Profligate Practices —
The Role of Information Technology

in the Debasement of Knowledge

David Vernon
Deparment of Computer Science
Maynooth College
Ireland

Introduction

Richard Feynman, one of the most influential physicists of the late 20th century, once wrote that 'Poets do now write to be understood'[1]. Perhaps; perhaps not. But sometimes poets manage to express truths in a remarkably lucid and resonant manner. Certainly, when T.S. Eliot penned the words[2]:

> 'Where is the wisdom we have lost in knowledge?
> Where is the knowledge we have lost in information?'

he illuminated a disturbing but undeniable truth that there is a qualitative difference between wisdom and knowledge, and between knowledge and information, and, unexpectedly, that not only are the latter insufficient for the former but that their presence can perhaps inhibit the former.

If Eliot begs the question, then we must surely attempt an answer. I wish to contend in this paper that the loss (or not, as the case may be) of knowledge in information is totally dependent on the human context or activity and, indeed, that information technology is acting as a powerful agent of this loss. I will argue that this need not necessarily be the case and I will suggest a number of ways in which this tide of knowledge debasement can be stemmed.

The basis of my thesis, which takes Eliot as its ground, is that there is indeed an important distinction to be made between data, knowledge, and wisdom. Few would argue with the distinction between knowledge and wisdom, but the difference between data and knowledge is often misunderstood. Worse still, I would contend (and will argue) that the term information is used ambiguously as a synonym for either data or knowledge (or something in between the two). Consequently, in order to avoid confusion, I wish to make one important substitution of words at the outset: I will use

---

[1] *The Quantum Universe*, T. Hey and P. Walters, Cambridge University Press, 1987.

[2] T.S. Eliot, *Choruses from 'The Rock'*.

the term 'data' in place of Eliot's 'information' and we will reintroduce the term information later on in an unambiguous manner.

Knowledge and Data

Let us begin our scrutiny of these words with 'knowledge'. All knowledge is known. That is, it is known by a knower: it is of necessity grounded in the experience of a knower. There is an inalienable human aspect to knowledge: it reflects our understanding of reality. If we remove the human knower, we reduce knowledge to mere data: a collection of ungrounded — but perhaps valid — facts. Facts, such as π being 3.14159 (a fact that is meaningless to anyone who is not acquainted with geometry), facts such as the statement that Lugnaquila is 165 degrees south south west of Mullagcleevaun mountain in Wicklow (a fact that is not only useless but also dangerous to a hill-walker navigating by compass and not knowing whether the bearing is with respect to grid north, true north, or magnetic north and not knowing that there is a nasty cliff in Glenmalure directly of this bearing).

Raw facts - data - are simply statements that an individual takes axiomatically to be true; in general they require no corroboration with anything outside the experience of the individual other than that they form a consistent part of some otherwise axiomatic system. Knowledge, on the other hand, differs fundamentally. Knowledge has as its ground, not an individual, but a group of individuals (or a society, in general). Knowledge is a reflection of an agreed consensus which necessitates dialogue and the establishment of a common set of semantics amongst the individuals sharing the dialogue and grounding the knowledge. Knowledge is very closely tied to the creation of an epistemology amongst the members of a society, the establishment of a consistent and faithful world-view. There is a very great difference between a deluded individual 'knowing' he is a great musician and our (collective) knowing that O'Riada was a great musician. Whilst I don't wish to make anything of it in this paper, it is perhaps worth noting that the transition for knowledge to wisdom requires one to go beyond even this, to transcend the epistemology of the society specifically by invoking a value system — which, importantly, is not fact-based — that constitutes a different orthogonal domain of discourse: wisdom is that by which society becomes humanity.

In essence, then, data is an epistemologically shallow term whereas knowledge is epistemologically deep (and wisdom is value-laden), and, at a somewhat trite level, what distinguishes knowledge from data is our understanding — our know-how — of how we and others can use the data to accomplish some useful task.

Let me recount a story to help clarify the distinction.

