

REVISTA DE HISTÓRIA DAS IDEIAS 18

HISTÓRIA • MEMÓRIA • NAÇÃO



INSTITUTO DE HISTÓRIA E TEORIA DAS IDEIAS
FACULDADE DE LETRAS

COIMBRA 1996

THE INTERNET: WHICH FUTURE FOR ORGANIZED
KNOWLEDGE, FRANKENSTEIN OR PYGMALION?* ***1. What we might have known*

In 1963 Arthur C. Clarke published a story called "Dial F for Frankenstein", in which he imagined the following scenario. On the 31st of January 1974, the last communications satellite is launched in order to achieve, at last, full interconnection of the whole, international telephone system. The day after, all the telephones in the world ring at once. It is the horrible cry of an artificial supermind, born out of

* Luciano Floridi (floridi@vax.ox.ac.uk) is Research Fellow in Philosophy at Wolfson College, Oxford, and Lecturer in Philosophy of Jesus College and St Anne's College, University of Oxford. He was educated at the University of Rome "La Sapienza" and received his M. Phil, and his Ph. D. in Philosophy from the University of Warwick. He has published numerous articles on epistemology, the transmission of texts, the history of scepticism and humanities computing. He is author of *Scepticism and the Foundation of Epistemology* and editor of the *Iter Italicum* on CD-ROM (Leiden, Brill, 1996)

** This article is a revised version of an invited paper I gave to the first international conference, promoted by the UNESCO Philosophy Forum, to celebrate the fiftieth anniversary of the founding of the organization (Paris, 14-17 March, 1995). Several electronic versions were made available through the Internet in the last few months, and I wish to acknowledge the suggestions and comments sent to me via e-mail by hundreds people who helped me to improve it. Hard-copies were published in *The Electronic Library* and in *Bollettino della Società Filosofica Italiana* (in Italian).

the global net work of 80 billion individual switches in the world's automatic exchanges. The story was a piece of science fiction, and nobody remember it in 1975, when Apple launched its first models of personal computer. We were at the beginning of the second age of Information Technology. During the following ten years personal computers reached mass diffusion, and became a comodity. In 1982 *Time* declared the computer "Man of the year", and even the space age started to be seen as an episode of the computer age.

In 1984 Intel and Microsoft were still building their empires when the newspapers suddenly rediscovered George Orwell's negative utopia. In *Nineteen Eighty-Four* he imagined a possible world dominated by propaganda and lies, without a language to express critical thought, and controlled by a Big Brother. In search of inspiration, many intellectuals reached for their copies of McLuhan's classic, *Understanding Media*. Most people were confident that Big Brother could be identified with the television. The global village was going to be the effect of a TV culture. Few people seemed to recollect *War Games*, a very popular film which had been on showing the year before. The plot was very simple: a clever teenager plays with the main frame of the Pentagon by means of a personal computer and a modem, and in so doing he runs the risk of starting the third world war. One could simply dismiss *War Games* as one more example of the gap between the two generations. And one would have misunderstood what the real novelty was: the remote control of a computer, the interactive play through the network, the exciting adventure represented by the possibility of navigating into a virtual reality.

War Games was a piece of science fiction only for Europeans. Since 1968, the Department of Defence of the U.S.A. had supported the Advanced Research Projects Agency (ARPA) for the construction of a network, known as ARPANET, which might still provide a system of communication after a nuclear attack. ARPANET had to be completely decentralized, so that each packet of information could still travel in the direction of its final destination through any of those nodes that were still functioning, no matter which particular route would still be available. ARPANET provided the first decentralized structure, which was going to remain the fundamental model of the subsequent Internet. The growth of ARPANET and many other networks throughout the country like USENET had been so successful that in 1981 another grandparent of the Internet had

appeared, BITNET, which provided electronic mail and servers to distribute information. In 1982 the Transmission Control Protocol (TCP) and the Internet Protocol (IP) had been established as the two standards whereby different networks could communicate. It was a lingua franca for the entire network, and in more than one sense the new, anarchic state of the Internet was "virtually" born. So in 1983 War Games was for the American public just an imaginary interpretation of a rather common reality.

