Business Computing: the Second 50 Years

The Guildhall conference for business leaders

The Guildhall, London

5 and 6 November 2001

Conference Proceedings

Knowledge Management and Digital Information

John M Ashworth

Former Chairman of the British Library

Knowledge Management and Digital Information

John M Ashworth

First, an apology. The programme describes me as the Chairman of the British Library which was quite accurate when the programme was printed but my term of office finished on 31 August and so I am strictly the ex-Chairman. However, I will be basing what I want to talk about this morning largely on what I have learned and worried about over the past five years when I was the Chairman of the Board of the British Library.

Let me begin by talking about libraries. I believe that they are THE key institutions of civilized societies. The reason is simple – without some systematized way of conveying from one generation to another previously accumulated wisdom, knowledge and experience there is no possibility of successive generations learning from the experiences of their predecessors and without that there is no possibility of cultural evolution. If we had had to depend simply on oral traditions and memories then our social development would have been constrained by the rate at which biological evolution proceeding by Darwinian selection of the capacity of our brains to remember and recall occurred. Clearly our amazing capacities to remember and recall suggest that for several millions of years of our evolutionary history that process did go on but for at least the last 50,000 years this capacity has been supplemented by records, first visual then written of one kind or another. Thus uniquely amongst living things we have complemented our Darwinian biological evolution by Lamarckian cultural evolution – that is by one generation inheriting from another some at least of the wisdom and knowledge that that generation acquired. However, not even I would wish to claim that inscriptions on the rocks could usefully be described as a "library".

The first libraries that we know of that resemble ours were those built by the classical empires of China and Rome. I do not want to say much about the Chinese Imperial library – it is a fascinating and rather specialized area of scholarship all of its own – but we would do well to remember that it was certainly the first and arguably the best in the world until the C18th century. We have, for example, catalogue records for manuscript scrolls that are now lost but which were in the collection at least two thousand years ago. The problem was that successive dynasties tried to edit the historical record to serve their own contemporary agendas by systematically rewriting or destroying scrolls in the library but often then forgot to edit the catalogue records appropriately. A typically bureaucratic error which has provided happy employment to generations of scholars of Chinese & Western scholars.

The classical world of Greece and Rome was always a lot less rigidly controlled and centralized than Imperial China and nothing like the Imperial Library was created. There were however, a number of libraries like that of Alexandria that attempted to collect copies of all literary and other records that the educated felt were significant. Often the State helped in this process – there is evidence for example that for many centuries every ship that docked at Alexandria was searched for scrolls and if the library did not have a copy of them then a copy was taken as part payment of the normal harbour dues. Deliberate collection policies and funding were also provided by the authorities in order to enhance the collection. However, these classical libraries shared with the Imperial Chinese one two great problems. The first of course was that they were based on scrolls – papyrus ones in the case of Greece and Rome and paper in the case of the Chinese –

and the second that each scroll was individually copied by a scribe. Scribes can and did make mistakes and an inordinate amount of scholarly time was spent in trying to reconstruct original texts from corrupt copies. Also, of course, scrolls are a nightmare to catalogue and store. Printing – invented by the Chinese probably in the eighth century AD – partially solved the first problem but the second required the invention of a new format: the book, or strictly the codex – a manuscript in book format. We know that this happened in the eastern Mediterranean in the first century AD but we do not know exactly how or why it happened. What we do have is tantalizing hints that it was connected with the decline in the pagan religions and the rise in the new Christian religion.

Analysis of the thousands of papyrus fragments found in the rubbish heaps of Egypt has shown that whereas pagan texts can be found in both scroll and codex formats Christian texts are found only in codices. One of the earliest and most intriguing is a fragment of the Gospel of St Thomas - a collection of stories and sayings of Jesus Christ of which some 60% can also be found in the canonical gospel books as we know them in our Bible. A few leafs of this are currently on display in the British Library Ritblat Gallery.

Quite why there should be this marked aversion for early Christians to use scrolls is unknown but it may have something to do with the need for concordances to be made between the different gospel texts – certainly it is very much easier to do this with codices than scrolls.

The codex is a magnificent invention and has advantages other than that of helping with concordances. It is more easily stored than a scroll and it is easier to find things in; lends itself to cataloguers because it can have simple metadata of chapter/paragraph form keyed to pagination. I am sure that classical scholars used scrolls more efficiently than we do but the total replacement of the scroll format by the codex by about C5th at least in the West shows that the convenience of the codex was recognized by all. Outside the Mediterranean the use of parchment rather than papyrus also probably helped with the domination of the codex since thick parchment is easier to use in sheets rather than rolls.

