12

Spinning the disks – lessons from the circus

John Robinson

Introduction

There are several assumptions which will inform this discussion. The primary one is that it is feasible and useful to attempt to describe the infrastructure required by the digital library in terms which can be understood by a non-technical reader without in any way masking the complexity of the requirement. The second assumption is that the description is enhanced by use of comparison with the pre-digital library. A third assumption is that working from the 'outside-in' is helpful for a non-technical reader. The final assumption depends upon the patience of the reader: the reality of the digital library cannot be separated from the complexity of the infrastructure required to deliver it and a clear understanding of the one requires at least a grasp of the scope of the other. Financial and economic considerations are inextricably bound up with technological complexity.

A great deal of normalisation of electronic information systems has occurred, especially since broadband Internet access became ubiquitous along with the widespread availability of low-cost, powerful personal computers. As with any pervasive technology, normalisation tends to mask underlying complexity and so long as the technology works, we are content to rely upon it despite having no idea what goes on 'under the hood'. In the case of the digital library, I will argue that a naive approach is not always the best one, especially when resource allocation choices must be made. Choices have consequences and a significant consequence for all of us is dependence upon complex technologies for access to the information which underlies our approach to knowledge. This, I will argue, has consequences which we cannot ignore.

Spinning plates

The spectacle of a master practitioner of plate-spinning is a wonder to behold. A dozen or more thin bamboo poles hold dinner plates aloft, five or six feet above the stage. The plates only stay up so long as they continue to spin. As they slow, they start to wobble. The practitioner, having started each plate spinning, must stay vigilant, running between the poles, catching the wobbling plates and bringing them back to speed. While the spinner is attending to one plate, another will start to wobble. The audience holds its breath. If the spinner is not fast enough, there is a crash and the audience lets out a sigh. The spinner retrieves another plate from the reserve stock, starts it up and the show goes on.

In ICT there is a notion of data which is held on 'spinning disks'. This is shorthand for saying that data is made accessible in real time from disk arrays attached to servers which are online to the Internet providing a web interface to which an authorised user gains access from a web browser on her desktop computer. Spinning disks are not the only means of storing data but are currently the most cost-effective method of holding data for real-time access. Other methods such as CD, DVD or tape archive require human intervention.

Looking into the various methods of data delivery (and the term 'data' here is itself shorthand for a wide variety of material which may present to the user as text, numbers, images, sounds or any combination in a multimedia format) it is possible to build a comparison with the 'old-fashioned' library or archive: books sit on shelves; special books sit on shelves behind library counters; archives sit in special environments – and in all cases the common thread is that the reader of the book or user of the archive must attend in person and be authorised before gaining physical access to the material at which point real-time delivery commences. By comparison, the web access which we have come to take for granted in the decade or so since it was 'invented' (around 1994) allows us to look at a screen and press a button. The data is sitting on spinning disks in a data centre somewhere out there 'on the net' and – given the right level of authorisation – we access it without having to leave our desks.

What we may not have grasped is the masterly performance required to keep those disks spinning. We may have stopped holding our breath at the spectacle, but we still notice when one comes crashing to the ground. This chapter seeks to map the scale of infrastructure which is required to bring the data to our desktops and the economic models which underpin what is by any account the extraordinary levels of

investment required to keep that infrastructure going. It will also examine the real and present danger of a crash and some of the strategies evolving to mitigate risk. It will suggest that there is an epistemological challenge presented by our increasing reliance upon real-time online access to information which can be put – quite literally – beyond grasp at the flick of a switch.

Before the digital library

Within living memory it was possible to obtain a university degree without resorting to electronic information systems except at the margins. A student – let's call her P – enrolled on a course. She saw her details entered onto a mainframe computer via a 'green screen' (text mode video terminal), then proceeded to purchase textbooks, writing pads, notebooks and other course materials. Lectures came with handouts; tutorials were supported by collections of journal articles helpfully copied and bound within the terms of the university's copyright agreement. There were more green screens in the library for the 'OPAC' (online public access catalogue) but once the details of a book or journal were noted down on scraps of paper, it was a matter of walking through the book stacks in search of the material. Photocopiers were available but often it was just as easy to write out the quotes from the books and articles in long-hand, as they were going to have to be entered manually into the essays and assignments in any case. A large part of the discipline of study revolved around transcribing accurate bibliographic details in order to make citation easy when writing up.

