PAUL MOBBS

A practical guide to sustainable IT

Unit 6

APC

This unit is one of 12 sections to a "A practical guide to sustainable IT", a hands-on guide to working with everyday technology in an environmentally conscious way. The guide has been written by environmental activist and ICT expert Paul Mobbs, and was commissioned by the Association for Progressive Comunications (APC) with the support of the International Development Research Centre (IDRC). To download the full text of the guide, or any of the other units, please visit: greeningit.apc.org

Author
Paul Mobbs

Copy-editing
Alan Finlay

Layout proofreading
Lori Nordstrom

Publication production
Karen Banks and Flavia Fascendini

Graphic design
Monocromo
info@monocromo.com.uy
Phone: +598 2 400 1685

Commissioned by the Association for Progressive Communications (APC).

Conducted with support from the International Development Research Centre (IDRC).

A practical guide to sustainable IT

Published by the Association for Progressive Communications (APC) with support from the International Development Research Centre (IDRC).

South Africa
2012

APC-201206-SU-R-EN-DIGITAL-162
ISBN: 978-92-95096-71-4

INFORMATION STORAGE AND SECURITY

Managing information storage securely and reliably is one of the ways that we can reduce energy use in IT. Apart from the impacts of creating and storing information, a significant factor in the energy and carbon footprint is the human user; and if the user has to spend time searching for information, or recreating lost information, that represents a waste of time and resources. Learning to manage our use of storage devices is where we begin to tackle this problem. Then we can move on to looking at solutions that protect sensitive data from disclosure, important data from being lost on the hard drive, and all data being accidentally lost, wiped or corrupted.

This unit examines various methods to improve the security and reliability of computer systems and the information that they contain by considering the work of backing up.

The simplest rule to secure information on a computer system is to back up.[1] Backing up works on the principle of *redundancy*; if one copy is lost or damaged you can revert to the back-up copy. There are many different "pathways" for backing up. Which is the best suited/most convenient pathway to use will depend upon how your data is organised, what the objective of the process is, and whether you are trying to secure just a few files or an entire computer hard drive.

1, Wikipedia, Backup. en.wikipedia.org/wiki/Backup

6.1. DATA SOURCES AND REDUNDANCY

The purpose of backing up is to preserve important data; if we have to spend money and time backing up superfluous files then that represents an unnecessary waste of resources. Therefore the first step in developing a system to back-up data is to identify what *needs* to be backed up, and avoid including data that does not need backing up.

6.1.1. Identifying what to back up

If we characterise the types of files manipulated or stored by computers, and the need to back up that data, we can broadly identify six categories:

- ***Freely available downloaded data*** – *for example web pages, commonly available files and data.* While this data may be important, it is not irreplaceable. Unless it forms an important part of the information we keep, backing up this kind of information is not a priority because we can easily and cheaply – using less time and resources than keeping our own back-up copy – download another version if we need it.
- ***Paid-for downloaded data*** – *for example downloaded music, subscription publications and paid-for data/files.* As this is not irreplaceable data, the question is whether backing up would cost more than the value of the data you may potentially lose. Another difficulty is that music or other downloaded files may only work on the "authorised" computer they were downloaded onto; backing up digitally-locked files such as this only has value if you can transfer them to another machine later (e.g. the iTunes service allows you to authorise up to five machines to play your downloaded music).
- ***Stored data (offline)*** – *for example software installation discs, bought CDs/DVDs or your own data back-ups.* As with downloadable data, if you already have an offline copy there is no need to back up this data again unless there is a reason to do so (for example, if you have edited/changed the contents of a stored file since it was read from the storage medium).
- ***Stored data (online)*** – *for example websites, files kept in online systems (e.g. social media) or information stored in online services (e.g. web-based email and file storage).* One of the problems of backing up is the scale of the data that is now routinely stored on many different computer systems. With the growth of online services, this data is increasingly not under the control of the person who created the file or data. While backing up to an online system is a way to avoid backing up using physical media (DVD, USB sticks, etc.), the question is what would happen if the online system was unavailable. For example, if you run a website and the server is hacked, do you have an up-to-date back-up copy of the site to quickly restore the online service? Or if you routinely store data online and lose your internet service, could you access that data if it is urgently required? All data stored in online systems should, if it has value, be backed up locally too.
- ***Replaceable personal data*** – *for example emails, circulated reports/files and information shared between groups of users.* This type of data isn't irreplaceable because copies will be held on other computer systems. The issue is how much effort it would take to reconstitute this information if you were to lose it. In most cases keeping back-ups of this information is a simpler option than trying to recover it from many other sources/locations later.
- ***Irreplaceable personal data*** – *for example draft work, personal images or recordings, and files which are not held on any other storage medium or computer system.* This is the most important data to back up because there are no redundant back-up copies to replace this data in the event that the computer system is lost, damaged or fails.

What this list illustrates is the importance of valuing the data we propose to back up. If we're trying to minimise the ecological impacts of IT, being able to plan how we back up and minimise the use of resources as part of that process is an essential part of a green IT strategy. There is one key rule to take note of in this process: *The more common and easily accessible a resource the less we need to back it up; the more unique or expensive a resource the greater the need to create a back-up copy.*

Box 6.1.

Backing up mobile devices

Today there are a variety of mobile computing devices in use, from mobile phones to personal digital assistants and fully functional laptop PCs. Mobile devices need special attention when it comes to backing up their contents because they are more likely to be lost, stolen, or damaged.

