Internet users by A.C. Nielsen (referred to in Ref 9) includes 34% women.

These indications of an increasing female use of the Internet should be compared with the way women are using it (see Section 2.3.).

Further, a recent American survey made by the Online Research Group has shown that Internet users form a younger group than those of other online services (Ref 10). It is understandable that the young users are attracted to the more exciting colours, graphics and variety of services of the World Wide Web. The so-called 'digital kids' will become the future Internet users, and marketers have already begun to focus on the millions of children who are Internet users. In 1995 an estimated three million children in the US had access and it is estimated that 12 million children will have access by the year 2000 (Ref 11, p. 45).

2.3. Internet use — how much and by whom?

An important factor is the period of time the user has been occupied with the Internet. The respondents in the author's survey had been users for 18.4 months on the average. Sixty-eight percent of the respondents had 12 months' experience or less. The average for women was 15.3 months and for men 19.1 months. Concerning the size of current use, the respondents are connected to the Internet 6.14 hours/week on the average. Women and men, however, seem to have different patterns of Internet use. In a comprehensive survey carried out by A.C. Nielsen (referred to in Ref 9) the average Internet user is connected 5.47 hours/week.
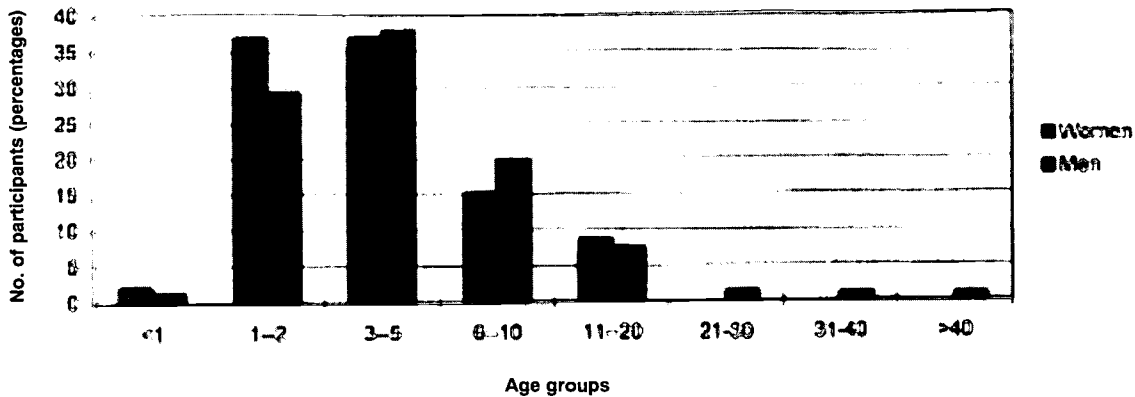
Figure 2: The author's survey: Internet use in hours per week for women and men (percentages).

Figure 2 shows that women are in the majority in the group as people who use it less than three hours per week, while men in general are connected for longer periods of time. On the average women use the Internet 4.79 hours/week and men 6.35 hours/week. This is comparable with a result from an earlier pilot study where it was found that Danish information professionals used the Internet for 4.1 hours/week on average in May 1995 (Ref 4, p. 65).

The Mediacenter survey has some findings that support the hypothesis of women using the Internet for shorter periods of time than DO men (Ref 6, p. 25):

Table 1: Duration of Internet sessions in minutes: Distribution of women and men (Mediacenter survey).

	1-5	6-15	16-30	31-60	60+
Women	23%	21%	28%	16%	11%
Men	16%	22%	27%	22%	13%

Concerning sessions lasting 1-5 minutes, women are in the majority while men clearly prefer longer sessions (more than 30 minutes).

The above findings concerning female and male Internet use are in harmony with several comments from the respondents of the author's survey stating that women tend to use the Internet to solve well-defined problems, while men are inclined to do some net surfing (called 'net watching' by the serious users!) as well. When it comes to information retrieval, female Internet users seems to be more goal-directed and convergent. In most cases they stop when the information searched for has been found. Male Internet users, on the other hand, have adopted a more divergent/explorative attitude which may lead to experiences of the serendipity[1] type once the information searched for has been found.

In the Mediacenter survey, net surfing is the most popular Internet activity followed by database searching and e-mail in the second and third places (Ref 6, p. 18).