P had good handwriting and submitted many of her essays and assignments in handwritten form. So long as the papers were legible, the tutors did not object and it certainly made life easy when it came to sitting exams as she was able to write quickly and neatly without fatigue. When the pressure started to mount for typewritten essays, her typewriting skills were so poor that she found it quicker to write out by long-hand and then pay to have the material typed up. The rhythms of reading and writing were central to P's time at university. The library was a fantastic place to study; book stacks took up almost every available inch of floor space, with the study desks crammed in to every spare nook and cranny. Books were arranged across eight floors, a vertical rabbit warren, and the smell of books was everywhere. Each morning she would enter the library and head directly for a favourite spot, always a little anxious until she found it as there was competition for the quietest spots in the corners of the most remote stacks. She particularly liked the tables near theology but if those were taken, it was possible to find something quiet near philosophy or – at a pinch – archaeology. For some reason, literary and cultural theory (her double majors) were always busy and she preferred to find the books and take them back to the drier reaches of the library rather than sit amid the (albeit hushed) chattering of her fellow students.

An entire discipline of librarianship evolved around the business of building and maintaining libraries and it is not necessary to go far into that field for the purposes of this argument, other than to make some general observations. By and large the economics of the classic library was the economics of books and journals, bricks and mortar, and the deployment of skilled human labour. The library was a fierce physical presence designed to secure and control access to its contents. It was a repository of human knowledge, a sanctuary of scholarship and learning, the heart and soul of the academy. Libraries were the repositories for the crown jewels of academia and librarians the custodians and gatekeepers, the keepers of the keys to the kingdom. Lest this starts to read like a nostalgic lament for a lost golden age of the library, consider that librarians could be described as the inheritors after the Enlightenment of the mantle previously held by priests and scribes who controlled access to knowledge via the written word at a time when reading and writing was the preserve of the privileged few. The profession of librarianship is now having to adapt to a new age and some might say that the barbarians are at the gates, seeking to usurp or overwhelm the role of the gatekeepers.

The advent of the digital library

Let us consider P's son – we'll call him D – who enrols on a university course in 2009. D makes his selection of preferred universities online by browsing their websites and chattering about the options with his friends on social networking sites. He has his own PC, connected via the broadband network at home (which he finally managed to persuade his mother to pay for by getting her hooked on e-mail and online shopping at the PCs in the local library). She now has a PC at home and is very happy to use it for one thing at a time (mostly for writing books); she cannot fathom how he manages to do his homework, chat on line with instant messenger, do e-mail, browse the web and listen to music, all at the same time. She thinks he has the attention span of a gnat but has grudgingly to admit that he got good enough A levels in spite of that to be able to take his pick of some of the better universities (although she can't understand why he would want to study something called 'multimedia production' rather than a 'real' subject like English or computing).

Before the day dawns when D must pack his bags and head for his new hall of residence, he has already enrolled online, paid his fees, obtained a university log-in and e-mail address and chosen his first semester classes. Top of his list of reasons for choosing his university (apart from being a respectable distance from home) is its promise that he can bring his wireless enabled PC with him and use it anywhere on campus in addition to the hall (where every bedroom has an ethernet socket) and a selection of cafés in the local town where the university has teamed up with the council to provide a range of 'wifi hot spots'. D has been reassured that the university's firewall policy allows for use of his favourite social sites and has already seen the long list of electronic information services to which his new log-in provides access. There is some information about the university library which D doesn't pay much attention to, being more interested in the specifications of the machines in the multimedia section of the Integrated Learning Resource Centre (ILRC) to which (he gathers) the library is attached. The second-year student showing him around shares his enthusiasm for the quad-processor Macs and says, with a vague wave of her hand, 'the books are over there'. D – examining the phenomenal archive of digital multimedia projects and websites which she has just shown him, fully indexed in the university's 'digital repository' – can't remember when he last read a book, so doesn't pay a lot of attention.

Once he starts his course, D is given a reading list which is mainly URLs. He can browse them from anywhere, but soon discovers that the multimedia computer rooms in the ILRC have the best network connections and are the only machines powerful enough to play back the more complicated resources. There is also a very nice café attached, and his fellow students all seem to hang out there. Sometimes the computer rooms are a bit noisy, but he soon notices that most of the students bring in headphones to connect to the Macs and in any case everyone is sending instant messages back and forth and doing e-mail so a lot of the interaction is on the screen (even with the student at the machine next to him).