Many smart phones have built-in back-up software to copy the files they contain to a PC. Connect the phone to a PC and the contents of the phone can be copied either to a single large file (full back-up) or a directory containing the files on the phone (selective back-up). It's important to use the internal back-up software of the device in order to capture the operational and configuration files the device requires, but which are not normally available for the user to manipulate. Android phones are technically already backed up to "the cloud" – the data storage system operated by Google. It's not technically possible to make a local back-up, although there are third-party applications which allow an Android phone to back up to other online services. There are also an increasing number of commercial services, usually run from pre-paid mobile applications, which will automatically back up the data on your phone. In the event of loss or damage to the phone, the data can be recovered from the service and downloaded to a new phone.

For ordinary mobile phones there are methods to back up the contents of the SIM card, but these tend to be expensive as they often require a physical device to read the card. Some mobile operators give the option of backing up the address book and other features of the phone to the operator's system – and these can be downloaded to a new phone in the event of loss or damage.

Backing up iPods and similar music-playing devices is more complex, in part because it runs into problems of digital rights/copyright. In most cases you are loading data from another machine so keeping a back-up copy is unnecessary – unless files are only stored on the device and nowhere else. Most music download services will register a number of playing devices, and so again in the event of loss or damage the old device can be unregistered and the new device registered in order to transfer files onto it.

Finally, laptop and notebook PCs can be updated just like any other desktop system. The machine can perform a back-up over a network, either to another PC or to a network server, or connected to an external storage device (such as a hard drive, USB stick, external hard drive or memory card) to copy the files that require backing up. iPads can back up to another computer via a cable or Wi-Fi connection, or you have the option to store data in Apple's cloud storage system.

What's important is that we organise files in a way that assists the process of making back-ups. This means organising information in a way that separates the important data – which requires regular backing up – from other less important data. Separating important data from the less used and superfluous files on the computer system is also good practice because it allows low value files to be regularly deleted from the system to free up space on the hard drive. Organising data by its importance is easily done using a clear structure of directories/folders on desktop (or server machine) to isolate the most important or regularly edited files from other data which has a low value or is infrequently changed. For example:

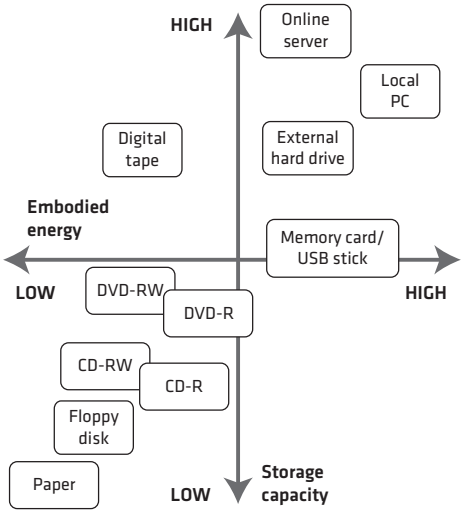
- Users should store their current work/important files inside a single identifiable "work" directory. This simplifies the process of deciding which files to back up – backing up all current files can be easily carried out by copying the "work" directory to another machine on a network or to a removable storage media;
- Other user files that are used infrequently, or which have already been archived, can be segregated into an "archive" directory;
- Low value or superfluous data, such as web pages or downloaded files, should be kept separately from the user's recent and archived data – and these folders can be regularly sorted and their contents deleted to free up hard drive space;
- Where digital rights/copyright is an issue, data which presents a legal problem if copied or backed up (for example, commercial music or video files) can also be segregated from the bulk of user data, and excluded from the back-up process to avoid any legal difficulties which result from making copies of that data.

6.1.2. Ecological impacts of backing up

How we decide to back up will have an impact on the ecological footprint of our IT needs. For example, if we use reusable back-up media, such as digital tape or external hard drives, that can over time have a lower impact than options where we use the storage media once. The difficulty is that there is little detailed information on the impacts of different back-up technologies, and there has been no comprehensive life-cycle analysis of backing-up options to compare one option directly to another. While there are some studies

Figure 6.1.

Relative storage capacity and embodied energy of back-up technologies



which show that downloading music is less ecologically damaging than buying music on a CD,[2] or that buying software online is better than getting it on DVD,[3] once you back up those files onto a CD/DVD or other offline storage media to keep them secure, most of these benefits are lost.

Much of the recent research on the impacts of computing to date has focussed on the use of servers and cloud computing rather than the impacts of everyday data storage technologies such as CDs, external hard drives or tape storage. Just as many green IT studies do not consider the embodied energy[4] of the equipment involved, studies on the electronic distribution of data do not consider the human resources involved in creating or purchasing data – and the relative value of backing up that data via different means to prevent its loss or corruption.

2. Koomey et. al. (August 2009). The energy and climate change impacts of different music delivery methods. download.intel.com/pressroom/pdf/CDsvsdownloadsrelease.pdf
3. Accenture/WSP (October 2009). Demonstrating the Benefits of Electronic Software Distribution: A study of greenhouse gas emissions reduction. www.digitalbychoice.com/en-gb/483648_CarbonFootprint.PDF
4. Wikipedia, Embodied energy. en.wikipedia.org/wiki/Embodied_energy

While we can't make decisions with certainty, what we can do is form general rules on the impacts of different storage options based upon the characteristics of the technologies involved:

- For long-term storage (years rather than months) passively held data has a lower impact than actively maintained data (e.g. storing data for long periods on optical discs like CDs or DVDs requires less energy to maintain than storing data live using online services).
- Where data is not updated on a regular basis, the embodied energy of complex or semiconductor-based technologies is higher than other storage options (e.g. keeping archived data on magnetic tapes or discs, or on optical CDs/DVDs, has a lower impact than using external hard drives, memory card/USB sticks or online storage).
- Where data is regularly updated, meaning that any static back-up would quickly become out of date, then online storage, external hard drives and memory card/USB storage are a better option.

