

## The No-Nonsense guide to The Digitization of the World

*‘Digitization is the process of converting information into a digital format. In this format, information is organized into discrete units of data (called bits) that can be separately addressed (usually in multiple-bit groups called bytes). This is the binary data processed by computers and devices like digital cameras.*

Both text and images can be digitized: a scanner captures an image (which may be an image of text) and converts it in to an image file, such as a bitmap. A software programme designed for optical character recognition (OCR) analyzes a text image for light and dark areas in order to identify each alphabetic letter or numeric digit and change each character into ASCII code (see Box 1).

Audio and video digitization uses one of many analog-to-digital conversion processes in which an analog (continuously variable) signal is changed, without altering its essential content, into a digital (multi-level) signal. A process of sampling measures the amplitude (signal strength) of an analog waveform at evenly spaced intervals and represents those samples as numerical values for input as digital data.

Digitizing information makes it easier to preserve, access, and share. For example, an original document may only be accessible to people who visit its physical location, but if the document content is digitized, it can be made available to people worldwide. There is a growing trend

towards digitization of historically and culturally significant data.

Digitizing creates what is called a ‘non-rivalrous resource’. From an economist’s point of view, digital information products are unlike the products of commercial manufacture (toothpaste and chocolate) which are consumed. In theory, they exist infinitely and can only be sold once unless, that is, they are ‘updated’ to become a new product.

Digital information products can be duplicated (copied) with complete fidelity to and without damaging the original. This is why protection devices such as encryption technologies are given such prominence and why so many battles are being fought over copyright and patenting.

### Background

The first digital computers were developed in the USA in the 1940s. Rapid developments concentrated on reducing size and increasing speed and capacity. Today’s computers use miniature integrated-circuit technology in conjunction with rapid-access memory. Computers are desk-top, lap-top, hand-held and will soon be ‘embedded’ in other technologies. They are connected by cable or infrared links to global communications systems, the Internet and ‘cyberspace’. Future generations of computers are expected to use forms of ‘artificial intelligence’.

Audio and video developments converged with digital technologies that evolved via the computer industries. Digital was quickly taken up for audio by Philips and Sony, which introduced com-

puter machines’ (computers that can solve problems and achieve goals in the same way that humans can) – and it is a small leap of the imagination to arrive at a digital replica or clone that has the exact physical and mental characteristics of a particular individual.

By the end of the 21st century, it may be possible to fabricate digital replicas of people and to invest them with a complete biological and social life-history. These replicas will be able to enter into dialogue about their lives and even to reproduce certain abilities, for example, playing chess or a digital piano. No soul – perhaps (see Box 3) – but every other human attribute.

### Convergence

The logic of digital convergence leads inexorably to digitization of the full range of all data (enabling that data to be used with increasing ease), integration of data handling systems, integration of different components in the chain, and the global reach of data, devices, content, communication and users.

If a portable device starts out as a hand-held diary or mobile phone, it quickly morphs into both and, in turn, becomes an electronic organizer/book/newspaper/communications centre. By adding biometric data, financial profiling, security details, it becomes an identity tag/health card/passport that begins to be a complete record of who we are and – given today’s sophisticated tracking devices – where we are and what we are doing.

One miniaturised device becomes a digital alter ego that both identifies and privileges those who have it. Those who do not have it, are the excluded. ■

patient’s skull to facilitate preoperative planning... Anti-terrorism experts are trained in virtual environments peopled with virtual hostages and virtual terrorists, to hone the quick response of the trainees... The employment of virtual humans is limited only by our imagination.’<sup>3</sup>

Or limited only by our technology. What if it were possible to create a virtual human being with all the attributes of the real? Little of the research that led to the Human Genome Project and the Visible Human Project could have been done without computers. The outcome of both projects was a complete digital blueprint of the human being. Couple this with work being done on artificial intelligence – the science and engineering of making ‘intelli-

### Box 3

The Massachusetts Institute of Technology has a Computer Science and Artificial Intelligence Laboratory (CSAIL) which seeks to understand and develop systems – living and artificial – capable of intelligent reasoning, perception and behaviour. Theologian Anne Foerst, who worked there as a research scientist, has commented that if it were possible to construct a machine with humanlike intelligence, she would regard it as having the attributes of personhood and dignity. It would have the right to search for its own soul. ‘Instead of insisting on a qualitative difference between us and the machines Artificial Intelligence will create, it seems more reasonable to turn the question around. Not reflections on “why a machine never can become like us,” but instead the question of what might be the conditions under which God would accept such a creature as God’s child.’<sup>4</sup>

same treatment. Her identity is unknown, but she, too, was reincarnated as a 40-gigabyte record. Both digital bodies can be accessed via the Internet.

