Editorial: Big data archives

Luke Tredinnick and Claire Laybats

In 1965 Gordon Moore – co-founder of the microchip manufacturer Intel – first made the prediction that subsequently bore his name. Moore observed that the density of components on integrated circuits had tended to double every year, and predicted that this would continue for at least the next decade. Ten years later Moore revised-down his estimate to a doubling every two years. Because of a broad correlation between the two, Moore's Law as it came to be known is usually stated in terms of increasing computer processing power. It is the most famous law of computing; almost everybody who works with technology has encountered it in one form or another. Nevertheless the ramifications of that prediction are unfolding today.

Moore's law is an example of exponential growth – a function where the rate of change is proportional to the current value of the function. One of the characteristics of exponential growth trends is that they are difficult to visualise. They start slowly but rapidly accelerate beyond all comprehension. it perhaps no longer seems shocking that the entire computing power behind Apollo moon was less than that contained in a modern smartphone. But Moore's law still has the power to unsettle us. There has been a trillion-fold increase in computing power since 1958. A modern smartphone is now broadly comparable to the Cray 2 Supercomputer produced during the mid 1980s, the fastest computer of its time. We carry in our pockets machines as powerful as those used in nuclear weapons research only thirty years ago.

Moore's law is in fact not inviolable. It broadly held until 2012, but over the past six years the rate of change in microchip development has begun to slow. Intel has for example stated that the doubling of power now occurs over a period of roughly thirty months rather than eighteen. The reason for this slowdown is that as more component are packed into smaller spaces it becomes incrementally more difficult and more expensive to shrink them further. There is a physical limit to the density of microchips, and as we approach that limit production begins to butt-up against fundamental physical laws. Most commentators think that Moore's law will remain broadly reliable until about 2025, but after then things become less clear. As a consequence it may be that we are approaching an age in which the rapid advancements in computing technology that have been a defining feature of the last fifty years begin to stall. The well of ever increasing computer power that drove economic development through the late twentieth and early twenty-first centuries may dry-up just surely as the oil fields of industrial production.

Moore's law is of course a well-known trend and its consequences have been widely extrapolated in sometimes extraordinary technological prediction such as for example the idea of the technological singularity (Shanahan, 2015). But there is another technological trend that is similar to Moore's law and that has similarly profound consequences for the ways in which we manage technology and information in the contemporary age and into the future. It is a trend that has drawn much less interest and no apocalyptical predictions. In parallel with the growth of computing power the density of data storage density has followed a similar trend, and its consequences of this are almost as profound. The power of the microchip may drive our fantasies about the technological future, but it is the power of mass storage that in many ways both defines the way that technology is used today, and describes the kind of information rich society we are emerging into.

In 1967 a 1mb hard drive cost about $1 million. Today the cost of that storage is around $0.002 and the price is falling all the time. The trend in reducing data storage costs has been recognised for some time, but in 2005 Mark Kryder was featured in an article published in Scientific American that drew on Moore's Law and offered its own observation and prediction. That paper observed:

> Since the introduction of the disk drive in 1956, the density of information it can record has swelled from a paltry 2,000 bits to 100 billion bits (gigabits), all crowded in the small space of a square inch. That represents a 50-million-fold increase (Walter, 2005: 32)

Kryder's Law as it has come to be known described the exponential increase in density of data storage on magnetic disks doubling over a thirteen month period, outpacing the growth in raw computing power.

In fact declining information storage costs and increasing density of information storage media are trend that predates the invention of magnetic storage. Paul Conway (1996) has observed that the density of recording media began to increase in the 19th century and into the early twentieth century. Nevertheless as with the rise in computing power, it is the exponential nature of this trend that creates its extraordinary impact. We tend to think of the information age as being driven primarily by increasing processing power, but if anything it is our capacity to store and retrieve ever vaster collections of data that has allowed the influence of the microcomputer to penetrate every aspect of our lives. We live in an age of big data, not in the sense that big data sets are here, but in the sense that big data sets are becoming the norm. One reason for this is an interesting corollary to Kryder's law and its thirteen-month period: if you can afford to store an incremental data set today, then you can continue to afford it indefinitely into the future.

Today's tech giants depend as much on low cost data storage for their business models as they do on raw processing power and the effects of declining data storage costs are built-in to their business model. Google for example relies on an incomprehensibly large database that all but reproduced every webpage, every printed work ever published, every music recording ever made, and vast volumes of personal and corporate emails, photographs, sound recordings, videos, documents and files. Facebook is not principally a social networking service, but a vast database containing every interaction anybody has every had through any of its services. The volume of data involved in these services alone is phenomenal, and they represent just the tip of a titanic data iceberg that has been quietly growing thanks to the exponential reduction in price and increase in capacity of digital data storage. The fundamental driver of this model is not only that incrementally growing data sets are worth collecting and storing, but much more importantly that they can be stored indefinitely thanks to the ever-declining costs of data storage.