Figure 6.1 illustrates the relative embodied energy of different storage technologies and their relative storage capacity. Certain technologies offer a higher storage capacity, but often this is associated with a higher ecological impact. Even similar technologies can differ – for example a local PC or server machine has a higher impact than an online service because in most cases online services will be optimised to operate more efficiently than a comparable small server or PC. There are also differences between compatible technologies – for instance, because a re-writeable CD-RW or DVD-RW can be reused many times, it has a comparatively lower impact than a single use CD-R/DVD-R.

6.2. DEVELOPING BACK-UP SYSTEMS AND PROCEDURES

Deciding how often to back up is a balance between practicality and the cost of data loss. Where back ups can be automated – for example copying a whole hard drive to a digital tape – the process is less demanding of time and so can be carried out more frequently. Where individual users back up their files to removable storage, such as a USB hard drive, it requires more time and so would be carried out less frequently. What's important is that some form of back up is put in place, and then procedures are agreed to ensure that these systems are used on a regular basis.

6.2.1. Backing-up pathways

Figure 6.2 illustrates various methods of backing up. There are three different roles in this process which are defined by who has control over the system hardware:

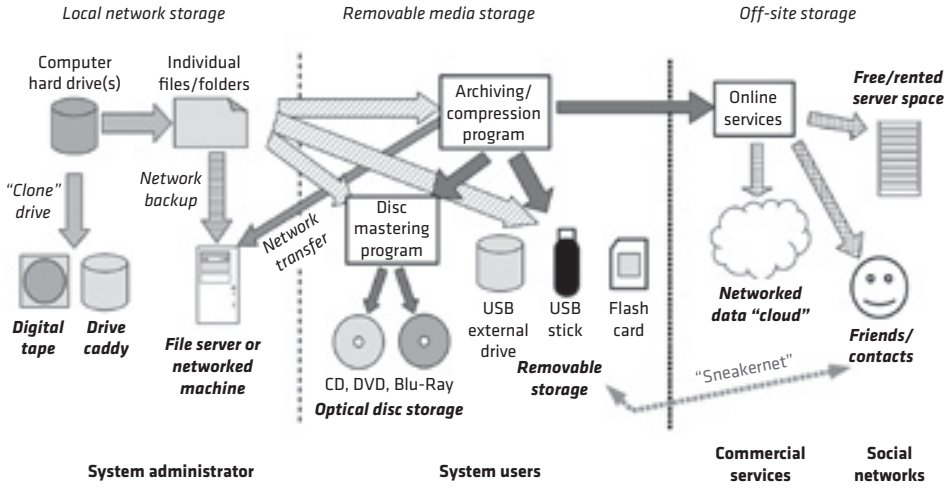
- *System administrators* are responsible for back-ups which require special actions or security privileges – for example, backing up a hard drive to digital tape. Where a single user looks after their own system they would carry out this role, but for larger installations where there are defined roles, these actions are the responsibility of the individual who has responsibility for the IT infrastructure.
- *System users* are responsible for the files which exist within their own storage space, whether that be on their own machine or on a networked server. While their machine can be backed up centrally, encouraging users to undertake their own back-ups of essential files adds an extra level of security to the process – and helps discriminate between essential data files and other superfluous data.
- For off-site storage the system user/administrator is involved, but it also relies upon other agencies in order to provide these services. These might be free services, contracted services, or informally organised storage offered by virtual communities or social networks. The main concern with all off-site storage is how secure the data is, and whether it can be accessed when required.

6.2.2. System-level back-ups

We'll work through figure 6.2 from left to right. The process begins with the hard drive inside

Figure 6.2.

Back-up pathways



the computer. Information on a computer hard drive is vulnerable to hardware failure, corruption of the operating system (for example by malware)[5] or user error – and of course theft of the machine or a disaster which befalls the building the machine is located in. There are various ways to protect the information stored on a machine, and each option gives a differing level of protection.

Firstly, it is a simple procedure to duplicate the contents of the hard drive, most straightforwardly by operating a second hard drive within the same machine. There are systems which can do this automatically, such as RAID[6] hardware which automatically duplicates data on two hard drives. This is effective at preventing loss due to hardware faults but will not prevent losses from malware, user error or physical damage to the hardware. Additionally, running hard drives in parallel increases energy consumption in the machine. Therefore, for most small servers and desktop computers, the simplest option is to copy an *image*[7] of the hard drive to a back-up storage system:

- Using a drive caddy,[8] a second hard drive can be inserted into a machine – but as it is removable it doesn't require a power supply at all times, only when it is being used to record a back-up. This is often a cheaper option than using USB hard drives, especially if you are reusing old hard drives from other machines. When the machine is booted up the operating system can read and write from the hard drive, allowing individual files or the entire hard drive to be copied across. When the operation is complete the machine can be shut down and the second hard drive removed and securely stored in another location. In the event of a loss of data the removable drive can be re-inserted and the required data restored.
- Using a digital tape drive,[9] a copy of the hard drive can be written to tape and then removed and securely stored. As with the removable hard drive, data can be read back from the tape, although the process for doing this can take longer.

Removable media represents an extra level of data security because the data they contain

5. Wikipedia, Malware. en.wikipedia.org/wiki/Malware
6. Wikipedia, RAID. en.wikipedia.org/wiki/RAID
7. Wikipedia, Disk image. en.wikipedia.org/wiki/Disk_image
8. Wikipedia, Drive enclosure. en.wikipedia.org/wiki/Disk_enclosure
9. Wikipedia, Tape drive. en.wikipedia.org/wiki/Tape_drive

Box 6.2.

Removable storage media for backing up

For offline storage and backing up of data there are a number of different technologies available. Which is the best-suited depends primarily upon cost, the storage capacity, the viable lifetime of the storage media, and whether or not the media can be reused. Selecting which technology to use should be considered part of developing backing-up procedures.