Meanwhile, the international network had kept on growing at a fantastic speed. Further improvements in computer technology and in the telephone lines allowed faster, more reliable and cheaper way of communication, which in turn opened the network to more users. The growing mass of users made it more convenient to provide further services and resources. More and more CD-ROM catalogues became searchable through the network, and an increasing amount of free software was soon available online, thus attracting even more users, and so forth. Commercial services started to use the network, offering access and further applications. After 1988 the number of "host" machines with direct TCP/IP connections began to double every year. In 1989 the RIPE (Reseaux IP Europeens) was established in order to coordinate the development of the Internet in Europe. By then, several national networks were fully developed and interconnected, while linking each day more and more sites. The escalation reached its critical point in the early 1990s. The end of the cold war opened the Internet to Eastern European countries. In 1993 there were 176 millions of computers in the world, and theoretically each of them could have been connected to the Internet. In fact many of them were. So last year the Internet finally appeared to the general public as the most revolutionary innovation since the invention of telephones, though in this case *Time* missed the opportunity to elect the Internet man of the year. Today, twenty five years after the first exchange of e-mail on ARPNET, over a hundred countries are online, there are thousands of nodes providing entire libraries of any sort of digitized information, the Internet is growing at a rate of twenty percent a month, and it has become a vital instrument for the development of research, communication, business and education. A whole population of several million people interacts by means of the global network. It is the most educated intellectual community ever appeared on earth, a global academy that, like a unique mind, thinks always.

The Internet is a completely new world, about which we seem to know very little. Despite Clarke, McLuhan, Orwell or films like *War Games*, its appearance has found most of us, and especially the intellectual community, thoroughly unprepared. In the last issue of the *Philosopher's Index* on CD-ROM the word *Internet* did not even occur. We do not know what to think of this new phenomenon, or how to deal with it. What is the Internet exactly? What can it be used for? And what will be the effects of such a radical revolution in the way we handle the world of information? These are the three fundamental questions that will determine the future of organized knowledge; three questions that, in this paper, I shall attempt to answer only with an increasing degree of imprecision.

2. *What we know*

The first question is the least difficult to answer. By means of the word "Internet" we usually refer to the international system of digital communication, emerging from the agglomerate of thousands of networks that interact through a number of common protocols all over the world. As a supervening entity, the Internet is an epiphenomenon that cannot be physically perceived, meaningfully located in space and time, over above the set of interacting networks that constitute it. Nobody is finally responsible for it as a single enterprise. It is a collaborative initiative, a set of services and resources, and each network that is part of it is accountable only for its own proper functioning. Nobody is earning money from the whole service either, because the Internet is free, in the same way as you would expect the information in a public library to be free. Of course, there are commercial services on the Internet as shops in a museum but, again, the museum itself is free. One may have to spend money to reach the place and stay there, as one may need to pay a telephone bill or a fee to a commercial company that gives you access to the service, but then the resources are there at one's disposal, free of charge. Nobody is running the system or will be able to control it in the future. When Bill Gates speaks of his plans to dominate the Internet, what he really means is that he wants to make sure that the new version of Microsoft Windows and its network will result the most popular interface and route for single users who wish to manage their own access to the network through their personal computer. In other words, he would like to run the main bus line that takes one to

the library, not the library in itself. It is already a very ambitious project.

The second question — what one can do with the Internet — must be left slightly more open, though for a different reason. It is not that we do not know how to use the system; it is rather that the variety of things that one can do through the Internet increase literally every single day, and cannot be finally fixed. Usually, we are told that four typologies of communication are implemented through the network: electronic mail, discussion groups, remote control and file transfers. More simply, this means that through the Internet we can exchange private messages with a friend, create an electronic club to discuss a particular topic with other people on the network, publish an electronic journal, setup a so-called slow reading group on Voltaire's *Candide*, access and control other computers in other places and so acquire data in all possible forms, from free software to millions of bibliographic records, from thousands of electronic texts (all the major philosophers are online) to the splendid images of the Louvre paintings, from statistical graphs, to musical sounds and whole data banks concerning an enormous variety of subjects. Any exchange and manipulation of symbols, images and sounds is already possible on the Internet, or soon will be, so much so that in the future even television will probably be remembered as just another episode of the computer age.