2.4. Electronic mail — If I send a letter to you what will happen?

The classic use of the Internet is communicating by sending and receiving electronic mail, including non-textual files. In the author's survey the average user sends 13.3 messages/week (ranging from 0-170) and receives 21.3 messages/week (ranging from 0-1000). The trend is that users who send up to 20 messages/week receive fewer than they send. In the case of over 20 messages/week sent, more messages are received than sent. Thus, there seems to be a 'critical mass' of electronic mailing. Over 20 messages/week could bring about snowballing! And that might happen in step with an increasing use of, for example, distance work/telecommuting.

According to a survey from October 1995 (Ref 12), 20% of US workers will be telecommuters in 2003. Such findings should always be taken with a grain of salt. In the author's survey respondents reported a total of 13.2% of their working hours as distance work. In five years the figure was estimated to a total of 24.4%, which means an increase of 85%, or 17% annually on the average. So far, things have not developed that fast concerning distance work.

2.5. Electronic conferences

Another popular Internet facility is participating in electronic conferences/discussion groups. The respondents in the author's survey subscribed to 5.2 electronic conferences on the average. Fifty-nine per cent, however, did not subscribe at all and 24.3% had limited themselves to fewer than six conferences, which seems to be the maximum for most users provided that regular activities are going on. The information professionals in the above pilot study participated in four conferences on average.

Concerning the number of messages, the respondents only sent 2.7 messages/week (ranging from 0-40) and received 181 messages/week (ranging from 0-3500) on the average. This high number of received messages was mainly due to a few users who gave very high numbers. In the pilot study the information professionals sent 0.8 and received 107 messages/week on average. As was the case in electronic mail, a similar 'critical mass' could be identified in electronic conferencing: Users sending fewer than ten messages/week receive fewer than they send, and users sending more than ten messages/week receive more than they send.

2.6. Quality in information retrieval

A general outline of quality issues in connection with information retrieval can be found in Ref 13, pp. 46-54. The question is further developed in Ref 14.

Most of the respondents in the author's survey were using Internet resources for information retrieval. In their opinion this had improved the quality of their reference work to some extent. On a rating scale from 0 (not at all) to 5 (very), the three special groups had the average scores shown in Table 2.

Table 2: The author's survey: Average ratings of improvement of information retrieval quality for three special groups.

Groups of users	Average rating
Business people	3.65
Researchers	3.46
Students	3.41
All respondents	3.52

Figure 3 shows the different experiences of quality change more clearly. In the business peoples' opinion the improvements have been remarkable, while the researchers and the students are more reserved.
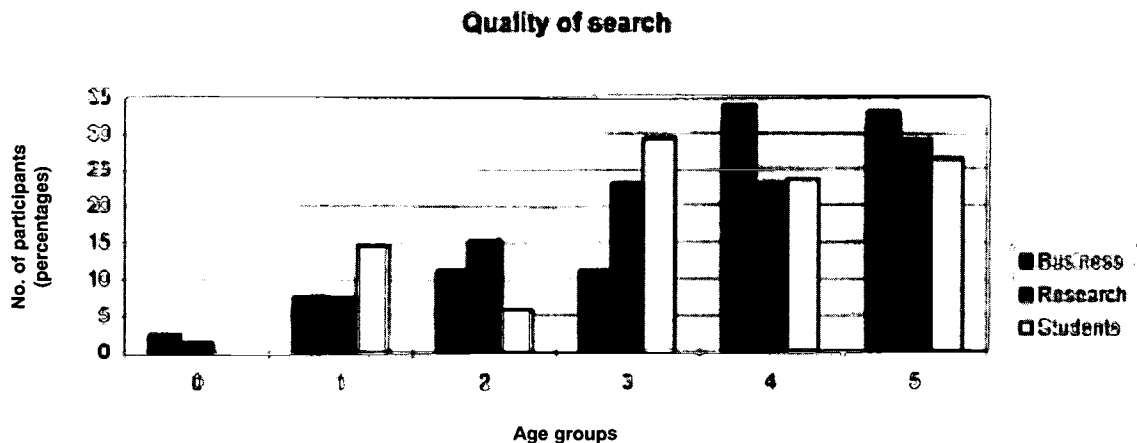
Quality of search

Figure 3: The author's survey: experienced improvement of information retrieval quality (in relation to the situation before use of Internet). Numbers as percentages of respondents.