For day-to-day backing up of small volumes of data – such as the daily/weekly backing up of current work cheap reusable USB sticks or flash memory cards are the simplest option. For the amount of data stored they are faster to use than DVDs, and although more expensive than DVDs when you factor in the number of times they can be reused, it is a fraction of the cost of a DVD.

For longer-term archiving of data, CDs, DVDs and Blu-ray discs are the cheapest and easiest option. For infrequently used data, or for regular back ups of user files, they are a simple way of preserving files. Optical discs can be a reliable means of storing data for a decade or more if the contents of the discs are indexed to indicate their contents clearly, and are stored in optimal conditions. While you can reuse CD-RWs and DVD-RWs, they can take more time to wipe and re-record than using other reusable technologies such as USB sticks.

For larger quantities of data, external hard drives, digital data storage (DDS) tapes and hard drives mounted in a drive caddy are the best option. Digital tapes are good for cheaply backing up a single large volume of data, but take longer to record and to read compared to hard drives. External hard drives are a simple option to mirror a user's files on a large drive, and are very simple to use because they are organised in the same way as the computer's internal hard drive. Internal drives mounted in a drive caddy are cheaper to use than consumer-oriented external hard drives, but require a little more knowledge to use as it may involve re-configuring the hardware or operating system permissions in order to access the drive.

Table 6.1.

Profile of removable storage media

Media	Capacity, gigabytes	Cost*, £UK	Cost per gigabyte	Reusable**	Life, years***
CD-R disc	0.7	0.2	0.3	no	30-50
CD-RW disc	0.7	0.6	0.9	x1,000	15-30
DVD-R disc	4.4	0.3	0.07	no	30-50
DVD-RW disc	4.4	0.5	0.1	x1,000	15-30
DVD-R dual layer (DL) disc	8.0	1.5	0.2	no	30-50
USB memory stick	0.1-256	1-600	2-4	x100,000	10-15
Compact flash/Smart media	0.25-128	6-750	2-3	x100,000	10-15
Micro-drive	0.1-8	20-60	4-7.5	yes	5
Blu-ray disc	25	2	0.1	no	30-50
Blu-ray dual layer disc	50	7	0.2	no	30-50
Digital data storage (DDS) tape	2-160	2-25	0.2 1	yes	10-12
USB external hard drive	320-2,000	60-300	0.12-0.18	yes	8-10
Hard drive in drive caddy	500-3,000	60-200	0.06-0.12	yes	5-12

* Cost per unit – for CD/DVD/Blu-ray discs this is based on the price for a single disc when purchased in packs of 10 to 25. Price is the recommended retail price, averaged across a number of brands, sourced from amazon.com

** An x figure represents the number of times the media can be reused before errors are likely to make it unusable. If "yes" then the media can be reused over the expected working life of the technology. For USB sticks/memory cards, this refers to the number of write operations which, because of the way the storage is configured, gives perhaps a fifth of this figure as complete write, delete and re-write operations due to the way the internal electronics of flash media function.

*** Figure represents the optimal storage lifetime of the technology before the media becomes unreadable – can be much less if not optimally stored and cared for, especially optical discs.

can be stored securely – for example, inside a fire-proof safe. This protects against theft and disasters. The difficulty is that restoring just a few files from removable bulk storage devices can take time. For this reason it is often more convenient to use a file server machine connected to the local network and allow users to back up files from across the network. The data held on this machine can also be backed up to a bulk storage device, such as a tape or removable drive. If restoring files on other computers on the system, it will be relatively faster to read back files from the central server than having to locate and install removable media. However, if there is a disaster, then the back-up of the central file server can be used to restore all the important files held on the network.

Where there are a number of machine sharing a network, using a file server is likely to be more efficient than individually backing up each system on the network – both in terms of the administrator's time and the hardware required. *Windows 7* has a built-in software application for network-based back-ups; for *Windows 7* and earlier Windows systems there are also various third-party applications that run across a network to automate backing up to a central server. For Linux systems there are a number of free network back-up and archiving applications, such as *Amanda* or *Bacula*. For Mac OS there is also a built-in application, *Timemachine*, which will back up across a network or to an external hard drive, and third-party applications are also available. Another advantage of network back-up software is that it can incrementally back up the contents of a desktop PC while it is in use, meaning that PC need not be left on to perform back-ups when the operator is not using the machine.

The concern with any backed-up data, particularly removable media, is the security of the information stored. With file servers it is possible to configure additional security measures, such as hard drive encryption,[10] to preserve the security of the data stored on the machine.

Similar encryption options are possible with removable media. The difficulty is that encryption uses more processing power, and that in turn increases power demand overall if routinely used for servers or desktop PCs. Also, if the encrypted media degrades or is corrupted, compressed or encrypted data is also more likely to suffer a catastrophic loss of the whole block of data rather than the corruption of one or two files within the back-up.

6.2.3. User-based back-ups

Now let's look at the computer user. The benefit of the user backing up is that, because they know which files are the most important, it is possible to target which sections of the hard drive are backed up. This can generate a much lower demand for data storage, making it a popular option. Because of this a wider range of back-up technologies are available for use. How, and how often the user backs up should integrate with the types of activity carried out on the system, the sensitivity of the data, and the frequency with which that data is changed. What's important is that those carrying out the process know how to create reliable back-ups, how to configure the programs used to create the right kind of data format, and ensure that the back-up media is tested afterwards to be certain it is usable.

Box 6.2/table 6.1 outline various storage media and their characteristics. Which is the most appropriate depends on the skills of the user, on the costs of the media, and most importantly on the scale of data to be stored. Some media can be reused to reduce costs and ecological impacts. Due to their high environmental impact, USB sticks and memory cards should be regularly reused many times for as long as possible. Although single-use optical discs cannot be reused, their benefit is that they have a long storage life. For these reasons USB sticks and memory cards are better for routine daily/weekly backing up, while optical discs are preferred for the long-term archiving of data and the offline storage of infrequently used data. While there is little hard data available, CDs, DVD and Blu-ray discs have a similar environmental footprint even though they have very different storage capacities. For the greatest efficiency, use the largest capacity disc that's suitable for the scale of back-up operation required.