3. *What we do not know*

Such an extraordinary range of possibilities makes our last question — how the Internet is going to affect organized knowledge — almost impossible to answer. In this case, it is even hard to give an initial shape to our ignorance, since there may be much more we do not know than we could guess.

One could start by remarking that, as a synthesis of several innovative technologies, the Internet is fostering the growth of knowledge, while at the same time generating unprecedented forms of ignorance. This is a common process in the history of technology. Whenever a radical change occurs in the way we handle our knowledge, some individuals are left behind, uninstructed, and the new technology makes those who do master it suddenly aware of other domains of knowledge still to be explored. What is truly

remarkable, in our case, is the impressive scale of the phenomenon in question. Not only has the Internet already enlarged our notion of illiteracy, and produced new forms of cultural isolation and discrimination, but, because it intensifies and amplifies the effects of the digital revolution, the Internet is also transforming some of our most radical conceptions and habits. Take our conception of a text for example. The enormous importance of the new model of "spineless textuality" represented by the hypertext, the virtual ubiquity of documents, the appearance of online services and electronic sources that need to be catalogued, have all changed a discipline like librarianship in the most radical way. Even the library itself may disappear, as we move from the holding and lending library, which stores knowledge physically recorded on paper, to the consulting library, which provides access to electronic information on the network; then from an object-oriented culture, which produces multiple copies of physical entities like books and CD-ROMs for each user, to a culture which appreciates the relationship between time and information, and provides services charged per period of use; and finally from the library as a building to the library as a gate-node in the virtual space of the digital encyclopedia.

The Internet has transformed the physical citizens of a modern society into the disembodied netizens of a postmodern cybercommunity, as some hackers like to say. The jargon maybe be a bit extravagant, but the changes are almost tangible. In the electronic Agora of the global village, publicity has assumed an international scale, while privacy means electronic privacy in our e-mail conversations. Our good manners are evaluated on the basis of a "netiquette" of communication. Civil rights concern the way in which information about ourselves can be created and stored in databases, and then accessed and used through the network, whereas crimes range from electronic pornography to viruses, from the illegal reproduction of software to illicit intrusion into and mishandling of electronic systems, from the infringement of copy right to electronic plagiarism. Even the way we think may in the long run be affected, for relational and associative reasoning is nowadays becoming as important as linear and inferential analysis, while visual thinking is once again considered to be at least as indispensable as symbolic processing. And as the skill of remembering vast amounts of facts is gradually replaced by the capacity for retrieving information and discerning logical patterns in masses of data, the Renaissance conception of erudition and

mnemotechny is merging with the modern methods of information management.

In the electronic village implemented by the global network, sectors of activities like communicating, writing, publishing and editing, advertising, selling, shopping and banking, or teaching and learning are all being deeply affected. Such transformations are of the greatest importance, as they will determine our lifestyle in the coming decades. If I leave them aside now it is only because in the rest of this paper I wish to explore what such an epochal change in our culture will mean in one special field: the future of the system of encoded knowledge, the human encyclopedia.

4. The digital revolution

In order to understand the new problems that our organized knowledge is likely to face, and the new possibilities of which we may soon be able to take advantage, it is essential to put the appearance of the Internet into the context of the history of the human encyclopedia, and to outline the three principal paths followed by the digital revolution.

The human encyclopedia as a whole is an intellectual space whose density and extent are constantly increasing, although at disparate rates in different ages and cultures. The relative extent to which such a domain can expand at any particular time is directly proportional both to the quantity of information stored until that time, and to the current degree of accessibility of the memory of the system. There is no solid and constructive advancement of knowledge without an accurate memory of the past, but the interminable process of encoding and recording the human encyclopedia would obviously be fruitless, if efficient and economical methods were not concurrently devised in order to make past knowledge adequately available. So an effective management of organized knowledge aims at establishing a satisfactory equilibrium between the increasing degree of magnitude and complexity of the system of knowledge on the one hand, and the equally expanding amount of information required by its constant growth on the other, the concomitant level of accessibility representing the degree of success reached by such efforts at any particular moment in the history of knowledge. This explains why the rather brief history of western culture could be outlined in terms of the evolution of the methods whereby the encyclopedia has been made potentially

available to the individual mind. Organized knowledge moved from the unstable memory of oral transmission to the written text, from volumina, providing only a linear access to their contents, to codices, which made possible a random access, from the appearance of the tables of contents to the elaboration of indexes, from the establishment of public libraries to the elaboration of a science of bibliography, from the appearance of new reference books, such as encyclopedias or lexicons, to the publication of reviews and ever more specialized forms of literature.