The other great technical solution to the problems of classical libraries was as I have said the invention of printing. This mechanized the problem of text reproduction and also seems to have been associated with the introduction of a new religion – in this case Buddhism into China. Buddhism arose in North Eastern India and was taken to China by missionaries who first went West, through what is now Pakistan then North to the Silk Road round both sides of the Taklaman desert and then to the town of Dunghua, the key entrepot in the eighth century for the Silk Road trade. We can trace the centuries long course of Buddhism as it traveled this tortuous route and the oldest printed texts [by woodblock technology] that are known come from caves in the cliffs above Dunghua. There seem to have been printing factories established there to reproduce sutras and other sacred Buddhist texts for the use of the missionaries who were going on into China. If you go to the BL you can see in neighbouring display cases in the Ritblat Gallery the oldest printed text yet discovered, a version of the Diamond Sutra from Dunghua and also the near contemporary manuscript text of the Lindesfarne Gospels – the one produced as part of the project of the conversion of the English to Christianity and the other part of the project of the conversion of the Chinese to Buddhism.

Putting these two innovations of the codex and printing together to produce printed books as we know them of course did not occur until the C15th in Germany. The first

western printed texts [using moveable type technology] were indulgences, with spaces conveniently left for the insertion of the purchaser's name, but the rapid development of printing was greatly stimulated by the need, thanks to the Protestant reformation, of cheap and accurate gospel texts. Other printed materials interestingly were not usually produced in the form of books but rather in the form of loose leaves which the purchaser then had bound by his own binder according to his own taste. If you want to see the result of this go into the BL and look up at the central tower which houses the Royal library of KGIII 60,000 volumes of C18th books all leather bound with sumptuous bindings – decorator books to die for and still a collection much used by scholars. The speculative production of non-religious books and the development of the book trade progressed slowly but by the C18th century were much as we know it today and it was no accident that comprehensive libraries as we know them today can all trace their origins back to that period.

The British Library [BL] was originally part of the British Museum. The Museum was a quintessentially Enlightenment project to assemble all of human knowledge under one roof. Intellectuals like Banks, Sloane and Cotton gave their collections of books and the "cabinets of curiosities" that all C18th gentleman collected to the Nation in order to begin the collections of the British Museum. Over time the "curiosities" have been spun off and form the basis of the collections of the Natural History, Science and Geological Museums in South Kensington. The books were spun off to form the BL in 1972 and the construction of the new building for the Library at St Pancras marks the latest phase in this three hundred year process. The collection of ancient statues, prints and archaeological items remain with the British Museum.

I have spent some time on this rather simplistic canter[1] through the history of libraries and the book because I want to draw one very important lesson from it and one rather alarming conclusion. The lesson is that innovations in the format and storage of information, knowledge and wisdom seem to have been associated with the occurrence of new intellectual movements, either religious like Christianity or Buddhism, or intellectual like the Enlightenment. This steady progression over two millennia has given us the intellectual infrastructure that we have all grown up with. The alarming conclusion is that we have just thrown over if not all of those key innovations then certainly most.

By developing screen based text handling systems we have of course reinvented the scroll and thrown away the codex. With this we have rediscovered all those metadata standardization problems for which the codex was the solution. Despite the valiant efforts of librarians in international conference after international seminar and the enormous step forward that the Dublin Core protocols represent there is still no satisfactory system of metadata standards that satisfy the requirements of librarians. Cataloguing is thus still a very laborious and labour intensive activity and still far too parochial. The IT equivalent of the library catalogue the search engine is still worryingly temperamental and inaccurate. Last year the BL did an experiment and tried searching the patent literature – one of the most digitized of any specialized body of text – by search engine and by librarian. At best only slightly more than half of the relevant literature that was found by our patent librarians was found by the search engines which might make one a little worried by the increasing practice of patent agents of relying solely on digital sources of patent information. We are going to have to do better than that and quickly.

1 For a fuller treatment see Lionel Casson "Libraries in the Ancient World (New Haven, CT and London Yale UP, 2001)" and E. Eisenstein "The Printing Revolution in Early Modern Europe (Cambridge UP, 1983)

To give you another example the scientific basis of the wobble that caused the Millennium bridge across the Thames to close as soon as it was open and the formula that underlies the way of fixing the problem were available to the builders, in English, in a highly reputable journal that the BL takes before the construction of the bridge was begun. Yet they proceeded as if they did not know of its existence. Libraries are supposed to prevent that sort of mistake from being made. Now that science, technology and medical journals are available in electronic formats it should, in principle, be possible to catalogue and index text and other material automatically but it is not. We are in much the same positions as some of the scholars who Umberto Echo described in his wonderful book "In the Name of the Rose" we often suspect that what ewe want to know exists we suspect we know where it is but due to the difficulties of the catalogue we cannot find it reliably.