Access infrastructure

It is at this point that we need to pull back from the vision of D at his Mac, cocooned by headphones, furiously interacting via the screen with those around him while viewing ten different web screens at once, cutting and pasting material into a word processing file for his first assignment. We pull back and think about the infrastructure to which D is connected, upon which all of his activity depends. We'll call it the access infrastructure to distinguish it from the resource infrastructure discussed later. His Mac is one of a thousand machines the university bought at its most recent procurement. The university expects to get five years' life out of the machine. It has about 5,000 student terminals installed across its campus, which means that it expects to replace around a thousand every summer. This visible tip of a veritable infrastructure iceberg accounts (when installation and software costs are added) for about £1 million of the university's capital expenditure, every year, year upon year, indefinitely. The Mac at which D is sitting is one of 450 student workstations in the ILRC (counting staff terminals, this makes 500 seats). Tracing back from the Mac, it has an ethernet cable connected to a wall socket which connects to a wiring closet where the distribution network equipment is installed. When the ILRC was refurbished, it was 'flood-wired' using 'Category 6' ethernet. This means that there is an average of three outlets for every one actually in use (to enable furniture to be shifted and rooms rearranged without rewiring), along with adjacent mains electrical outlets. With space in the building for around 1,000 networked devices (including the staff PCs, printers, IP telephones and other networked devices including wireless network base stations), this means about 3,000 outlets connected to three wiring closets (one for each floor). This 'passive wiring plant' at refurbishment is expected to have a useful life of around ten years. Each wiring closet has active network equipment and fibre optic links back to a central server room. In order to make the sockets live, active network equipment is required. The central server room has the primary network connection to the outside world, a mix of small utility servers to keep the network going, medium-sized servers to provide operating system and application support for the PCs and Macs and a large server containing the 'digital repository' for the ILRC.

Table 12.1 summarises the capital investment by the university in the ILRC facilities along with the rate at which it is written off, the years in which reinvestment is required and the annual costs for electricity and maintenance. It shows a capital investment in the first year of £868,000 with running costs of £139,583. Major capital replacements fall due in years 5 and 10. If replacement capital was available year upon year at 20 per cent of the initial investment, a figure of £173,600 could be added to the recurrent, giving an annual facilities cost of £313,183 from year 2

Table 12.1 Investment by the university in ILRC facilities

Item	Unit price	Qty	Deprec	Yr 1 cost	Yr 5 cost	Yr 10 cost	Annual cost
PCs	£800.00	500	20%	£400,000.00	£400,000.00	£400,000.00	£82,120.00
Network edge devices	£1,200.00	30	20%	£36,000.00	£36,000.00	£36,000.00	£5,369.04
Network core devices	£7,500.00	2	20%	£15,000.00	£15,000.00	£15,000.00	£3,858.72
Small servers	£800.00	5	20%	£4,000.00	£4,000.00	£4,000.00	£1,775.92
Medium-sized servers	£5,000.00	2	20%	£10,000.00	£10,000.00	£10,000.00	£1,786.24
Large server	£25,000.00	1	20%	£25,000.00	£25,000.00	£25,000.00	£3,679.36
Data outlets	£65.00	3,000	10%	£195,000.00		£195,000.00	
Fibre-optic backbone links	£750.00	12	10%	£9,000.00		£9,000.00	
Floor boxes, trunking and electrics	£75,000.00	1	10%	£75,000.00		£75,000.00	
Room-cooling units	£3,500.00	10	10%	£35,000.00		£35,000.00	£13,608.80
Machine-room cooling units	£18,000.00	3	10%	£54,000.00		£54,000.00	£21,729.60
Leased-line to JANET	£10,000.00	1		£10,000.00			£17,000.00
Totals				**£868,000.00**	**£490,000.00**	**£858,000.00**	**£150,927.68**
Average capital spend per year				**£173,600.00**			**£324,527.68**

onwards. The first sustainability issue which has to be addressed is the fact that capital investment in ICT facilities must be refreshed at least every five years. A significant risk to the sustainability of ICT facilities, therefore, is a tendency in some institutions to treat capital investment as a 'one-off' and for capital spend to be managed through 'bids' which are treated competitively. One consequence of failing to plan a rolling capital replacement programme is the increasing unreliability of equipment and a major 'one-off' bill when the position becomes unbearable. Another consequence, given the differing lead-times for replacement of different bits of ICT infrastructure, is that 'emergency' spend (e.g. the need to spend a large sum of money towards the end of the financial year to avoid going underspent on the revenue budget) can lead to suboptimal procurement or even the procurement of the wrong equipment when an order is written for the kit which the supplier just happens to have in its warehouse ready for next-day delivery. The other significant risk to sustainability is running costs. The annual cost to sustain the facilities in our model ILRC appears at first glance high and may not tally with the experiences of readers used to managing an ILRC budget. In order to obtain these figures, I have included the cost of the electricity to operate – and cool – the devices in use in the ILRC. (I haven't taken any account of costs to heat and light the building.) This is a significant cost which is often accounted for separately, e.g. in a campus electricity bill which is not broken down by building or facility. Table 12.2 shows the electricity costs separated out from the other running costs of the ICT facilities.