Within such a process, the invention of printing has usually been considered a turning point, and it certainly deserves a special place in the history of human knowledge, as long as its importance is not misunderstood. The printed book represented a powerful new medium whereby a text, now mechanically fixed on paper, could be reproduced more quickly and cheaply, and hence be more safely stored and more widely diffused. Its innovative character consisted therefore in the tremendous acceleration it brought in the recovery, conservation and dissemination of knowledge among an increasingly large number of people. But such an enlarged availability of knowledge did little to improve the degree to which an individual could take full advantage of the entire encyclopedia, since the process of information retrieval remained largely unaffected by the appearance of books. Quite on the contrary, in terms of efficient utilization of all the available data, the publication of million of texts made problems of management only more complex, while their solutions became every day more crucial. The wider the extension of organized knowledge became, the higher and more specialized the hierarchy of the maps to the universe of knowledge had to be. In the long run, the printed medium showed itself to be capable of coping with the extension of knowledge only vertically, by producing more and more specialized guides and then further guides, in an endless hierarchy. Soon after Guttenberg, the development of new, automatic methods to manipulate, access and control the encyclopedia became first critical, then indispensable. As a consequence, the history of modern technology is full of attempts to reproduce, at the level of the automation of the information process, the same mechanical improvements already achieved at the level of the reproduction of knowledge (one can think of the academy in *Gulliver's Travels*). That they all failed is well known. The reason why they failed is that such an enterprise required a much more radical approach than the simple

extension of a mechanical solution to the problem of semiotic management. A modification both in the language of the encyclopedia — the binary logic — and in its physical nature — the electronic implementation of such a logic — was necessary.

Only the passage from printed paper to digital data made possible a thoroughly new way of managing information, and a much more efficient control over the system of knowledge. This explains, by the way, why Information Technology, as the long awaited response to the invention of printing, has been much more pervasive than any previous technology. Our societies live on information, and wherever there is information to process, from military applications to business, from bureaucratic word processing to the proper functioning of white and black goods (like a fridge or a stereo), Information Technology has become an element of vital importance.

5. Three Steps

The press mechanically enlarged our intellectual space; the computer has made it electronically manageable. The required process of conversion of the entire domain of organized knowledge into a new, digital macrocosm began in the fifties. And since then, it has followed three fundamental directions: extension, visualization and integration. The constant growth in the kinds of information that could be digitized has led to the construction of a domain that has come to include not only numbers and texts but also sounds, images and animation. The growing extent of such a binary domain has soon required forms of access far more congenial to the mind than the digital. The invention and improvement of visual displays units, together with the development of graphic interfaces and WIMP (Windows, Icon, Menu, Pointer) applications, have made possible a spectacular return of the analogical, as the fundamental threshold between the binary macrocosm and the mental microcosm. Finally, the translation of alphanumeric texts, images and sounds into the simple language of bytes has made possible an increasing integration of the various domains of knowledge into an ever wider and more complex encyclopedia. The process has further developed in two directions: the qualitative integration of the various types of digitized knowledge by means of multimedia and virtual reality; and the quantitative integration of the local domains into an ever wider environment of networks, which tends towards a global, multimedial and unique macrocosm of digitized knowledge.

6. Novelties and Problems

What, then, is the Internet's place in such a cultural scenario? As the most recent form adopted by the organization of the system of knowledge, the global network is only a stage in the endless self-regulating process through which the human encyclopedia constantly strives to responde to its own growth. As a consequence of the three processes of extension, integration and visualization, the Internet has made possible a management of knowledge that is faster, wider in scope, more complete in terms of types of information and easier to exercise than ever before. Finally, as a stage in the life cycle of the encyclopedia, the network has already given rise to unprecedented innovations and to new fundamental problems, some of which are especially relevant to the future of scholarship and organized knowledge.