The British Library takes about 50,000 journals of which some 6,000 are available in electronic form. We store them in a digital store & have or shortly will have the technical means of receiving, processing & transmitting copies automatically. Currently, however, we are forbidden by the copyright holders from doing that. They insist that we produce a printed version even if we then fax or email it to the customer. Not for the first time it is the social or legal infrastructure that is holding up technically possible productivity improvements. But actually such delays might be very useful because they give us time to work out how to manage digital libraries with the efficiency & reliability that we have, over the past 500 years, come to expect with traditional libraries. It will be fascinating to see in the next few years whether the evolution of the publishing industry & the copyright legislation & the frantic efforts of the standard setters keep more or less in step during this process. I suspect that they will not because the underlying IT technology will continue to evolve as it has in the past 50 years & that will produce new problems as fast as the old ones are fixed. The reason for this lies in the way that the technology has evolved.

This conference focuses on business computing but business has rarely, if ever, been in the radical innovation process in IT. The first programmable machine was, after all, Colossus, developed to crack the German military codes produced by their Enigma machine & the first users of the successors to Colossus were the military and administrative arms of government – not business. The high energy physics and related communities and the atomic weapons & rocket programmes of the US, Soviet Union & the UK drove the development of the hardware in the 1950'2 and 1960's & the Pentagon's desire to have a command & control system that would survive a nuclear first strike gave us the internet. The www came out of the needs of the high energy physics community at CERN – arguably the most valuable "spin-off" of all times. None of these developments were part of the market driven business sector of our economies. The business community struggled to adapt the kit that the military & research communities had developed for their own uses & pretty poor shots they made of some of it. I have always been puzzled by the complaints of the business community that it has been so difficult to develop commercial businesses on the web – since it was not developed with that in mind it is not surprising to me that they have not proved very fit for commercial purposes.

This pattern of innovation has continued. The high energy physics communities still demand ever more storage and ever faster machines & they have been joined by the molecular biologists. The need to acquire, store, access & manipulate the billions of DNA base sequences that make up the genomes of living organisms is bad enough but genomics is at least a 2D problem. Proteonomics – the study of the proteins for which those DNA sequences are the code – requires the representation and manipulation of 3D images and is

even more demanding in computing terms. The social scientists have also joined in with their requirements to handle ever larger data bases and models and even the humanities scholars now wish to put those data bases together with GPS data and develop a new kind of historical analysis. One of the most interesting projects that I have come across recently was the Electronic Cultural Atlas Initiative [ECAI] managed my Prof Lancaster at Berkley University, California which tries to do just that. These demands of the research communities are necessary – they have given us SuperJanet for example – but they are not sufficient to drive the technology forward.

There is a need for massive injections of risk money to drive the research that will underpin the next generation of devices and realistically this will have to come from the taxpayer, at it always has. Traditionally it has been the military and security needs of nation states that have provided the required stimulus and right on cue, we have just decided to declare a "War on Terrorism". If this is to be a serious campaign then we know from the skirmishes that we have fought in Northern Ireland & elsewhere that it will take at least 30 years, certainly as long as the Cold War and that got us from Colossus to Microsoft XP. It will only be won by our intelligence services deploying sophisticated surveillance & monitoring systems that will contain the terrorist threat whilst political and intellectual/ideological pressure undermines the will of those who are our adversaries. In other words it will be a knowledge management war & whatever kit that is needed will be provided. I think the investment needs of those who wish to make a business out of broad band communications will suddenly be met as a spin-off from whatever it will take to provide real time transmission of data and pictures of those going through airport checkpoints to central data bases etc etc. The librarian's problems with the Dublin protocols are going to be solved pretty quickly, I suspect and the funding needs of those academics struggling with ECAI are shortly to be met quite handsomely I would predict.

But remember – we have been here before. We cannot be sure of course of the relationship between cause and effect of events centuries ago. But the switch from scroll to codex certainly facilitated the rise of Christianity and thus destroyed the confidence of pagan Imperial Rome. The invention of printing equally certainly facilitated the spread of Buddhism into Confucian China – altering the development of that culture just as profoundly. The Protestant revolution in Western Europe altered for ever our own intellectual horizons & the Enlightenment is inconceivable without the encyclopaedias and comprehensive libraries and museums that movement both required and inspired. We are of course living in what historians will call I suspect, the end of the Age of Enlightenment/Reason. I certainly see no reason to believe that the present revolution in knowledge management, which is just as profound as the other three that I have used as examples, will be any different. What religious or intellectual movement will benefit from or be facilitated by these changes in technology is another matter of course – and what is cause and effect is equally problematic as always. Personally I would not put money on Islam but some amalgam of environmentalism and New Age Greenery with Buddhism [the world's fastest growing religion at present] certainly looks a possibility. I hope that I live to see it if it does happen for what I am sure about is that it is these revolutions of the mind that are ultimately the ones that really change our lives not those of technology.