Table 12.2 Electrical costs of running ICT facilities

Item	Annual electricity cost	Qty	Total
PCs 12 hr per day 312 days per year	£84.24	500	£42,120.00
Network edge devices on all the time	£58.97	30	£1,769.04
Network core devices on all the time	£1,179.36	2	£2,358.72
Small servers on all the time	£275.18	5	£1,375.92
Medium-sized servers on all the time	£393.12	2	£786.24
Large server on all the time	£1,179.36	1	£1,179.36
Room-cooling units 12 hr per day 312 days per year	£1,010.88	10	£10,108.80
Machine-room cooling units on all the time	£5,443.20	3	£16,329.60
Total			**£76,027.68**

These figures are based on some rough averages gained by studying the specifications of various pieces of equipment and should not be taken as authoritative. (The only reliable way to calculate the actual power consumption of any particular piece of equipment is by direct measurement when it is operating under load.) Although equipment has become relatively more energy-efficient in the past few years, the cost of electricity has risen dramatically. For the purposes of these figures I have used a rate of 9p per unit (kilowatt hour) which is the average domestic rate in 2008 after an initial allowance (about 16p per unit) is used up.[1] Large organisations may obtain a better price than this at wholesale prices, but the fact remains that the price was around 3.5p per unit in the early 2000s and every indication is that the price will continue to rise. Another way of interpreting these figures is to say that about a third of the electricity is to operate the PCs, about a third to keep the network and servers going and about a third to pump out of the building the heat generated by the equipment. This last point is worth pondering. In a typical building there are gas- or oil-fired boilers in the basement which heat water that is circulated around the building to heat it whenever the ambient temperatures fall below around 20°. In summer, especially with the presence of a lot of PCs, the public areas may require some cooling (accounting for most of the room-cooling units in the table). The heat will, quite literally, be pumped into the air above or around the building via heat exchange units. Except for the ambient heat from the PCs, which may reduce the heating bill during winter (assuming they don't overheat the building), the bulk of this heat is wasted. Money is spent heating the building and more is spent to cool the building, often at the same time, without any serious attempt to offset one against the other (e.g. by using waste heat from the machine room to heat the water which circulates around the building). This is a matter which has only recently started to be addressed under the heading of 'green computing': see, for example, the JISC project on managing environmentally sustainable ICT,[2] the Wikipedia entry on 'green computing'[3] and *Computing's* 'Green computing campaign'.[4]

A further reason for focusing closely on electricity consumption is to point out that, unlike the traditional library, the digital library ceases to function without electricity. At one level this is an operational point: electricity supplies to the library need to be secure, reliable and backed up by a contingency plan in cases of failure. But there is a more fundamental point which leads us to a further issue: the 'material' which makes up the *content* of the digital library is not only inaccessible in the absence of electricity, but in fact it *does not exist*, insofar as that 'material' itself may not be physically present in the library at all. The digital library, in other words, for all that it is inordinately resource-intensive, *has* and *holds* nothing – no stocks and no stacks – at all.

Resource infrastructure

The librarian of fifty years ago dealt with things which he or she could see and touch: books, journals, manuscripts, theses, research materials, archives. Stocks, stacks and stuff. Libraries were full of 'stuff' and librarians knew how to catalogue and store it so that others could find it, and how to get stuff from other libraries when local copies were not available. Librarians had budgets to purchase new stuff and 'weeding policies' for thinning out old stuff when space had to be found for new stuff. No disrespect at all is intended towards librarians, by the way; nor is this an essay on librarianship and the academic and professional expertise of librarians, whose contribution to scholarship and the pedagogic life of the institution is beyond dispute. On the other hand, it might be worth noting here that the 'changing university' in the mid-1990s, at least for the academics concerned, was widely identified as one which would focus on delivery rather than on discipline. For the moment, though, let's just look at the stuff which has been the stock in trade of the library. An example of 'bread and butter' library materials which can stand in for a multitude of cases is academic journals; an example of the challenge posed to the maintenance of archives comes from amateur photography.

Journal collections stand out as one of the earliest and most successful examples of mass migration from paper-based to electronic form. Journals historically were expensive to procure, expensive to maintain, labour intensive to archive and they took up a great deal of shelf space. I propose to refer to journals here not in order to consider the phenomenon of online publishing but as an exemplar of the resource infrastructure challenge, and this brief consideration of academic journals will allow me to review a range of relatively recent attempts to *hold digitally onto*, more generally, data that once held its material place in one or another archive. Plainly, accurate and reliable citation of journal sources remains a prerequisite for scholarly writing, and it is a fundamental assumption that the reader may, if he or she chooses, go to the original source to check the citation, perhaps to undertake further research. A measure by which a library has been judged is the ease with which a journal can be located, a volume found, an article retrieved. A large library in a large institution covering many disciplines is expected to carry full runs of journals covering hundreds of titles going back tens or even hundreds of years. For a copyright library, the physical burden assumes immense proportions. Lost journals or gaps in a collection can cause havoc for scholars and trigger desperate attempts to locate copies in other libraries and to negotiate copying rights, especially where the publisher has vanished or the journal is long since out of print.