7. What we can do: ideometry

A lot could be said about the important innovations that the Internet has brought in the field of the production and management of organized knowledge. One may think, for example, of the reduction of the time-lag between the production and the utilization of knowledge, the promotion of international cooperation and free share of information among researchers and scholars, or the possibility of remote teaching online. Yet most such novelties are actually less revolutionary than they seem to be, since their principal task is to make easier and quicker whatever we used to do more slowly and with greater difficulties. There are other possibilities, however, which do represent a more radical change with respect to the past. The global network, for example, is weakening the concept of specialization. We have seen that the era of the printed book, providing a rigidly structured context, invited vertical specialization, especially in the humanities. The electronic encyclopedia, on the other hand, now promotes forms of horizontal interdisciplinarity and multidisciplinary which will probably start bearing fruit in the next decade. The more flexible the space of knowledge becomes, the more "diatopic" our approach to it can be. It is difficult to limit oneself always to the same confined space, if one can navigate so easily across the disciplinary boundaries. Among such radical innovations, the most substancial concerns the way in which we shall be able, in

an ever easier way, to acquire further knowledge about the encyclopedia itself. Consider once again the intellectual space of organized knowledge. We can distinguish at least three different dimensions. The first one is represented by the ordinary field of primary data sets, what we usually perceive as the encyclopedia tout court. This is the principal information we can acquire when we have access to the encyclopedia, and it is also the information the encyclopedia is generally designed to convey to the user in the first place. We should keep in mind that primary data are meant to be multimedial, and so can range from a photograph to an ordinary text, from a mathematical formula to a statistics diagram, from an entire database to a piece of music, from a manuscript to a collection of drawings. The second dimension is represented by the so-called metadata, that is secondary indications about the nature of the data sets constituting the first dimension. Here we can find (possibly multimedial) information about copyright restrictions for example, or the collocation of our data sets in a physical library or in a virtual domain, the subject covered by the data sets, the quality of the information conveyed, and so forth. You can think of metadata as library records. The third dimension is represented by the derivative data that can be extracted from primary data sets, when the latter are used as a source for comparative and quantitative analysis. Let me first provide a slightly technical explanation. I hope the very simple example will be more intuitive.

In the book age, primary data sets were collected and organized in structures which were necessarily rigid and unalterable. Not only did the ordering principles create a domain, they also established de facto the limited range of primary questions which would be meaningful to ask. For example, if the ordering principle stated that the primary data should be all the poetic texts of any time written in English, the final edition in several volumes of all English poems provided the means to answer properly and easily only a limited range of primary questions, like who wrote what or when. There was a direct correspondence between ordering principles and possible questions. Information Technology has transformed all this. For it is now possible to query the digital domain and shape it according to principles which are completely different from those whereby the primary data were initially collected and organized. The structure of our particular set of digital data can be modified to fit an infinite number of requirements, and hence provide answers to secondary

questions which were not meant to be answered by the original structure. The new patterns that emerge from the application of quantitative and comparative queries may turn out to be meaningful and interesting for reasons that are completely extraneous to the initial ordering principle. Ideometry is the morphological study of such significant patterns resulting from a comparative and quantitative analysis of the field of knowledge, that is clusters of primary data like data-banks, textual corporal or multimedial archives used as extensional sources. Derivative data, that is the third dimension of the encyclopedia, are the outcome of an ideometric analysis of whatever sector of organized knowledge has been subject to investigation.

The difference between primary and derivative data can be more easily illustrated if we return to our collection of English poems. In 1994 Chadwick-Healey published the database of English Poetry on CD-ROMs (version 3.1). The structure of the digital collection is thoroughly flexible and we can re-organize it at will. As a simple example, we could study the presence or absence of the two popular figures of Heraclitus, the weeping philosopher, and Democritus, the laughing philosopher, through the entire set of documents. A quick survey shows that the joint motif of compassion for human misfortune and derision of human ambitions was very popular between the second half of the sixteenth and the first half of the seventeenth century, as it is in this period that we find most of the poets using the philosophical couple as a literary device. The pattern becomes even more interesting once we notice that during the seventeenth century the two Greek philosophers were portrayed in many Dutch paintings. Through a quantitative and comparative analysis we have made the encyclopedia speak about itself