In the pilot study the information professionals only scored 2.1 on the average. This is quite remarkable compared with the above ratings. A hypothesis could be that increasing information professionalism results in a more sceptical attitude towards the Internet-based resources, their lack of organisation and the primitive search tools available (compared with traditional online and CD-ROM searching). The above findings are not very suitable for a proof or disproof of the hypothesis, but the information professionals at least know where the shoe pinches.

6

Further, the Internet users were asked to indicate their expectations of the quality of Internet information sources in three years' time. The rating scale used ranged from -5 (considerably worse than today) over 0 (same as today) to +5 (considerably better than today). The three special groups of users had the average ratings shown in Table 3.

Table 3: The author's survey: average ratings of expectations of change of quality of Internet information resources for three special groups.

Groups of users	Average rating
Business people	3.07
Researchers	2.52
Students	2.58

Figure 4 gives the detailed findings of this part of the survey. It is notable that only very few respondents express negative expectations. Further, business people are more optimistic than academics and students.
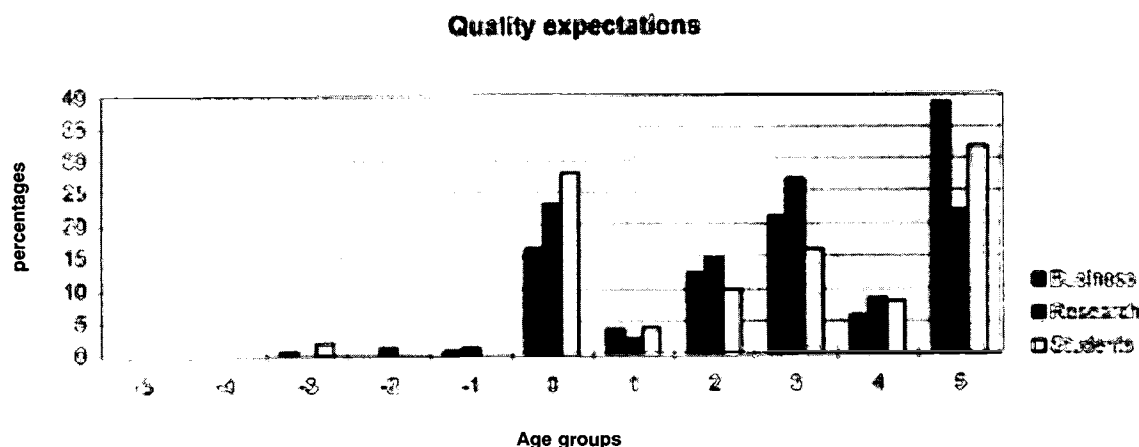

Quality expectations

Figure 4: The author's survey: Expectations of change in quality of Internet information resources for three groups of respondents. Numbers as percentages of respondents.

Students, on the other hand, seem to split into two main groups: the *moderate pessimists* (rating 0) and the *optimists* (rating 5). In this way the students have adopted an attitude similar to the information professionals in the pilot study. At this rating, where the average was 3.0, the same two separate groups could be identified. Elements of en explanation may be found in the fact that students have more contact with information professionals than do the other two groups.

There is no denying that there is a widespread need for quality control of Internet information resources. This need has been formulated in a clear-cut way by Floridi (Ref 15, p. 51) who maintains that:

'Unless academic and cultural institutions provide some form of quality control, we may no longer be able to distinguish between the intellectual space of knowledge and a very polluted environment of junk mail and data.'

This important concern about the very content of the information field leads to a closer analysis of the situation, and an attempt at pointing out some possible themes that should be brought into focus of the information professionals' endeavour to overcome the above problems.

3. Two models for the library information network environment

Why are many people displeased with the Internet as an information source? A recent survey (Ref 16, pp. 136-139) among British business librarians points out some of the most serious disadvantages (Table 4).

Table 4: Disadvantages of the Internet as an information source.

