

MANAGING THE INFORMATION LIFECYCLE

INFOKIT



www.jscinfonet.ac.uk/infokits/information-lifecycle

Managing The Information Lifecycle.....	1
What Is Information Lifecycle Management?.....	4
Taking A Holistic Approach To The Management Of Information	4
The Advantages Of Taking A Holistic Approach.....	4
The Information Lifecycle Model	5
Variations To The Lifecycle Model.....	5
Rationale for model selected	5
From theory to practice	6
Creation	7
Creating The Right Information.....	7
Creating Reliable Information	8
Data Quality.....	8
Data Currency	8
Status And Version Control	8
The Right People Creating Information	9
Proof Of Provenance.....	9
The Dangers Of Over-Restricting Access.....	9
The Benefits Of Unfettered Access.....	9
Creating Information In Appropriate Formats	10
Capturing The Right Metadata.....	11
Metadata Types.....	11
Active Use.....	11
Define The Purposes For Which Your Information Can Be Used	12
Locating And Accessing Information.....	13
Who Needs Access To The Information?	13
Restricting Access.....	14
Semi-Active Use	14
What Information Is Held And Why?.....	15
Defining The Journey From Active, To Semi-Active Use.....	15
Retaining The Context.....	16
Continued Maintenance	16
How Long Should Information Be Kept For?	16
The Drivers Governing Information Retention.....	16
Risk Management	17
Assessing The Value of Information To the Organisation	17
Defining Value	17
Appropriate And Cost-Effective Means Of Storing Information	18

Storage Options	18
Out Of Sight Shouldn't Mean Out Of Mind	18
Is Your Information Safe?	18
The Overlooked Importance Of 'Good Housekeeping'	19
External Service Providers	19
Final Outcome	19
Ensuring Continued Access To Information	20
Controlling The Disposal of Information.....	21
When Has Deleted Information Really Gone?	21
An Auditable Process	21
Disclaimer	22

What Is Information Lifecycle Management?

The information that your institution creates and uses can either represent an asset or a liability. Into which of these camps it falls is largely dependent on how it is managed. Put simply, the concept of information lifecycle management is about making sure you ask yourself the right questions at the right time regarding the management requirements of internally produced information. It does this by breaking down the 'lifecycle' that all information moves through into four distinct phases and identifying what are the most pertinent issues that influence how information should be managed during each phase.

Consideration of these issues at the outset helps ensure the effective management of your internal information throughout its entire lifecycle: from cradle to grave.

Taking A Holistic Approach To The Management Of Information

Unfortunately there is not a 'one size fits all' specification for what is required to ensure good information management. A host of factors, both from within the institution (operational) and from outside it (regulatory, legal) will need to be considered. Moreover, these influences will vary over time and are dependent upon the type of information being created, its purpose, content and usage. What is therefore required is a consistent framework within which the management of information can be considered. This framework needs to be flexible enough to accommodate the variety and range of information now being created and ultimately to ensure that the right decisions regarding its management are being considered and made at the right time throughout its 'lifecycle'. This kit outlines the basic lifecycle framework, underpinning the related Records Management and Email Management strands in providing a practical approach to Information Management.

The Advantages Of Taking A Holistic Approach

Following the approach outlined within this infoKit will help ensure that the management of information created within your institution is consistent, proportionate and fit for purpose.

It is not dependent on the installation of expensive new technology nor does it assume that all information will be captured and managed in a single place by a single system. Instead it seeks to provide a conceptual framework which can be applied whenever and wherever a new system or process is to be introduced, or as part of a review of the management of information created by existing systems and processes.

The following table illustrates some of the key advantages of managing information via the lifecycle method and how they are realised.

Advantage	How?
Consistency	By following a common model and addressing the same fundamental questions, this approach ensures a consistency in the way in which information is managed regardless of the particular system it is created by.
Inclusiveness	This approach is equally as useful when considering the management of ephemeral information and raw data as it is formal business records. It also applies regardless of the format of the information, be it electronic or hard copy.

Pro-activeness	Following the model requires you to look to the future and what the management implications for a particular type of information are likely to be in the months and years ahead, as well as those you face now. This helps you to forward plan and avoid unpleasant surprises.
Proportionality	It is up to the user to decide which elements of the lifecycle model are relevant to the information concerned and to their individual circumstances. It does not seek to impose a heavy management burden (and cost) where such measures are not warranted and where a more lightweight approach is called for.
Flexibility	Because of the generic nature of the model it is not dependent upon any particular technology. Its approach is likely to be equally valid as and when new technologies emerge, thus further demonstrating the advantage of consistency. The way in which the model's recommendations need to be enacted will inevitably vary from system to system and process to process, but the underlying principles on which they are founded will remain constant.