Ideometry is perfectly in line with the development of the history of thought, which through time becomes progressively more and more self-reflective. As a cluster of multidisciplinary methods, it has become popular in many disciplines since the sixties. Scientometric historiography, lexicography, stylometry, prosopography, citation analysis, bibliometric studies, econometrics and quantitative history in general have all used forms of quantitative analysis in order to investigate their topics. During the second age of Information Technology, however, scholars could perform ideometric analysis only on a limited scale and with enormous efforts. The trouble was, quite simply, that IT was not yet up to scholarly expectations and needs. It

was not that the Humanities were not sufficiently "scientific" to allow the application of IT tools, but rather that IT was too primitive to be of any real service for the highly sophisticate tasks required by scholarly research. The radical change brought about by the third age of IT and the Internet is that an ideometric approach is now becoming an increasingly easy option for any researcher. It is obvious that primary data need metadata in order to be manageable, so the second dimension of the encyclopedia can never be really separate from the first. Derivative data, however, are not so directly available, and the third dimension emerges only when large amounts of primary data are collected in digital form, are made easily accessible to the user, and can be rapidly queried and thus re-structured via electronic tools. Today all these conditions are being more and more adequately fulfilled by the Internet.

When simple and economical tools for studying visual and acoustic patterns also become available, ideometric analyses may be extended to the entire domain of the enlarged encyclopedia. At the moment, ideometry shows that digital texts, though they maintain some of the basic features of printed books and can therefore be used as surrogates, should not be understood as if they were meant to fulfil the same task. We do not convert printed texts into electronic databases in order to read them better or more comfortably. For this task the book is and will remain unsurpassed. We collect and digitize large corpora of texts in order to be able to subject them to comparative and quantitative analysis, and extract knowledge they contain only on a macroscopic level. So what is revolutionary in an electronic bibliography for example, is not that I can find what I need in a few seconds, for this is trivial, but that I can check when historical books on Analytic Philosophy started to be written, and discover how their number increased while the movement was becoming more and more scholastic. Corpora of electronic texts and multimedial sources are the laboratory for ideometric analysis. The larger the domain, the better it will be, for the ideometric value of a extensive corpora is given by the product rather than by the simple arithmetical sum of the ideometric value of each single document. ⁸

8. What we need to solve

The time has now come to turn to the new problems that the growth of a network of information and communication has already

caused or soon will give rise to. My impression is that there are at least ten principal questions worthy of attention. I shall deal with them very briefly and in what I take it to be their increasing order of importance.

1) The digital discrimination

We have already entered the stage in which digital information is unduly preferred over non-digital simply because it is available online, not because of its quality. However, the more resources undergo the conversion, the less serious the problem will become.

2) The disappearance of the great compilers

By greatly increasing the supply of data, the Internet satisfies, to a large extent, an ever growing demand for information. In this process, the use value of information has increased steadily in parallel with the complexity of the system. Its exchange value, however, has been subject to a radical modification. Because of the great and rapid availability of data, the Internet has caused a devaluation of some intellectual enterprises, like compilations, collections of images, bibliographical volumes and so forth, whose original high value depended mainly on the correspondingly high degree of inaccessibility that afflicted information in the book era. Today, some of the data, that in the past had to be discovered and collected with great expense of time and energy, are freely available on the Internet. The result is that users will look more and more frequently for value added to works done on raw data and then on electronic sources and tools. The era of the great collectors on paper is over. Raw data simply won't travel through the printed medium any longer, they will be (more or less freely) available through the digital channels of the network.

3) The emergence of the computerized scholar

So far, the academy has been slow in recognizing that new forms of scholarly activity have appeared, like moderating a discussion list, keeping an online bibliography constantly updated, or publishing a paper in an electronic journal. The sooner such activities will be properly recognized and evaluated, the easier will become for individuals to dedicate more time and efforts to the digital encyclopedia, the more the latter will improve, compelling further recognition for

those who have worked on its creation, and so forth, until a full balance between work and reward will be reached.