Creating CDs, DVDs and Blu-ray discs requires the use of a disc mastering application.[11] These come as standard on all current operat-

10. Wikipedia, Disk encryption. en.wikipedia.org/wiki/Disk_encryption

11. Wikipedia, Optical disk authoring. en.wikipedia.org/wiki/Optical_disc_authoring

ing systems, although third-party applications are often used as they give more functionality. The mastering program bundles up the data files into an image of the disc to be created, then writes that image to the blank disc in a single operation. Third-party applications usually provide extra, often proprietary options to configure the format of the disc, and to enable the creation of other disc formats such as audio CDs and video DVDs.

CDs/DVDs created on Windows machines have traditionally been a problem due to the use of Microsoft's proprietary format for creating data discs – which restrict their compatibility with non-Windows systems. Since Windows Vista, users have the option of creating a *live file system* format, which is only compatible with Windows Vista/Windows 7; or a *mastered* format, which is more widely compatible with Mac and Linux machines. MacOS and Linux machines usually create discs using the ISO9660 international standard, and so are more broadly compatible across different machines and operating systems.

A problem with keeping an archive of many CDs or DVDs is that it's not possibly remember what is on every disc – and it's very difficult to handwrite the contents on the disc itself. The solution is to capture a directory tree listing of the directory and file names on the disc and store it as a text file. Then, instead of searching the actual discs one at a time looking for a file, search the text files containing the directory trees using the "find" tool of the word processor/text editor. Microsoft keeps a guide to capturing directory trees on its website.[12] For Linux and MacOS, the Unix command *tree -if 'path_to_directory' > 'file_name.txt'* will create a directory tree which can be captured as a text file.[13] While it takes a few minutes to make a directory tree and store it, keeping a tree of each disc in the archive can cut the amount of time it takes to find a file. That's because the search process can be automated by word searching the text files to find which disc contains the data, rather than manually searching each disc to find the data.

12. Microsoft (April 2011). How to add the Print Directory feature for folders in Windows XP, in Windows Vista, in Windows 7. support.microsoft.com/kb/321379

13. See the manual page for tree command at linux.die.net/man/1/tree

6.2.4. Off-site and online storage

The risk with holding data in a single location such as a home or office is that disasters can happen. Buildings can catch fire, flood, or the equipment can be stolen. For this reason keeping data in another location is advisable, especially irreplaceable data. The simplest method is to create two back-up copies and store one of those in a different location. The difficulty is organising how the data will be transported to the other location, whether it is secure enough to hold the data, and accessible if it becomes necessary to retrieve the data.

Another option for desktop users is to manually back up their machine to a file server in a more secure location, either within the same building via the local network, or in a different building using an internet connection. This can be done in many different ways:

- A formal network archiving program, which will bundle up data and move it to a server;
- A network service, such as file transfer protocol (FTP), to allow more secure access to storage space on a local or remote server machine – for sensitive data this transfer can be made using an encrypted connection;
- A networked file system, such as shared folders on a Windows network or a networked file system on a Mac/Linux network, to share files directly between computers; or
- A more secure virtual private network (VPN),[14] which allows files to be shared to other linked computers across the internet (both *Windows 7*, MacOS and Linux systems are supplied with the software required to configure VPN connections, allowing computer systems to securely back up to a remote server).

The problem with network back-ups is that, even for a handful of machines, this represents gigabytes of data being routinely transferred across the network – and that requires a lot of energy (we'll investigate this in unit 7). The shift to faster networks, which use comparatively more power compared to the older/slower standard, is in part driven by running more operations across local networks – such as data back-ups. While the centralised/automated backing up of PCs from a server is very simple to organ-

14. Wikipedia, Virtual private network. en.wikipedia.org/wiki/Virtual_private_network

ise, the use of the network for backing up can create a heavy drain on the network's capacity.

Seeking the least ecologically damaging route to backing up requires us to value to the data we are copying against the impact this action creates. In many cases backing up across a small wired network, using 100 megabit rather than gigabit speeds, will use less energy than a wireless connection. Backing up more than two or three gigabytes of data to an online service that keeps files live 24/7 is likely to consume more energy overall than reusable storage media.

In devising a policy for backing up it is necessary to weight these different factors to produce the least ecologically damaging option.

6.2.5. "The cloud"

The greatest movement in computing at present is the storage of data in "the cloud". Cloud computing[15] has evolved with the ever greater use of mobile handsets and computers. Ten years ago this wasn't viable, but with the greater availability of broadband services, and the development of cheap high-capacity servers, storing large quantities of data online is now a viable option.

Apple's mobile devices, Google's Android operating system, and Microsoft's new Windows 8 system are tied to the use of cloud storage. Online services, such as banking or social media, also use data stored on many machines which are part of the cloud. Unlike an identifiable server, where you "know" where your data is stored, how the cloud handles and stores data isn't managed by the user – it's automatically determined by the rules that govern the cloud system. Cloud storage operators often run multiple data centres, and files can be spread across one or many of those locations depending upon which represents the most efficient way to move and hold the information concerned.

However, the important issue about "the cloud" is that it's not just a data storage system; it represents a whole business model for online commerce. If we look at the services using cloud storage, enforcing intellectual property rights in the digital domain is often an important element of the way they operate. This is best explained by looking at Apple iTunes, or Amazon's Kindle e-book system. In order to enforce intellectual property rights in the digital world it is necessary to track the use of data. The most secure way to do that is to link the storage of data on an individual's computer to the data stored on the organisation's cloud servers. This is enabled by having to register the device(s) which use these files with the company's information systems – for example Apple's music files or Amazon's e-book. In this way services can be provided, the movement of data tracked, and the use of intellectual property policed.