Thirty years ago, an encyclopedia sales-person knocked at our door on a wet November evening. Invited in, it was not long before the sales pitch had been made and our family were offered 'free' a magnificent set of twenty tomes, replete with marvellous celluloid overlay sections rendering various parts of human anatomy, and

detailed facts on every possible subject, provided we 'subscribed' to the year-book for a set number of years. Though I am convinced now that we couldn't afford them, the offer seemed too good to miss and the sale was closed. Shortly afterwards, we took possession of a massive collection of volumes, any one of which I was barely able to lift. Up until then, our family had relied on a more modest set of encyclopedias, the Arthur Mee edition, which we regularly consulted and which were read by adults and children alike. They were used for homework and to resolve arguments. When the new encyclopedias arrived, the Arthur Mee set was relegated to a back shelf and twenty black and gold volumes stood imposing and proud on the main book-shelf. I can remember them being consulted only once in all the years I lived at home after that. In retrospect, the reason is clear: we had traded up from a much-used if somewhat out-dated collection of facts to a modern glitzy exceedingly-detailed and barely-accessible collection. The amount of data on our bookshelves had increased some ten-fold to a profligate extent and our use of it all but disappeared. The quantity of data available to us had exploded but our ability to exploit it had plummeted. The essential common-usage, the mutual dialogue and reference to this data, had disappeared, and with it the growth of knowledge (at least from this source) in my home was killed forever.

The Role of Information Technology

The present convergence of communications technology and computer technology which is, I believe, bringing about an almost equivalent profligacy of data within individual organizations and across the globe. The individual databases of the '70s and '80s are now inter-linked, creating massive distributed repositories of data. These repositories are manifest as inter- and intra-institutional distributed databases, as local-area and wide-area network-based client-server systems, and, of course, as that anarchic repository itself, the world-wide web. Information technologists are keenly aware of the problems caused by the bridging of these historical data-islands and by the creation of these often poorly-structured repositories. They have given rise to new endeavours, such as data-mining and intelligent search agents, which are designed to find useful data-sets in the vast tracts of facts and figures.

I now come to the central part of my thesis. It is this. Information technology is giving rise to a massive increase in the amounts of data, in the home, in schools, in universities, in organizations of every size and type, with data being drawn from local, national, and international sources, through the media, the internet, CD-Roms, on-line databases. But the level of organization of this data, and even more importantly, the amount of dialogue on this data is either static or decreasing. In any case, the relative amount of organized, ordered, and human-centred knowledge is diminishing significantly. We have access to more data — facts — than every before but it is not at all clear that we know how to make use of these facts. Our knowledge base is being drowned in data.

In all of this I have studiously avoided referring to information. Let me know explain why. Eliot uses the term in the same sense as I have used the word data, as a form of degenerate or shallow knowledge. On the other hand, most people today, and

especially information technologists, speak of information and assume an equivalence with knowledge as something which conveys, and is grounded in a common epistemology, a shared understanding.

This is the difficulty with the term: it is inherently ambiguous and can be used as a synonym for either knowledge or data. I would argue, however, that there is a very strong drift - indeed a rip current - toward the lower end of the spectrum of information as data. Why? Because of information technology which by definition deals with data and not knowledge. Information technology can facilitate the dialogue, discourse, and mutual understanding by which data becomes knowledge but of itself it does not deal with it. Consequently, we become more and more used to associating information with debased knowledge rather than true knowledge.

Does this necessarily have to be so? Is information technology necessarily a degenerative influence on knowledge? Is there any way in which information technology could work the other way: as a positive influence in the creation of knowledge rather than as an agent of its debasement? I believe that the answer is 'yes'. If we accept that mere facts become knowledge when they are organized as an organic body of inter-related concepts which people know how to use, and which form the pivot or reference point in a process of human understanding of some particular subject or other, then the resolution requires two critical components.

It requires (a) the organization of the data and (b) human interaction. It is here that we encounter the great paradox in all this: communications. On the one hand, as I have argued, communications, which together with computer-based information processing systems form the engine of information technology, are the cause of our profligate practices in broadcasting (data-centred) information. On the other hand, communications technology can, in principle at least, facilitate the requisite human-human interaction. I hasten to add that they also facilitate human-computer interaction which is the norm today but this is insufficient. The need is for human-human interaction. By the way, e-mail or news groups rarely if ever approximate the required degree of human-human interaction. The time scales are much too long and the mode of interaction much to onerous to bring about the emergent understanding normally associated with human discourse.