Faced with this challenge, the attraction of online journals becomes, at first glance, obvious, but at second glance, a question rapidly emerges: 'online *where*, precisely?' If the answer is, 'online at the publisher', two questions follow. Firstly, 'what if the publisher goes under?'; secondly, 'what if the publisher ceases publication of the journal and takes the back issues off line?' In the days of paper-based journals the answer in each case was always the same: 'we've got copies of everything we paid for, sitting on our shelves'. In the online world, the simple answer might appear to be, 'we downloaded everything and put copies into our local repository'. Putting to one side the licensing and copyright issues which this response raises, let's focus for a moment on the sustainability issues which arise from the notion of 'keeping a local copy of an online journal'.

Keeping 'safe' copies of important material

You may recall that the model budget for an ILRC, set out above, included an entry for a 'large server', costing about £25K to purchase and about £3.6K p.a. to run. You may also have noted that the expected useful life of the server is five years. My outline of ICT costs has not yet included the labour costs involved in setting up and running the service, so I propose for the moment to leave these to one side in order to concentrate on the infrastructure costs and specifically the implications of 'replace after five years'. What this means in simple terms is that the option of 'keeping a local copy' is not actually the equivalent of keeping paper copies of subscribed journals. The hardware upon which the local copy is kept has only a limited life, not just because it gets old and is relatively uneconomic to maintain, but – more importantly – because the system itself becomes obsolete and the material held on it has to be 'migrated' to a new system every five years or so. The library's 'holdings', in other words, are inherently unstable, 'un-held', or 'de-held'. Clearly, there can be backups, but for all the technological might of the computing industry, no mechanism for the archival storage of text which is more durable than printed paper has actually been invented.

Anyone who has used a computer since the late 1980s is already familiar with the number of storage media which have become obsolete: 5.25" floppy disks, 3.5" floppy disks, zip disks and DAT cartridges have all become redundant with rapid changes in the marketplace. Interesting old files and other material stored on them often can no longer be accessed because there is no functioning device with which to read them. CDs and DVDs have emerged as data storage media, but these disks, especially the rewriteable varieties, degrade. With increasing data volumes, it has recently become much easier to buy external hard disks and store all the material there (so long as one remembers to keep making copies onto extra disks in case one of them fails). Now that recordable DVDs have become the norm for off-air TV recording, most of us have a collection of old VHS tapes gathering dust. We may not be aware until it is too late that these tapes have only a limited shelf life and – even if we still have a functioning VHS player – they simply won't play. Perhaps the starkest 'popular' example familiar to most of us of the change from permanent to volatile storage is digital photography. Images taken with digital cameras reside initially upon memory cards and are then transferred to computer disks. How long are these images going to last? An entire genre of cultural history revolves around the secrets of the family photograph album and there are numerous examples of albums, dating from 50 or 100 or more years ago, backed up in some cases by the original negatives. What is the prospect for the equivalent family snaps taken today on a digital camera? This issue is central to the scenario which led the EU's Lund Report (European Commission, 2001) to speak of the risk of a 'digital dark ages', thereby engaging explicitly with this issue of sustainability. Of course, the overwhelming conclusion of such reports has been that there must be collective action. I return to that notion below. Before I do so, it is worth considering a little longer some of the lessons provided by popular culture. This example provided by the transition from celluloid and chemical based to digital photography might have something to teach us when our pressing concern is with the holding of scholarly materials in university libraries and learning centres.

Functionality, complexity, sustainability

Mass production of cameras and film stock and the widespread availability of processing facilities at low cost enabled the family

snapshot to evolve to the point of being a widespread artefact of daily life during most of the twentieth century. The film camera itself is a relatively simple piece of technology involving a light-proof container for the sensitised material, a lens to focus an image and a shutter to control the exposure of the film. The photographic process provides a permanent record of the scene captured by the camera which is a primary artefact: a negative or transparency embedded in the same piece of celluloid which was present at the moment the image was captured. The current digital camera, by comparison, is a veritable marvel of technology: a computer with a lens attached. The image captured by the camera lens is focused onto a sensor (an analog device) and is present upon its surface for a fraction of a second, after which time an 'analog-to-digital converter' generates a computer file which describes in a complex digital format (possibly proprietary to the manufacturer of the sensor) the image which was momentarily projected onto the sensor. Beyond that moment the image has no physical presence and is not therefore intrinsically stable. One of the particular attractions of this technology, especially in photojournalism, is that the image can be transferred immediately to computer, manipulated, printed, e-mailed and published on the Web. Putting to one side the dependence upon high-tech industrial processes for the manufacture of the camera and peripheral devices, it offers the user a strikingly increased degree of control over the process of taking and handling images. Increased functionality comes at a cost, however, and one of the costs is that the process does not produce an artefact. There is no negative or transparency which – after processing – can be safely stored for tens or hundreds of years; there is only a computer file. Secondary processes are required to secure the image file and these can be complex and expensive. In the domestic context, the process of storing image files in such a way as to make them available to future generations is almost infinitely more complex than the keeping of a family photo album in the sideboard and most likely beyond the capability of even the most dedicated amateur photographer.