4) Stored knowledge > knowledge accessible

There is more knowledge on the Internet than we can access. In the macrocosm represented by the Internet, the fundamental imbalance between the extraordinary breadth of the system and the limited amount of knowledge that can be accessed by an individual mind at any one time arises again because the technology responsible for the processes of management is far less efficient than the technology responsible for the accumulation of knowledge. The quantity of information potentially available on the Internet has increased beyond control, whereas the technology whereby the network actually allows us to retrieve our data has improved much more slowly. The result is that nowadays we are once again very far from being capable of taking full advantage of the full extent of our digital encyclopedia. Information Technology developed in the fifties as an answer to the explosion of the system of knowledge originated by the book age. After a few decades, it can be seen to have defeated itself by extending the encyclopedia beyond its own control. The challenge of the next few years will consist in narrowing the gap between quantity of information and speed of access, before it becomes even greater. Projects like the American Information Superhighway or SuperJANET in Great Britain are of the highest importance not because ultimately they will resolve all our problems of management, but because they can restore once again a better balance between real extent and potential access to the encyclopedia. We should keep in mind that closing the gap between the two factors is impossible, as the fundamental lack of balance is rooted in the very nature of the system of knowledge. We shall always lay slightly behind the growth of organized knowledge. ^{5**}

5) Knowledge accessible > knowledge manageable

This is the problem of "infoglut", and concerns the degree to which retrievable information can actually be managed. Compared to our present age, any other moment in the history of thought has been a context of shortage of data. This led in the past to a rather voracious attitude towards information. Today, we face the opposite risk of being overwhelmed by an unrestrained, and sometimes superfluous, profusion of data. The old heuristic rule "melius

abundare quam deficere" has become self-defeating. A critical form of censorship is therefore a precondition for the individual mind to be able to survive in an intellectual environment, in which exposure to the human encyclopedia is greater than ever before. Plato and Descartes thought that knowledge could be compared to the food for our mind. If this is the case then today for the first time in the history of thought, we definitely need to learn how to diet. Without a new culture of selection, and tools that can help us to filter and refine what we are looking for, the Internet will become a labyrinth which researchers will either refrain from entering or in which they will lose themselves. In order to avoid both extremes, one can only hope that the care exercised today to avoid damaging the integrity of knowledge or losing information during the process of input — that is during the conversion of organized knowledge into a digital macrocosm — will soon be paralleled by equally close attention to the process of output — that is, efficient and economical ways in which we may select and retrieve all the information we need from the new encyclopedia.

6) The digital parricide

The global network may cause the loss of information on paper. Some libraries are already destroying their records on paper after having replaced them with OP AC catalogues. This is unacceptable, as would have been the practice of destroying medieval manuscripts after an *editio princeps* was printed during the Renaissance. We need to preserve the sources of information after the digitalization in order to keep all our memory alive. After all, the new encyclopedia is still too unstable to be trusted as a final repository. The development of a digital encyclopedia should not represent a parricide.

7) No epiphany on paper

Because increasing sectors of the new encyclopedia will remain forever digital, access to the network will have to be universally granted in order to avoid the rise of a new technological elite.^{8*}

8) The new language of the encyclopedia

IT is the new language of organized knowledge, therefore elements of IT will have to become part of the minimal alphabetization of any human being, if the freedom to information has to remain a universal right.

9) Intellectual space or polluted environment?

Because the Internet is a free space, where anybody can post anything, the Internet can become a dumping ground for any sort of intellectual rubbish, and organized knowledge could easily get corrupted or lost in a sea of junk data. In the book age, the relation between writer and reader is clear and mediated by a number of sociological, cultural and economic filters. One may be unhappy with it, but we know how it works and, after all, such filters do provide some positive selection. On the Internet, when there is any distinction at all, the relation between the producer and the consumer of information is direct, so nothing protects the latter from corrupt information. Now there is a lot to be said in favour of the free exchange of information on the network, and I do believe that any producer of data should be free to make it available online as he or she wishes. But I am also convinced that every user should possibly be protected from corrupt knowledge by an intermediary service, if he or she wishes to be so. 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. In spite of what many people are claiming nowadays, direct access to knowledge is on a temporary stage. In the future reliable intermediary services will become fundamental.¹⁰