Disadvantage	Importance weighting
Lack of organised information	28%
Lack of quality of information	21%
Lack of relevant information	19%
Lack of reliable information	18%
Too much information	14%
Total	100%

According to the survey business librarians consider Internet information sources *over-hyped*. The biggest concerns are retrieval problems, which are seen as rooted in the legendary and spreading lack of organisation in the Internet. Among further complaints are the lack of authoritative data, disappearing or moving information, and the fluctuating quality of information from site to site. All things considered, 'the Internet is for people who have enough time to browse through irrelevant information in search of pearls. But for business librarians this browsing becomes more like ploughing' (Ref 16, p. 138). The respondents admit (or rather, concede), however, that there are some valuable information sources on the Internet, and many of them have built up their own portfolio of useful sites.

A new phenomenon has come up in reference work in general: patrons are increasingly asking for 'Internet information'. They assume that the information searched for is (only) available on the Internet. Moreover, if the information is actually found on the Internet, they assume that it must possess a number of qualities for that reason alone. The reverse argument has also been stated: if something is not on the Internet, the information simply does not exist (Ref 16, p. 139).

In the author's opinion the above attitudes and trends concerning Internet information sources have the common denominators of:

- *ignorance* of what information literally is (at last, information is not dependent on the media);
- *confusion* of different things (e.g. real time and retrospective databases);
- *Internet superstition* (the glories of the Internet transferred to a single information source).

Many of these difficulties might be prevented if all persons involved would realise that today's information jungle (including the Internet sources) comprises several categories that differ essentially concerning information quality in the widest sense of the word.

A suitable model for putting the record straight is one inspired by L. Costers (Ref 17, pp. 319-320). According to Costers research libraries, in offering electronic services, should guarantee the quality of such services and of the documents provided. The libraries should cooperate in order to create a *Controlled Information Network Environment*. This will give the information professionals an opportunity of selecting Internet-based information resources according to some quality criteria (creating 'virtual reference desks'). Costers allows for three different service layers to develop (Table 5).

Table 5: Costers' model of service layers in libraries.

Service layer	Library services
1	Own electronic services (e.g. OPAC). Services bought to the library (e.g. CD-ROMs, full-text databases).
2	Electronic services of a regional/national library cooperative to which the library contributes.
3	The Internet and its information resources

It is easy to see that in these three different service layers or information network environments, the possibilities of exercising quality control are decreasing in step with development. A library has 100% control over its own OPAC, a little less for, say, a joint national OPAC network, and very little concerning the Internet. Consequently, the library cannot answer for the information quality in the same way in all three layers. Information professionals ought to have a good grasp of these matters but, unfortunately, this is not always the case. Library users may be in an even more unfortunate position by virtue of lacking a general view of the situation. Library users will only be able to make the best possible use of the different information resources if they fully realise the importance and consequences of the different quality of service offered by a particular library.

The concept of 'information' is not an easy one to think of. In order to describe and perhaps even understand

the trends in the development of contemporary and future information systems one also has to use the concept of 'virtuality'. Originating in the subject of computer simulation, this concept has been used increasingly in the information field during the last 6-8 years. The very subtle concept of 'virtuality' (not to mention 'cyberspace'!) may be even more difficult to understand than 'information', and thus a number of information professional/user misunderstandings are imminent. A usable and operational definition of virtuality is offered by Michel Bauwens, who has contributed extensively to the development of these matters. According to Bauwens (Ref 18, p. 2):

> '. . . virtualisation means that every piece of information produced everywhere in [an] organisation, is accessible from anywhere, anytime.'

When is all this going to happen? It has been on its way since the very beginning of the online age. According to a recent study of the future of libraries, free Net access in American public libraries will be available in 2003; the first virtual large library will open up in 2016; and half of the Library of Congress will be digitised in 2043. The figures are average estimates made by five experts (Ref 19).

Michel Bauwens maintains that virtualisation is going to take place in the libraries during a series of steps which will overlap (see Table 6).

Table 6: Degrees of virtuality in the development of libraries (Bauwens).

Degree of virtuality	Access	Documents
1	Electronic, to OPAC	Paper
2	Electronic, to virtual collections	Paper
3	Electronic, to virtual collections	Electronic

This development will necessarily include:
- the replacement of printed matter by electronic information;
- increasing use of computer-mediated communication;
- universal access to information.

In the author's opinion many pitfalls and talking at cross purposes could be avoided if information professionals and users shared a certain minimum of understanding concerning these different levels of service, virtuality and information quality. After all, this is only a bringing up to date and an extension of some traditional Library and Information Science core concepts. Until the online age librarians often experienced similar problems when patrons confused, say, books with bibliographies.