While there has been great interest in the efficiency of cloud storage, there has been little debate about the enforcement of tighter intellectual property rights over digital data and the effect this has on society. Human culture has traditionally been shared, and that's been the key to the development of knowledge and learning. In contrast, lodging information in "the cloud" enforces strict boundaries on people's ability to share and re-work/re-imagine the elements which make up our culture.

One of the most debated points about the cloud is the way in which these systems log large quantities of data about individual's patterns of data use and communication. This enables all sorts of profiling activities in order to identify people's attitudes and interests. Primarily this data has commercial value to marketing and public relations agencies – and that has a whole number of negative ecological impacts given that such a large part of the internet's capacity, and increasingly direct mail, is tied up conveying sales information.[16] More controversially, this same data has increasing relevance to the state security role of police and security services and by extending the powers of the state over people's data these systems could potentially be used to police the freedom of thought, expression and communication enabled by digital communications.

Finally, through various online services people are lodging more of their data online. As a result the demand for processing power and data storage is increasingly being driven by the

15. Wikipedia, Cloud computing. en.wikipedia.org/wiki/Cloud_computing

16. Taylor and Koomey (2008). Estimating Energy Use and Greenhouse Gas Emissions of Internet Advertising. imc2.com/Documents/CarbonEmissions.pdf

Box 6.3.

Basic rules for making back-ups on removable media

However you choose to back up files, there are certain basic rules to ensure that the process works well, and that the data stored is readable in the event of the loss of corruption of files:

When regularly backing up current work using reusable media, always store the back-up in a secure location – if storing sensitive information, the back-up should be stored with a level of physical security that reflects the importance of the data it contains.

If data is to be archived, and is of great value, two copies of the archive media should be created, and one of those should be stored in a different location. This ensures that the data it contains can be recovered even if the copy kept locally is lost, damaged or destroyed.

Always check the readability of back-ups after creating them

- When making CDs, DVDs and Blu-ray disks, always ensure that the disc mastering program performs a verification read after writing the disk to check its content.
- Back-up applications should automatically check that the files moved are readable, but if manually copying files to a USB stick or external drive, copy them back into an empty directory to ensure that the back-up is readable.

Mechanical external/internal hard drives and micro-drives can have their lives shortened if shaken or dropped, and must to be stored in a location that is well away from strong magnetic field (electric motors, video displays, etc.). USB sticks and memory cards should also be carefully stored in secure locations as mechanical stress (e.g. being dropped or subjected to heat and cold) can break the internal electrical contacts and render them unusable.

Magnetic media, such as digital tapes, must be kept at an even temperature, protected from extremes of heat and cold, and shielded from magnetic fields (ideally, if stored for a long period, they should be kept inside a metal container to exclude all magnetic fields).

When archiving data onto CDs/DVDs, always label the disc with an identifiable index – for example, the date the back-up was made and a few words to describe the content.

When making a large number of archive CDs/DVDs, create a listing of the directory tree on the disc and store it as a text file – these files can be searched when you want to find a stored file, rather than manually searching the discs.

Indexing the CDs/DVDs according to the date they were made, using the format year-month-day-disc number (e.g. 2012-03-12-01). This creates a unique index key to identify the disc, and keeping the discs in date order allows them to be accessed quickly when required.

Optical discs require careful storage to maximise their working life

- Store the discs in a CD/DVD folder to protect them from damage, or if making a large number of discs you can store them on the spindle containers that the blank discs are supplied in.
- Never expose the surface of the disc to strong light for long periods of time, particularly sunlight or near to bright fluorescent lights (ultra-violet light damages the polymers in the disc).
- Keep the discs in an environment that has an even temperature, and protect them from extremes of heat and cold.
- Handle the disc by its edge and centre hole, and avoid touching the surface of the disc as the grease/dirt from fingertips encourages dust and fine grit to stick to the surface.

needs of large data centres rather than individual computers. The ecological advantages of the cloud[17] are predicated on the basis that this new demand for computing power can be managed more efficiently in a large data centre than on a large number of small systems. The difficulty is that by keeping data in the cloud you are putting all your trust in the availability of the online services. If the service fails (e.g. the problems with the Blackberry service in 2011),[18] or the user's account is compromised or blocked, or the company providing the service collapses, access to data in the cloud can be lost.

To have secure access to our most valued information at all times, it is necessary to keep back-ups where they can be accessed locally. The difficulty for the cloud computing model is that if users keep back-ups of their information on their own machine for the purposes of security, many of the arguments for the ecological efficiencies of the cloud system disappear.

6.2.6. Manually transporting data – the "Sneakernet"

In the early days of personal computing dial-up internet connections were very slow. While today's broadband internet connections run at a few hundred to a thousand kilo-bits per second, early dial-up modems ran at 0.3 kilo-bits per second. For this reason the fastest way to move data wasn't via a network, it was physically carrying floppy disks in a bag. As a metaphor for the idea of moving data on foot, this method of data transfer became known as the "Sneakernet".[19]

While we may focus on the use of the internet, it is arguable that a person walking with a bag full of data can move information faster than many electronic networking technologies. That's because while the internet is fast for everyday small files and email transfers, when considering the movement of giga-bytes of data the throughput of the network can be very slow. The different ecological impacts of moving data via electronic networks, or manually carrying or sending storage media via the postal system or a courier service, are also significant.

There are various figures for how much energy it takes to move a gigabyte of data across the internet.[20] A general figure is somewhere around seven kilowatt-hours per gigabyte,[21] and, with power generation producing around 600g of carbon dioxide per kilowatt of power, that entails the emissions of around 4.2 kilos of CO_2 per gigabyte of data. An optical disc with a plastic jewel case takes around 16 kilowatt-hours of energy to produce, releasing about a kilo of CO_2 as a result of its production; and while there are no detailed figures, a DDS tape might take three times that because it is more complex product containing mixed materials.