While this example may not be adequate to substantiate a general rule, if an observation is permitted, it might be that with increased functionality comes increased complexity and that increased complexity is the enemy of sustainability, not only because the process of image capture is inherently more complex, but also because the process of securing that image as a permanent record is complex and requires constant updating. One obvious solution in the case of images which are judged to be important, and worth preserving, is to create archival prints. This has a dual irony: the method involves a reversion to the 'old-fashioned' methods of the archivist; it runs counter to the prevailing wisdom which is to scan old images and films into digital format for access and 'preservation'. Is it possible that despite immense technological development, the digital scans of, say, nineteenth-century photographs are actually less stable than the originals (which have, after all, already survived for more than a hundred years)? Extrapolating from this example, the two primary challenges to the sustainability of digital archives, including those 'containing' academic journals, are firstly the relatively short life of digital file formats and the hardware upon which they are stored (leading to the need for continual capital replacement and data migration programmes), and secondly the highly specialised labour required to configure, support and manage these systems.

Sustaining the infrastructure

Some might argue that a book is an example of information technology and therefore librarians have long been information technologists. It is certainly the case that librarians have always been engaged with the problem of sustaining their infrastructures. What has changed dramatically with the advent of the digital library is the extent to which the infrastructure has changed, imposing new costs and requiring new skills, while limiting local mastery of the material and processes concerned. Libraries which seek to be stand-alone digital libraries must take on board new facilities and the skilled staff required to build and operate them, if not actually to control them. Much more commonly (in the education sector at least), libraries are required to engage with and depend upon the institutional information technology infrastructure provided by computing service or information systems departments. In some cases (around 32 per cent of the HE sector and 21 per cent of the FE sector according to JISC research[5]), libraries have combined with IT and/or IS and/or AV services to form 'converged service' departments with a single reporting line into an institutional senior officer possibly described as a chief information officer (CIO).

In each case, the generic challenges are the same: capital-intensive infrastructure which must be replaced every five or ten years; teams of people with the high-level systems administration and programming skills required to operate the infrastructure; new approaches to revenue budgeting which take into account the new operating costs and the requirement by the institution to defend these costs against its core objectives in learning, teaching and research. What is most striking about this change is that, as the institution commits more and more of its intellectual property (either bought-in as books and/or created locally as learning objects, theses, research outputs and the like) to digital storage, the requirement to sustain the infrastructure extends out to an unlimited future.

Collective solutions for collective problems

As noted above, one response to this scenario is to seek collective solutions. Take the case of journals: when they were paper based it could be assumed that a number of libraries held copies and hence, should a library lose a copy and the publisher not be able to replace it, a copy could be sourced from another library. With online journals, the library may hold a local copy in its 'digital repository' in case the publisher goes out of business or ceases to deliver the title, but what happens if that local copy is lost or corrupted? Several approaches have been devised including the well-known JSTOR in the US,[6] and another which goes by the picturesque title of 'Lots of Copies Keep Stuff Safe' (LOCKSS).[7] In this model, a network of repositories is formed which synchronise with each other to ensure that local copies are clean, uncorrupted and up to date. Plainly, this 'solution' requires funding, and models are currently under investigation to secure funding on a long-term basis.[8] Both these approaches depend upon a fast and reliable wide area network. In the UK this network is called JANET[9] which has been centrally funded since the early 1990s (when it was used primarily by computer scientists) by grants 'top-sliced' from all the UK education and research funding councils (supplemented in some cases by compulsory institutional subscriptions). JANET is a secure, high-speed private network connecting all UK educational institutions to a common 'backbone' which has dedicated interconnections ('peering points') with many other education and research networks across the world[10] as well as commercial 'Internet service providers' (ISPs) including the BBC. During the 2008/09 academic year, the budget for the operation of JANET was £49.6 million, of which £39 million was passed directly from the funding councils with oversight by JISC and a further £5.7 million collected from higher education institutions (HEIs) within the terms of the JISC agreement. In return for this funding, JISC has a service level agreement (SLA) with the network operator which requires that every institution of higher or further education and every research council in the UK is provided with at least one dedicated connection to the JANET backbone, the current version of which (SuperJANET 5) was constructed using a separate capital grant of £50 million over five years from 2005/06. The SLA also mandates the international connections, the peering with commercial ISPs and a variety of other network support activities.