The Visible Human data sets are designed for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a model for the construction of network accessible image libraries. According to the U.S. National Library of Medicine, ‘the Visible Human data sets have been applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical, and industrial uses by nearly 2,000 licensees in 48 countries.’

To all intents and purposes these are digital replicas of human beings, lacking only their corporeality, personality and memories to convince us of their real presence. But they are not alone. According to Joseph J. Beard, new technologies have enabled other kinds of virtual humans to be ‘nurtured in the womb of the digital computer’. They populate film, television and computer screens, but are not limited to the two-dimensional world. They may also appear as holograms (three-dimensional images) or in animatronic form:

‘Virtual humans have found steady employment in the entertainment field. Actor Robert Patrick’s digital clone performed as the liquid metal cyborg in *Terminator 2: Judgment Day*. That same clone was later devoured by a digital T-Rex in *Jurassic Park*. The blockbuster *Titanic* was peopled with virtual passengers who strode the digital deck of the digital Titanic... But virtual humans can do more than simply entertain. For example, virtual humans play a significant role in ergonomics. Engineers using computer-aided design (“CAD”) software employ virtual humans in equipment design to optimize the human-machine interface. In medicine, virtual replicas of real tumors are presented in the virtual replica of a





### Box 1

American Standard Code for Information Interchange (ASCII) is a character encoding based on the English alphabet. It represents text in computers, communications equipment, and other devices that work with text. Work on ASCII began in 1960. The first edition of the standard was published in 1963, a major revision in 1967, and the most recent update in 1986. It currently defines codes for 128 characters: 33 are non-printing, mostly obsolete control characters that affect how text is processed, and 95 are printable characters.

compact discs (CD) in 1982. Sony went on to develop the mini-disc launched in 1992 and Hi-MD in 2004. Similarly, video cassette recorders (VCRs) were replaced by digital video discs (DVD), holding up to 26 times more information than CDs, and by High Definition DVD (see Box 2).

Today, digital technologies are being used in preservation programmes in libraries around the world. The Google Library Project is digitizing books in collaboration with the libraries of Harvard, Stanford, Oxford, the University of Michigan, and the New York Public Library. It is competing with Project Gutenberg, the Million Book Project, specialized sites like Perseus and the Alexander Street Press, as well as Microsoft and Amazon. Some of these enterprises are philanthropic. Others are purely commercial.

Preservation programmes attract controversy. Some say that digitization projects are no replacement for preservation programmes based on reformatting on microfilm or on deacidification, conservation treatment and improved storage conditions. Others point to circumstances, for example in developing countries, where an institution with no experience or facilities

wants to preserve a specific collection. It may then decide to invest in digital instead of microfilming equipment to avoid high expenditure on cameras and processors and because the digital equipment and staff skills acquired will serve other purposes as well. In all the above scenarios:

‘Digital technologies offer a new preservation paradigm. They offer the opportunity of preserving the original by providing access to the digital surrogate; of separating the informational content from the degradation of the physical medium. In addition, digital technologies liberate preservation management from the constraints of poor storage environments typical of the tropical and sub-tropical climates in which many developing countries are located.’<sup>1</sup>

### Digital books

Until recently, a digital book was a reproduction of digital images of texts and photos on a physical surface such as traditional paper, photographic paper, film, cloth, plastic, etc. Two early digital book-printing systems were the DocuTech system from Xerox (introduced in 1990), and the InfoPrint system from InfoPrint Solutions Company.

### Box 2

High Definition DVD shares the 12cm diameter and 1.2mm thickness of the current generation of DVD discs, yet it can deliver eight hours of high definition video on a dual-layer, single-sided disc. Enhanced interactivity, multimedia functions, secure content protection and the capacity to store 10,000 average MP3 tracks on one disc are highly attractive to the consumer market. For the IT industry, HD DVD-R (the writable variant of HD DVD) can hold up to 30GBytes of data.

Today, the portable electronic book is on the horizon. In 1991 Franklin Electronic Publishers marketed an electronic Bible on a handheld device with a four-line screen and a keyboard. Sony went on to develop the Data Discman, containing encyclopedias and a World Travel Translator, swiftly followed by the Bookman.

The arrival of ‘personal digital assistants’ (PDAs) manufactured by Palm Computing in 1996 initiated the widespread availability of small but powerful computers. From there it was only one step to developing ‘electronic books’. They mimicked traditional books by being portable and were designed to display black text against a light background on a high resolution screen whose ‘pages’ could be ‘turned’ by pressing buttons.

The Rocket eBook was launched by Nuevomeia (USA) in 1998, quickly followed by the SoftBook, which had a leather cover and held 250 complete books. Other companies quickly joined in, the most notable being goReader, offering downloadable textbooks to university students.