Figure 6.3 shows a comparison of moving data via the internet (shown as hatched bars) and using storage media via Sneakernet (shaded dots). For the given storage capacity of each media type the impacts of moving that data electronically via the internet or creating the back-up and then transporting it are calculated. There are studies of the ecological impacts of the postal service,[22] and conservatively these figures have been multiplied by a factor of five to take account of the higher package weight. The costs of moving data online have been assumed to be £1 per gigabyte, while the cost of sending an optical disc/DDS tape via post, and purchasing the blank media, is also calculated for comparison. What the results show is that:

- Producing an optical disc and a plastic jewel case to protect it takes around 16kW-h of energy and emits a kilo of CO_2; however, as moving data via the internet is also energy intensive, sending a DVD via the post has a slightly lower impact than sending the same data across the internet.

17. WSP/Accenture (2010). Cloud Computing and Sustainability: The Environmental Benefits of Moving to the Cloud. www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Sustainability_Cloud_Computing_TheEnvironmentalBenefitsofMovingtotheCloud.pdf

18. ZDNet (October 2011). BlackBerry issues statement over downed services. www.zdnet.com/blog/btl/blackberry-issues-statement-over-downed-services/60450

19. Wikipedia, Sneakernet. en.wikipedia.org/wiki/Sneakernet

20. Koomey et. al. (August 2009). The energy and climate change impacts of different music delivery methods. download.intel.com/pressroom/pdf/CDsvsdownloadsrelease.pdf

21. Taylor and Koomey (2008). Estimating Energy Use and Greenhouse Gas Emissions of Internet Advertising. imc2.com/Documents/CarbonEmissions.pdf

22. Pitney Bowes Inc. (2008). The Environmental Impact of Mail: A Baseline. www.pb.com/bv70/en_US/extranet/landingpages/Environ_Impact_Mail_Web.pdf

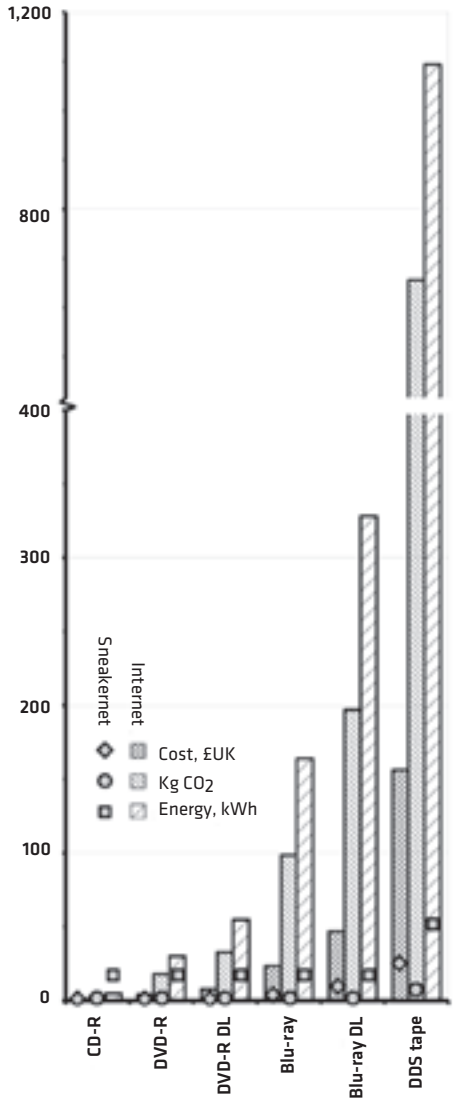
- Moving less than two or three gigabytes of data via the internet is more efficient in terms of energy and carbon emissions than posting an optical disc however, that's primarily because of the high energy and carbon density of the disc. If reusable media were used, such as a rewritable CD or DVD, that would significantly reduce the impacts. Even if only reused four times, a CD-RW disc would then be more efficient than the internet.
- Taking the cost of data at £1 per gigabyte, if using commercial postal services it is cheaper to send up to two to three gigabytes via a network than via an optical disc.
- While for everyday small-scale data movements (e.g. web pages download or email) this analysis doesn't have a great relevance, where these figures are very significant is the use of networks for backing up data. For example, Google has recently launched a large-capacity cloud storage service.[23] If used for backing up, routinely moving that quantity of data via the internet would take more power and emit more carbon than storing a DVD at a location where it could be held securely.

The implication of this is that backing up large quantities of data off-site has the lowest impact if carried out using storage media, such as optical discs, rather than using electronic networks. The difficulty is organising how the data will be transported to the other location, whether it is secure enough to hold the data, and accessible if it becomes necessary to retrieve the data. How these issues are dealt with, and solutions found, would have to be tackled as part of the process for planning and implementing a back-up procedure.

When data CDs first became widely available, but the use of computer networks was still in its infancy, the publication of digital content often used CD-based distribution as a means of moving large quantities of information in a structured format. Today, with the advent of high-speed networks, disc-based distribution of publications is going out of fashion in favour of network-based/web distribution of content – and more recently the movement of large blocks of data using peer-to-peer file sharing.[24] The relative ecological impacts of networked versus stored data distribution indicate that we may need to re-evaluate how information distribution is planned. As general rule, where the content is relatively static or is

Figure 6.3.