In summary, the existence of digital libraries, especially when resources are held or maintained external to the institutions, depend for their connectivity in the UK upon public expenditure in the order of £60 million per year (recurrent plus capital) which has been explicitly determined by the principle that pooling of funds into a common pot enables the procurement of a network which far exceeds in quality and reliability anything which individual institutions could purchase from commercial providers. A corollary of this is that a network which had its origins in computer science and research has now become a part of the mission-critical infrastructure for almost every aspect of the institutions which it connects. Other collective approaches have also attracted central funding. During the period 2006/09, JISC will spend around £12 million on programmes to digitise a variety of collections of material which will be freely available to the sector and a similar amount to support institutions in the development of their 'digital repositories'. It also spends around £5 million p.a. supporting its two national data centres (EDINA and MIMAS) which provide digital content and resource discovery services and act as exemplars in the hosting and presentation of digital material.

JISC has prepared comprehensive advice on digital 'preservation' and 'curation' which suggests that '[d]igital preservation should ... be addressed from as early in the object life cycle as possible, particularly as the manner in which a resource is created has a significant impact on its durability. This requires effort and input from an organisational and cultural perspective as well as a technical one.'[11] Given the economics discussed above, large-scale digital archives become attractive on the grounds that economies of scale set in, teams of people can just as well manage large archives as small ones and overheads for electricity and cooling can be shared, but these depend upon ongoing funding. Recent experience with the Arts and Humanities Data Service (AHDS)[12] has demonstrated that such funding cannot always be relied upon. In that case, a well regarded, centrally funded service to provide secure archiving of complex data sets lost its funding because one of the funding partners decided that its priorities lay elsewhere. At the time of writing, the service has relaunched itself with a mixture of institutional support (it is distributed across six partner institutions) and project grants.

You will notice that no comprehensive solution is offered here: this is because no such solution exists and any solution which does emerge will most likely not be simple. I will conclude therefore with two observations about the ways in which the landscape is changed – perhaps forever – by the advent of the digital library.

Keeping the spinning disks aloft

I started with the image of the circus performer, dashing about the stage, keeping plates spinning at the end of bamboo poles. When done well, this is a masterly performance, rightfully deserving of respect. When considering the digital library, the metaphor has resonance in the first instance because of the notion that the data which the library contains is held on spinning computer disks in order to be delivered online, upon demand. What is less apparent to the user of the digital library is that he or she depends, for access to the materials, upon the masterly performance not of individual, expert human performers, but of machines and ICT professionals, operating out of sight – perhaps in another city or another country. Their performance must be maintained at all times that access is required which in effect means 24 hours per day, every day of the year. The cost of access, then, must be able to be measured and examined for value in relation to the new capabilities which are enabled by this delicate, masterly balancing act.

One of the potential consequences of the move from low-tech, sustainable information technologies such as printed books and journals to high-tech digital technologies such as those described in this collection is the pressure to centralise resources, to obtain the economies of scale which can bring the cost of the digital archive within grasp, and to depend, no longer, on individual human expertise. Rather than each centre trying to maintain a local digital repository, the argument goes, why not build giant repositories, to be accessed across the network: places like the British Library,[13] for example, or the National Archives?[14] At first sight, this can seem to be an attractive solution to a difficult problem. But perhaps we should stop and think about this a little more. Victor Keegan, writing in *The Guardian*, points out a historical parallel: '[w]hen the 1850 public libraries bill was going through Parliament, opposition came mainly from MPs representing the universities of Oxford and Cambridge. They were appalled at the idea of unmediated information getting to people who probably shouldn't have it and in whose hands it could even be dangerous' (Keegan, 2008). Are we about to take a large step back to a time before the widespread distribution of information to local sources – or have we perhaps already taken that step? Isn't 'ultra mediated' information energy-intensive to maintain and control and, as a consequence, attractive to those for whom control is more important than content? Alternative approaches using distributed models such as 'Controlled Lots of Copies Keep Stuff Safe' (CLOCKSS) reduce the reliance upon a single repository of *content*, but not upon the surrounding infrastructure such as network connectivity and access control which continues to be most efficiently delivered through central funding and operation. As the JISC advice puts it, this is increasingly an 'organisational and cultural' and – dare one suggest – political, ideological issue.

Conclusion

A great deal has been written about the extent to which the digital, online age has 'opened up' the information previously held within the four walls of institutions, but the same case cannot be made for mediation itself. Information may be highly accessible, but accessibility is mediated through complex technologies, characterised by strong centralising tendencies which could well bring us full circle back to the notion of a small number of extremely well guarded, centralised knowledge repositories – and to those few who control them. Accessibility is also inherently fragile. I initially described this fragility as having more in common with the performance of a master plate spinner than that of a librarian, but we can now see, I think, that in the realm of the digital, neither of these professionals would have the means to take that fragility in hand. Even where repositories remain open and accessible, perhaps under democratic control by friendly regimes, our access is nonetheless mediated by a technology which can be taken away at the flick of a switch. Perhaps, on this basis, the most pessimistic among us would argue that the digital dark age is already upon us, held at bay only so long as we can keep a light burning.