But there is reason to believe that computers and communications technologies could provide a forum which would allow people to work symbiotically with colleagues, within the institution and world-wide in a cooperative organizing information (knowledge) centred environment. But it requires an explicit understanding that the benefits which can arise from the deployment of information technology are leveraged not in the domain of technology but in the organizational domain through the improvement of present work-practices, through the introduction of new, less profligate, work-practices, and especially through the effective organization of information so that it achieves and reflects the higher order of human-centred knowledge, with all that it entails, rather than the featureless tracts of data-based wastelands.

The Possible Evolution of Information

So what might the future hold? It is clear that, whether we like it or not, technology will continue to be an increasingly important influence on the way in which we work and relax. The information society is still in its infancy and, such is the present rate of change that, whatever the future holds, it is guaranteed to be different to our present situation. But, for the sake of argument, let us assume one of two scenarios:

1. we continue along the present path, creating ever-larger repositories of data, and easier and more wide-spread access to this data.

2. we recognize the reality of our present debasement of knowledge, drowned in our oceans of data, and we take steps to rectify the situation.

In the first case, we can expect to see the present tendency to produce 'bloat-ware' to continue and escalate. Bloat-ware is a pejorative term which is used to describe software systems which are extremely inefficient in terms of their size and utilization of hardware (such as memory). From a software development point of view, bloat-ware is the epitome of profligate practices in information technology. They are typically the product of repetitive version releases which offer some incremental functionality at a disproportionate cost in terms of the required speed and size of the host computer platform. It does not take much reflection to find examples of bloat-ware in, for example, the world of word-processing or operating systems.

The major challenge to the world-leaders in bloat-ware is the recently much-vaunted network computer which access its application software as well as its data directly, and in real-time, from the internet and then runs that software on a processor dedicated to the language in which the software is written. The most-often quoted examples of this new type of computing scenario are Sun's Java language and their Java workstations and the NetPC being promoted by Larry Ellison from Oracle. It is not yet clear that network computers will supplant the traditional PC with its massive local hard-disk storage, but it is highly-likely that they will be a force to be reckoned with, especially in the expanding intra-net market.

Since the availability of data will continue to explode, we will see 'intelligent' data mining applications becoming common-place (and, indeed, essential) filters on our access to this data. Some will be based on the familiar query forms which typically use keywords to access often poorly-indexed information; the difference in this instance is that they will use adaptive learning algorithms to tune the search both to the capability (or preferences) of the user and also the data-repositories which yield successful searches. These intelligent data-gathering agents use either neural-network systems or more conventional statistical models to effect their learning ability.

It is inevitable that the level of networking, and the number of people connected to these networks, will continue to escalate. There are limits to growth but they haven't been reached yet and it would be brave person who would place a bound on them at this point in the evolution of networked communication systems. The increased presence of networks will be accompanied by an increase in network bandwidth, *i.e.*

the data-carrying capacity of a network. Even now, organizations are specifying 100Mbit per second fast Ethernet technology as the minimum standard for new networking infrastructures and there are signs that the (not so new) ATM (Asynchonous Transfer Mode) networking is finally settling down and will be delivering the promised 622Mbits per second bandwidth cost-effectively in the near future.

The growth in the use of analogue and digital GSM cellular telephone is obvious and the ubiquity of this (often annoyingly intrusive) instrument is evident on every street-corner, cafeteria, pub, train, car, golf-course, tennis-court, and swimming-pool. Lamentably, they have even been seen wielded by hill-walkers in Wicklow! The move toward the integration of cellular telephony with very small light-weight low-power computers in well underway and will continue, offering more and more people easier and easier access to more and more data, anywhere and everywhere. And so it goes.

In all of this scenario, we see a continued increase in volumes of data being traded between human and computer with emphasis on faster, cleverer, and more ubiquitous human-computer interaction. It is a scenario driven by technology and centred on technology. Its evolution will, I believe, be typified by an inexorable increase in the volume of data and an equally inexorable decrease in the organization of that data; see figure 1. Ignoring all the signs and lessons, we will attempt to deploy technology to compensate for the lack of organization.
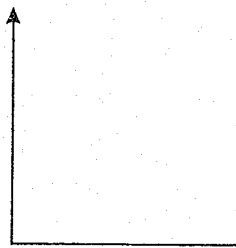

Figure 1: the evolution of data volume, organization. and validation in scenario 1 where we continue to deploy information technology as we do at present.