10) Decentralization vs. fragmentation

This is probably the most pressing issue on our agenda. By converting the encyclopedia into an electronic space, we are running the risk of transforming the new body of knowledge into a disjointed monster rather than an efficient and flexible system. In the past, inadequate efforts have been made to forecast how the human encyclopedia was going to be affected by the binary revolution occurring in the management of information, and so promote changes that could be found more appropriate in order to meet the rise of new intellectual demands. The consequence has been that the Internet has developed in a very chaotic if dynamic way, and today it suffers from a regrettable lack of global organization, uniformity and strategic planning. While we entrust ever vaster regions of the human encyclopedia to the global network, we are also leaving the Internet itself in a thoroughly anarchic state. We seem to be unaware of the consequences of what we are doing. Efforts at coordination are left to

occasional initiatives by commendable individuals, or to important volunteer organizations, but all this is still insufficient to guarantee that in a few decades organized knowledge will not be lost in a labyrinth of millions of virtual repositories, while energies and funds are wasted in overlapping projects. The Internet has been described as a library where at the moment there is no catalogue, books on the shelves keep moving, and an extra lorry load of books is dumped in the entrance hall every hour. Unless it is properly structured and constantly monitored, the positive feature of radical decentralization of knowledge will degenerate into a neo-medieval fragmentation of the body of knowledge, a fragmentation that in turn will only result in a virtual loss of information. Already it is no longer possible to rely on the mere speed of our networked tools (NIR) to browse the whole space of knowledge, and collect our information in a reasonably short time. If global plans are disregarded or postponed, and financial commitments delayed, the risk is that information may well become as easy to find on the network as a needle in a haystack. Some people in the past have compared the invention of the computer to the invention of printing. I have argued that the comparison can be very misleading, at least in one important sense: the appearance of the printed book belongs to the process of consolidation and enlargement of our intellectual space, whereas the revolutionary character of Information Technology has rested on making possible a new way of navigating through such a space. But there is another sense in which I would be inclined to accept a substantial affinity between the two phenomena. For in the same way as the invention of printing led to the constitution of national, copyright libraries that would coordinate and organize the production of knowledge in each country, so the Internet is in need of an infrastructure of centres which, through their coordinated efforts, may fulfil the following five tasks:

- 1) guarantee the stability, reliability and integrity of the digital encyclopedia;
- 2) provide constant access to it without discrimination, thus granting a universal right to information;
- 3) deliver a continually updated map to the digital universe of thought;
- 4) expand the number and quality of primary, secondary and derivative resources available online, especially those that won't attract commercial operators;
- 5) support and improve the methods and tools, whereby the

encyclopedia is converted into a digital domain and then networked information are stored, and accessed, retrieved and manipulated.

I hope that what I am suggesting won't be misunderstood. I am not advocating the creation of some international bureau for the management of the Internet, a sort of digital Big Brother nor do I have any wish to see national organisms take control of our electronic new frontier. Such projects would be anti-historical, contrary to the fundamental rights of freedom of communication, of thought and information, and moreover they would be impossible to realize. Far from it, I believe in the complete liberty and refreshing anarchy of the network. What I am suggesting is that the Internet is like a new country, with a growing population of millions of well educated citizens, and that as such it does not need a highway patrol but it will have to provide itself with an infostructure like a Virtual National Library System which could be as dynamic as the world of information, if it wants to keep track of its own cultural achievements in real time, and hence be able to advance into the third millenium in full control of its own potential. It is to be hoped, therefore, that institutions all over the world may soon be willing to take full advantage of the new technologies available, and promote and coordinate such a global service, which is essential in order to make possible a really efficient management of human knowledge on a global scale.

Today we are giving the body of organized knowledge a new electronic life, and in so doing we are constructing the digital heritage of the next millenium. Depending on how we meet the challenge, future generations will consider us as new Pygmaliions or as old Frankenstein.

Luciano Floridi (Floridi@vax.ox.ac.uk)

Copyright (c) 1995 by Luciano Floridi

Permission is granted for reproduction of this document in any medium, but only in whole, for non-commercial purposes, and as long as appropriate credit is given to the author and this notice is always included.