Comparison of Internet versus Sneakernet impacts



23.BBC News (April 2012). Google Drive to offer free storage in the cloud. www.bbc.co.uk/news/technology-17831725

24.Wikipedia, Peer-to-peer file sharing. en.wikipedia.org/wiki/Peer-to-peer_file_sharing

not regularly updated, the impacts of distribution via optical discs may be lower than holding that information live online. For example, the data required to install an operating system today fills a DVD, and arguably the ecological impacts of distribution on DVD would be marginally better than via the internet and yet the trend today is for greater network-based software distribution to install and maintain operating systems, or to run software remotely within the a cloud system.[25]

6.3. CONFIGURING THE DESKTOP TO IMPROVE DATA SECURITY AND USABILITY

Another significant area of everyday data loss is user error or system crashes which affect the work/program in current use. While many programs are getting better at recovering data following a system crash, it is possible to improve the way we use programs to reduce the likelihood of losing data.

Firstly, many programs – particularly word processor/office applications and text editors – can routinely create a back-up copy. This feature is enabled from the user configuration options for the program. While some will create a back-up file with a slightly modified file name, others will set up a specific directory to store back-up copies of the files/documents being worked on. Each time you save a copy of the file it stores the previous version as a back-up. If a crash corrupts the original file, or the contents of the file are accidentally lost or erased, or if saving the file erases some valuable data that had existed in the previous version of the file, it is possible to revert to the back-up copy and recover lost data.

Secondly, it's always a good idea to regularly save work. Some programs allow you to configure an auto-save option to save the document at a specified time interval. Rather than remembering to save, the program will take care of that for you. The difficulty is that sometimes, if you make a mistake, the auto-save option can over-write the back-up version of the file being worked on. The trick is to set the auto-save interval to be short enough that you will not lose data if the system crashes, but not so short that it repeatedly stores the current document, which restricts the ability to reuse the previously saved copy of the document if data is lost during editing.

Thirdly, most programs allow you to "undo" the editing or processing of data. The more steps that it is possible to undo, the greater the chance that mistakes can be corrected. The main restriction on the undo facility is the amount of memory it takes up. Your options are limited if using a machine with limited memory space. If the machine has a lot of memory then it is possible, from the programs configuration options, to increase the number of undo steps. How many steps you set depends upon your preference – 30 to 40 steps is probably sufficient for most applications.

Finally, most operating systems use a "waste basket" to store files which have been removed from the hard drive. This means that you have to occasionally go in and empty the waste. While it is tempting to simply "empty trash", or use a delete command to directly remove files from the hard drive, the waste basket feature serves a very useful purpose – *it stops the user deleting files accidentally*. Also, do not blindly use the "empty trash" command – always check what's in the waste bin before erasing the contents.

25.Wikipedia, Software as a service. en.wikipedia.org/wiki/Software_as_a_service

Box 6.4.

Information storage check-list

The more common and easily accessible a resource the less we need to back it up, the more unique or expensive a resource the greater the need to create a back-up copy – value the files on the computer using this general principle.

Separate the information held on the machine into different directories – one directory tree for current work, another for already archived files, and another for low value/superfluous files – to simplify the process of backing up and reduce the amounts of data requiring storage.

Identify roles and responsibilities for backing-up data, and agree policies and procedures for regularly backing-up data:

- System administrators/managers should be responsible for backing up at the system level, and co-ordinating the use of network-based software for backing up to a central server.
- Computer users should be responsible for backing up current work and other important data using removable storage.

Backing up with the least ecological impact requires that we match the type of data being stored with the characteristics of the storage media used:

- Data that is being archived for long periods, and other data which changes infrequently, should be backed up on optical discs as these have the lowest ecological impact and a long storage life.
- Routine daily/weekly back-ups should be made using reusable media, such as USB sticks, memory cards or removable hard drives.

When creating archives on optical disc, create a text file containing a directory listing of the disc's contents – searching this disc using a text editor (or integrating it into a desktop search system) takes less time and energy than physically searching each disc.

Using the internet for backing up has a high impact when moving many gigabytes of data. While cloud computing is becoming popular, always consider the alternatives to overcome the data security and ecological drawbacks of using large amounts of online data storage.

When moving large quantities of data between two systems, posting optical discs/digital tapes can have a lower ecological impact and financial cost than transfer over a network.

Configure the features of desktop applications to preserve data: Always create back-up copies of files being worked on; auto-save at regular intervals; set/increase the number of “undo” operations; and always use the waste basket rather than directly deleting files.

A practical guide to sustainable IT

This practical guide to sustainable IT offers a detailed, hands-on introduction to thinking about sustainable computing holistically; starting with the choices you make when buying technology, the software and peripherals you use, through to how you store and work with information, manage your security, save power, and maintain and dispose of your old hardware. Suggestions and advice for policy makers are also included, along with some practical tips for internet service providers.

Written by IT expert and environmentalist Paul Mobbs, the purpose of the guide is to encourage ICT-for-development (ICTD) practitioners to begin using technology in an environmentally sound way. But its usefulness extends beyond this to everyday consumers of technology, whether in the home or office environment. We can all play our part, and the practice of sustainable computing will go a long way in helping to tackle the environmental crisis facing our planet.

This is also more than just a "how to" guide. Mobbs brings his specific perspective to the topic of sustainable IT, and the practical lessons learned here suggest a bigger picture of how we, as humans, need to live and interact in order to secure our future.

The guide is divided into 12 sections (or "units"), with each unit building thematically on the ones that have come before. They can be read consecutively, or separately. The "unit" approach allows the sections to be updated over time, extracted for use as resource guides in workshops, or shared easily with colleagues and friends.

The guide has been developed on behalf of the Association for Progressive Communications (APC), with funding support from the International Development Research Centre (www.idrc.ca). It is part of a APC's GreeningIT initiative, which looks to promote an environmental consciousness amongst civil society groups using ICTs, and amongst the public generally. Other publications and research reports completed as part of the GreeningIT initiative can be downloaded at: greeningit.apc.org

APC

IDRC | CRDI