Turning our attention to the second scenario, where we recognize the reality of our present debasement of knowledge, drowned in our oceans of data, and where we take steps to rectify the situation, I would have to say that, like it or not, it will probably involve all of the technological developments outline in scenario 1. However, and it is an important 'however', there will be a fundamental shift away from technology-centred development to human- and organization-centred development which will place a much greater emphasis on human-human interaction, rather than human-computer interaction and albeit mediated by computer and communications

technology, and which will promote the use of effective and efficient work-practices rather than the present profligate practices.

Very importantly, there will be a significant improvement in the organization of data, with classical principles of librarianship and data-stewardship finally being imposed on our information-repositories. I use the word 'imposed' intentionally for I believe that the present 'democracy' of the WWW and internet can only lead to an inevitable increase in 'information entropy' with less and less organization and order overall.

As part of this process, there will be a return to the traditional validation of 'publications'; that is, data or information will only be accepted for publication if it is deemed by someone other than it author to have some value.

Equally, there will have to be some form of enforced 'data decay' mechanism whereby information which is not used, or validated, decays with time and eventually disappears (see figure 2). This will act as a natural inhibitor to the almost uncontrollable growth of data associated with scenario 1.

There will be a much more pervasive presence of verbal, visual, and aural communication, in real-time. To satisfy the consequent and huge demand for communication bandwidth, we will see at the very least all of the scenario 1 developments in communications technologies and probably the common-place introduction of ATM networking, if not nationally and internationally, the at least over local areas and especially for intra-nets. Somewhat paradoxically, and notwithstanding the move toward natural computer-mediated human-human communication, there will also be a move to make it harder to communicate, in the sense that people will only engage in communication when there is a value to that communication. The present culture of responding to the daily deluge of e-mail almost immediately will change and longer, more conventional, time-scales for this type of communciation will become more normal.

In summary, scenario 2 will subsume scenario 1 and will place much greater emphasis on more effective and efficient human-human interaction. It is a scenario driven by humans and centred on knowledge, organization, and order. Its evolution will, I believe, be typified by a controlled increase in the volume of data, an increase in the organization of that data, and an increase in the level of validation of that data (see figure 2).
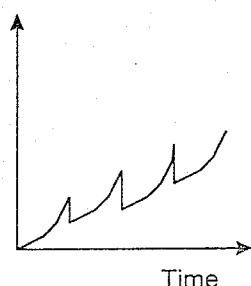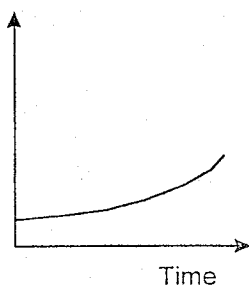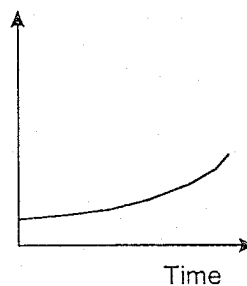
Figure 2: the evolution of data volume, organization, and validation in scenario 2 where we attempt to halt the present debasement of knowledge.

Whither Now?

Which of the two scenarios will prevail? It is impossible to tell; the information society is being borne on a information revolution which respects no trends and is manifest as an uncontrollable self-catalyzing process in the development of our society. But we are a society of humans — thinking, for the most-part rational, and reflective beings — and, whilst we many not be able to control or evolution, we can at the very least influence it. The influence we choose to bring to bear will depend on our recognizing the validity of Eliot's distinction between wisdom, knowledge, and information, on our attaching a value to the former, and our conscious will to ensure that technology serves mankind and, in the process, advances our knowledge and humanity.

Acknowledgement

ACADEMIC AND SPECIAL
LIBRARIES SECTION
LIBRARY ASSOCIATION OF IRELAND

INSTITUTE OF
INFORMATION SCIENTISTS
(IRISH BRANCH)

INFORMATION MILLENNIUM:
THREAT OR PROMISE?

FRIDAY, FEBRUARY 14th 1997

IRISH FILM CENTRE
6 EUSTACE STREET, DUBLIN 2

Sponsored by:
H.W. WILSON, HODGES FIGGIS, SWETS (UK) LTD., ITELIS Ltd.