What is the status of knowledge at that moment when the power supply fails us – because it might well fail us – or if electricity becomes an expensive and scarce resource? The question is urgent for those of us who know exactly where the light switches are located. At this point in time, we no longer ask: 'Can we sustain the infrastructure?' Instead the urgent question is this: 'What if we can't sustain the infrastructure?' What then happens to the digital library and to widening access – and to the knowledge/power hook-up itself? If you are tempted, by the way, to feed these sorts of questions into Google, pause for a moment and consider that Google, apparently here at hand, is actually running in a vast server farm in California. That farm consumes so much electricity that it has been rumoured that Google will shortly need its own nuclear reactor to continue to run it. The example of Google is apparently banal, but the infrastructure and sustainability questions return with that banality: how does Google get to our desktops? On what infrastructure does it depend? And what will our digital library users do if Google is unplugged? What should we conclude if it is indeed the case, as some colleagues argue, that we should build a digital library which is 'as reliable as Google'?

The loss by fire of the great library of Alexandria continues to hold a curious resonance for some of us, representing a loss of more than books and paper. The electronic age has broken through the walls of the library, with much of the disruption that this metaphor implies. As it did so, it has changed the role of both library and librarian as repository and custodian of human knowledge; it has diluted the power of the librarian as gatekeeper, and it might be argued that, by the same token, a knowledge community, and all that goes with that notion, has been dispersed. The notions of catching and of holding (knowledge or plates) may well have gone with these, and in this sense, the digital library's Alexandria (not a conflagration this time, but the curious sound of universal silence) is a little bit closer. In terms of sustainability, let's be clear sighted about what we have, what it means, what it costs and how rapidly it might slip from our grasp.

Notes

1. Since writing this (early 2008) the price of electricity has risen by around 40 per cent, which means than an average price per unit is now around 16p and my supplier has dropped the discount rate. Although oil prices (to which electricity prices are linked) have fallen following a mid-year peak, electricity continues to increase in price. This means that the figures in Table 12.2 could be underestimated by as much as 50 per cent and can be expected to increase year upon year well above the rate of inflation. The cost of electricity is emerging as a significant drain upon institutional budgets across the board, not just for ICT equipment.
2. *http://www.jisc.ac.uk/whatwedo/programmes/programme_jos/susteit.aspx*

3. *http://en.wikipedia.org/wiki/Green_computing*
4. *http://www.computing.co.uk/computing/specials/2162404/green-computing*
5. See p. 15 of the 'institutional preparedness' report at: *http://www.jisc.ac.uk/media/documents/themes/accessmanagement/ipsreport.pdf*.
6. *http://www.jstor.org*
7. *http://www.lockss.org*
8. See, for example, CLOCKSS at: *http://www.clockss.org*.
9. *http://www.ja.net*
10. For example, Internet2 (*http://www.internet2.edu*) in the US, DANTE (*http://www.dante.net*) for mainland Europe, AARNET (*http://www.aarnet.edu.au*) for Australia, HEANET (*http://www.heanet.ie*) for Ireland and KAREN (*http://www.karen.net.nz*) for New Zealand.
11. *http://www.jisc.ac.uk/publications/publications/pub_digipreservationbp.aspx*
12. *http://www.ahds.ac.uk*
13. *http://www.bl.uk*
14. *http://www.nationalarchives.gov.uk/*

References

European Commission (2001) *The Lund Principles: Conclusions of Expert Meeting on European Content in Global Networks*. Lund, Sweden, 4 April. Online at: *http://cordis.europa.eu/ist/digicult/lund-principles.htm*.

Keegan, V. (2008) 'Creativity shouldn't be limited to finance', *The Guardian*, 4 March.

13

The CREE project: a case study on the novel delivery of search-related library services and its economic implications

Chris Awre and CREE staff at the University of Hull

Introduction

The Collaborative Resource Evaluation Environment (CREE) project[1] was awarded funding by the Joint Information Systems Committee (JISC) in December 2003 and ran from February 2004 to August 2005. A second round of funding from the JISC was awarded for work taking place between January 2007 and July 2008. The work has been led by the University of Hull, and has involved a range of partners in both UK higher and further education.[2] This case study describes the work of the CREE project and how it has sought to understand novel ways of presenting search-related library services, work that was undertaken in the context of the development of library services generally within the University of Hull. In considering this context, the case study highlights the impact of the work on the management and value of delivering digital library services.

Why CREE?

The CREE project was funded within the JISC's Portals Programme,[3] an initiative that looked to exploit the perceived benefits of portals in helping to manage the increasing range of electronic resources available to end-users and access to these resources. The Portals Programme