In 2005 Start-up E Ink and LG Philips LCD announced an ‘electronic paper’ display screen measuring some 28 cm across and less than 300 microns thick. It uses microcapsules of oppositely charged black and white pigments floating in a clear fluid. The pigments rise or fall depending on an electrical charge. Thousand of microcapsules sandwiched between a piece of steel foil and a piece of clear plastic become the dots that form a black and white page. Unlike liquid crystal displays, the screen does not need to be backlit for an image to be visible.

More recently the Sony®Reader has hit the market, offering ‘electronic paper technology’ that provides ‘clarity and resolution that rival paper itself’. The Sony device is the size of a paperback and holds an average of 80 books. Its capacity can be extended with a memory stick or SD card.

For people who forget their reading glasses, the text can be enlarged by up to 200%.

In late 2007 Amazon went public with the Kindle device, which has a six-inch screen with no backlight and displays ink particles using electronic paper technology designed to overcome the glare and eye strain that users dislike about conventional electronic screens. Amazon has forged agreements with all the major publishers to make a large repository of titles available for rapid download. From Amazon’s Kindle Store, users can shop among 90,000 titles, which will download in less than a minute using Whispernet, the Kindle’s built-in wireless delivery system.

Amazon has partnered with a high-speed data network to make Whispernet available to mobile users without having to find a Wi-Fi hot spot. Whispernet delivers magazines, newspapers and blogs to subscribers automatically. Another avenue that Amazon is exploring for the Kindle is the digital text platform, where users can make their own content available for download on Amazon’s site. It is a form of self-publishing, in which a writer can upload content to Amazon’s site and then name a price for purchase by Kindle users.

Hybrid devices will be developed that combine books with personal organizers and communications. Similar to dedicated readers, they will have larger screens intended for long streams of information, buttons placed for easy turning of pages, and the usual capacities for bookmarking and annotating. However, they will also contain data organizers and do all the other tasks previously associated with PDAs such as e-mail, Internet browsing and MP3 playing.

### Digital newspapers

A growing number of newspapers have launched digital editions – replicas of printed newspapers that people can read on-screen or in newsprint in remote loca-

tions. In addition, most newspapers offer web sites to which access is by subscription or free. The two are perceived and managed differently. Web sites are more (inter)active, less static, and have the ability to combine text, graphics, sounds, and moving images. Critics of digital editions of newspapers say that they are only glorified screenshots of web sites, awkward to navigate and offering yesterday’s news.

Yet, digital editions have an increasing following worldwide. They offer readers familiarity, portability, and convenience. They allow readers to assess how a story ‘played’ in the newspaper (information that is often lost in cyberspace). Many digital editions are sent to a subscriber’s computer automatically. What is lost is the immediacy of ‘online’, replaced by news that is, at best, hours old.

In a recent article, Chilean journalist Lidia Baltra writes that these technological and cultural developments have profoundly altered editorial processes:

‘The characteristics of the digital newspaper are forcing the “paper newspaper” to make changes, both externally and internally. For example, as print information takes time to reach the reader through a long process of writing, designing, printing and distribution, it has to offer information that is more than plain facts, which has greater contextualization, analysis and scope. The editorial department today is one entity and open so that journalists, editors, photographers and designers from both newspapers (print and digital) can work in it in a more coordinated way. Everyone has to know how to manage multimedia and the department remains open 24 hours a day, seven days a week. Even “last minute news” has to be constantly and instantly updated.’<sup>2</sup>

The true promise of digital editions of newspapers lies in the future and in convergence with some of the technologies

described earlier. Digital editions and web sites will combine to offer breaking news, today’s commentary and tomorrow’s in-depth analysis in a more personalized format. Individual preferences will be catered to, with more on favourite subjects by favourite writers. Interactivity will be a determining factor. Readers will be able to communicate with journalists and editors to comment on the news, but advertisers will be able to target individuals according to their personal preferences. It will all be market-driven.

### Digitizing people

People can also be digitized. The first attempt in this direction, completed in 2003, was the Human Genome Project, a world-wide research effort aimed at analysing the structure of human DNA, identifying its approximately 20-25,000 genes, and determining the sequences of its three billion chemical base pairs. The information generated by the project is expected to be the source book for biomedical science in the 21st century. It will help scientists to understand and eventually to treat many of the more than 4,000 genetic diseases that afflict humankind. Digitizing DNA is the first step in replicating human life.

In the USA, the Visible Human Project (VHP) created a complete, anatomically detailed, three-dimensional representation of both the male and female bodies. The first ‘visible human’ was Joseph Paul Jernigan, a 39-year-old Texan convicted of murder and executed by lethal injection. His body was frozen to minus 160 F and ‘imaged’ with the same magnetic resonance and computer technologies used in medical diagnosis. He was then sliced into 1,878 millimetre-thin sections which were photographed and digitized. By late 1994 Jernigan had been reincarnated as a 15-gigabyte database. One year later, the body of a 59-year-old woman from Maryland who died of a heart attack was given the