Yet that future of functionally unlimited and limitless data is faltering. Kryder's law has not proved as robust as Moore's. The rate of decline of the cost of disc storage is faltering, and for similar reasons to the faltering of Moore's law. This fact is not just of academic interest, but has real-consequences for data management and archiving. Until recently it was a reasonable assumption that if you could afford to store incremental data in the present time, then you could afford to retain it indefinitely, because the declining costs of storage would always keep pace with the growing data set. The cost of storing one billion transactions over one year would be broadly the same as storing ten billion transactions over ten years or one hundred billion transactions over one hundred years. But as the rate of decline in the cost of magnetic storage slows that assumption no longer holds, and the costs of storing incremental data sets begin to rise over time. Furthermore, as the data set grows the rate of cost increase also grows. The practical upshot of this is that the assumption that you can keep everything indefinitely with no ongoing financial consequences can no longer be made. In an economy that is increasingly producing and relying on big data sets, that adds additional uncertainty to future business processes.

Interestingly one of the ways in which this problem can be managed is through application of traditional information management approaches. Archive practices are increasingly being applied to big data sets. The threat of growing future storage costs means that the assumption that big data sets can be retained unmanaged is being replaced by more traditional retention policies and practices. This renewed interest in archiving practice is an example of how techniques and practices developed to manage paper collections are finding relevance in the vast digital data sets of the contemporary age.

But what implications do the looming end of Moore's Law and Krydor's Law have for the future growth of technology and the development of the information age? Are we facing the limits of computing power? Have we reached peak computing? Historical trends are of course not predictors of future events; the facts of these patterns in the past does guarantee their continuation in the future. In fact Moor's Law and Krydor's Law are both predicated on specific technologies and materials: the silicone chip and the magnetic drive. Indeed the growth of Flash storage has thus far largely mitigated the decline of magnetic disks. As new ways of approaching the challenges of the information age are adopted, the smooth exponential curves of the late twentieth-century may perhaps be superseded by broken curves describing periods of rapid development punctuated by periods of relative stasis. The pace of development may be slowing for now, but that might herald a new beginning in the advancement of technology rather than a looming end.

December's *Business Information Review*

The first paper in December's *Business Information Review* touches on a topic that reflects a familiar experience in the contemporary business environment: organisational restructurings and its impact on knowledge retention. Andrew Lambe, Fiona Anthony, and Jo Shaw write about the experience knowledge continuity and organisational memory during NHS England's organisational restructuring in 2015. The paper examines the approach taken by the Knowledge and Intelligence (K&I) team of the Sustainable Improvement (SI) team at securing organisational knowledge following the Smith Review of Improvement and Leadership Development in the NHS and its consequences, charting the stages in the migration of content, the development of new retrieval tools, and the development of a new knowledge service.

The second paper in this issue is the second part in our series looking at professional bodies and their impact on the library and information profession, the first of which was publish last issue. That paper featured an interview with Simon Burton of CB Resourcing and president elect of SLA Europe about the value of professional membership. In this issue Claire Laybats talks to Nick poole, Chief Executive Office of CILIP (Chartered Institute of Library and Information Professionals). Nick Poole has been responsible for developing UK and international standards for information management, advising agencies on digital programmes and leading partnership initiatives with a range of leading digital companies, and discussed in Career's Corner the value of professional membership and active professional engagement.

Our first research article for December is entitled *Integrating Social Media and Grounded Theory in a Research Methodology: A Possible Roadmap,* and explored ways of applying grounded theory to social media research. Grounded theory is an approach to research that involves generating theoretical frameworks in a systematic way from research data, and is intended to close the gap between theory and empirical research. Written by Mohanad Halaweh, the paper provides practical methodological guidance in a step-by-step approach to Social Media Grounded Theory methodology, which can be applied by researchers.

Paul Corney returns to the pages of *Business Information Review* to discuss the ISO Standard 30401 Knowledge Management Systems Requirements. The paper is a follow-up to Pauls 2016 article about the development of the ISO standard, and argues that it has the potential to be a game-changer for knowledge and Information management. Paul charts the development of the standard from the early 2000s through to its publication, and addresses some of the controversies and benefits.

Our final paper in December's issue is entitle The Ever Developing Role of the Library and Information Profession, and reports and reflects on research undertaken by CB Resourcing and CILIP into skills requirements in the Higher Education library and information management field. The paper, written by Claire Laybats, provides an interesting companion piece to our recently published Annual Survey on trends within the business sector, and the finding have broader implications for the ways in which library and information management is changing to meet the challenges of the twenty-first century.

Last in this issue is Martin White's final *Perspectives* column. For the last four years *Perspectives* has bought a quarterly reflection on new research and publications outside the field of Library and Information Management that has broader implications for the profession. This column covers topics including the readability of infographics, text mining, and the International Monetary Fund (IMF). December's *Perspectives* is the final column in the series that has charted an astonishing range of research and publishing over the last four years. Martin has been an invaluable contributor to Business Information Review through both *Perspectives* and his other contributions, and we wish him well for the future.

References:

Conway, P. (1996), Preservation in the Digital World, available at: https://www.clir.org/pubs/reports/conway2/index/ [accessed: 24th October 1028]

Shanahan, M. (2015), *The Technolgical Singularity,* Cambridge, Massachusetts : The MIT Press

Walter, C. (2005), Kryder's Law, *Scientific America,* 293 (2005): 32-33.