4. Conclusion

It is an old sociological rule of thumb that technology is moving faster than the institutions, which in turn are changing faster than attitudes.

Anybody knows that information technology is developing very fast at present. In the above case of Internet users and use it was found that some groups of users (women, young users) are running fast in order to keep up with the times and, preferably, to gain on technology. That leaves behind the institutions some of which in the information field might need to rearm in order to survive; especially concerning distance work, which it is believed will develop rapidly in the near future and require new thinking on management, responsibility, and social issues.

A more profound understanding of information quality in the light of virtualisation could lead to less disappointment over the whole lot of the Internet.

Acknowledgements

9

Dr Helge Clausen
State and University Library
Universitetsparken
DK-8000 Aarhus C
Denmark
Tel: +45 89 46 22 64
Fax: +45 89 46 22 20
Private tel/fax: +45 86 99 63 54
E-mail: hcl@kumsb.dk

References

[1]   Nu skal Internettet tages alvorligt (1996) (Now the Internet should be taken seriously), *Datatid*, 5, 22.

[2]   URL: http://www2.mids.org/ids3/pr9510.html

[3]   Floridi, L. (1996) The Internet: which future for organised knowledge, Frankenstein or Pygmalion? Part 1, *The Electronic Library*, 14 (1), 43-48.

[4]   Clausen, H. (1995) Information quality as a factor in research library internet use, *Proceedings of the 19th International Online Information Meeting*, Learned Information, Oxford, pp. 61-69.

[5]   Clausen, H. (1996) Web information quality as seen from the libraries, *New Library World*(in press).

[6]   *Adgang til og brug af Internettet i Danmark*(Access to and use of the Internet in Denmark). Mediacenter Danmark, May 1996, 30 pp.

[7]   Barkow, T. (Ed.) (1996) Gender and the Web, *Wired*, 4 (4), 70.

[8]   GVU's 5th WWW User Survey: http://www.cc.gatech.edu/gvu/user_surveys/survey-04-1996/

[9]   URL: http://www.infokraft.se/si.htm

[10] Rupley, S. (1995) Do you fit the Internet profile? *PC Magazine*, 14(20), 30.

[11] Arnold, S.E. (1996) *Publishing on the Internet: A New Medium for a New Millennium*, Infonortics Ltd., Calne, Wiltshire.

[12] Pescovitz, D. (1995) The future of telecommuting, *Wired*, 3 (10), 68.

[13] Bang, T., H. Clausen, S.E. Dyrskov, C.G. Johannsen and O. Johnsen (1996) *ISO 9000 for Libraries and Information Centres: A Guide*. Report of a project supported by NORDINFO. FID, The Hague (FID Occasional Paper 13). Published in Danish in 1995 by NORDINFO.

[14] Clausen, H. (1994) ISO 9000 and all that: is the information sector ready for the big quantum leap? *Proceedings of the 18th International Online Information Meeting*, Learned Information, Oxford, pp. 37-46.

[15] Floridi, L. (1996) The Internet: which future for organised knowledge, Frankenstein or Pygmalion? Part 2, *The Electronic Library*, 14 (1), 49-52.

[16] Kelly, S. and D. Nicholas (1996) Is the business cybrarian a reality? Internet use in business libraries, *Aslib Proceedings*, 48(5), 136-144.

[17] Costers, L. (1994) The electronic library and its organizational management, *Libri*, 44(4), 317-321.

[18] Bauwens, M. *Cyberspace, Virtualisation, and the Role of Cybrarians*. URL: http://www.ecom.be/SPACE.TXT

[19] Pescovitz, D. (1995) The future of libraries, *Wired*, 3(12), 68.

Note

[1]   **Serendipity.** A happy coinage by Horace Walpole to denote the faculty of making lucky and unexpected finds by accident. In a letter to Mann (28 January 1754) he says that he formed it on the title of a fairy story, *The Three Princes of Serendip*, because the princes 'were always making discoveries, by accidents and sagacity, of things they were not in quest of.' Serendip is an ancient name for Ceylon (Sri Lanka). The Wordsworth *Dictionary of Phrase & Fable*, 1993.

NOTICE

REPRODUCTION BASIS