The Information Lifecycle Model

The model of the information lifecycle used as the basis for this infoKit is a simple one based on four main phases:

1. Creation
2. Active Use
3. Semi-Active Use
4. Final Outcome

Variations To The Lifecycle Model

It should be noted that this is not the only version of the information lifecycle in existence, nor does it in any way represent the definitive version. A quick search of the web will locate several examples of lifecycle models, some very similar to those outlined previously, others substantially different.

Alongside those using the lifecycle methodology as a framework for managing internal information are others who are adopting the same fundamental 'cradle to grave' approach to the management of [library materials](#) or [storage management](#). This helps demonstrate the wide applicability of the underlying concept.

Of those using the lifecycle model for information management, some have more complex models which include a [greater number of phases](#) whilst others [use fewer](#).

Indeed there are those who believe the concept of the lifecycle to be fundamentally flawed with regards to information management and argue instead that the notion of the records continuum represents a more useful and practical model, particularly when managing electronic information. This view largely originates from and is particularly [prevalent in Australia](#).

Rationale for model selected

The one thing which unites all of the different variations and approaches outlined above is their concentration on taking a holistic, 'cradle to grave' perspective. With this consistent core principle established, exactly how the lifecycle is defined can and should vary according to the purpose for which it is being applied. None should be considered wrong, so long as they are comprehensive and fit for purpose.

With this in mind we believe the version of the model we have adopted for this infoKit is the best fit for our purposes due to its following characteristics:

- a clear chronological structure

*"All records have a life cycle from creation/receipt (birth), through into the period of active currency (youth), thence into semi-currency, e.g. middle-aged closed files that are still referred to occasionally, and finally either confidential disposal or archival preservation. In the digital age it is especially important to introduce conscious management at the earliest possible stage as this will determine the ultimate extent of control over electronic material"*¹

- clearly defined phases which users can relate to
- a consistent focus which does not combine different concepts such as management operations (e.g. 'evaluation' or 'applying retention schedules') and chronological phases within the same model, and
- breadth of application through avoidance of narrowly defined professional jargon.

From theory to practice

The following sections map the four phases of the information lifecycle and raise the generic questions which it is suggested you should think about with regards to the information that is being, or will be, created. Of course the term 'information' covers a huge variety of formats, media, technology and content and it is likely that the generic questions outlined within the lifecycle model will need to be further refined to address differing needs.

That is why this infoKit also includes applied sections where the concept and framework of the information lifecycle has been used to provide additional and specific guidance on the management of particular types of information. At present this includes guidance on the [management of records](#) and [emails](#) but it is envisaged that this list will grow in the future as more candidates for inclusion are identified.

The objective of these sections is to provide guidance on the management of these particular types of information that has been further refined to make it as detailed and specific as possible. This guidance will always follow the same lifecycle model and will build upon the generic questions and issues raised within it, but will add further detail where the specific requirements of that type of information warrant it.

This approach ensures the consistency of management referred to earlier, avoids the duplication of material (as some guidance will always be the same regardless of whether we are talking about formal records, emails or databases) and ensures that the specific topics are covered in as comprehensive and relevant a manner as possible.

¹ University of Edinburgh - Records Management Policy Framework
http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RM_framework.htm

Hopefully these applied sections will make it even easier for you to apply the lifecycle model to your own individual situations and thus help make the leap from theoretical model to practical application.

Creation

Creating information is often the easy part and appears at first glance to require little thought or planning. After all the whole drive of the IT industry in recent years has been to empower the user to create and manipulate information as quickly and easily as possible. Quantity, however, does not necessarily equal quality. Indeed it can often help contribute against it. Sometimes this may not matter. A hastily scribbled 'to do' list serves its purpose so long as it reminds you of what you must do and when you must do it by - assuming of course that you can find it when you need it. It is of little consequence whether it is typed into a word document, listed in an online diary or scribbled on the back of the proverbial envelope. Nor does it much matter whether it contains spelling mistakes or abbreviations only known to you. A completely different set of rules and acceptable quality standards apply when drafting of your annual report, designing a system for handling contracts or entering data within your personnel database.

The questions to be addressed during this phase are largely aimed at ensuring that the information created is fit for purpose and that it is actually capturing appropriate and reliable content. They include the following questions:

- Are you creating the right information?
- Are you creating reliable information?
- Are the right people creating it?
- Are you creating it in the most appropriate format?
- Are you capturing the right metadata?

Creating The Right Information

Information may be created for many different purposes. It may be deliberately created solely to inform others through its content. Alternatively it may be that its creation is merely the by-product of a process where its true value lies mainly in providing evidence that the process in question has occurred (for example a receipt for goods purchased). Either way it is important that the information in question is fit for purpose.

An example of a typical set of information which generally serves its purpose well is that which is created as part of the administration of a business meeting:

- The agenda: provides required factual information such as details of where and when the meeting is to be held, what is to be discussed and in what order so that everyone comes to the meeting equally and fully prepared
- Background papers: provides attendees with the information they will require to make informed decisions during the meeting
- Minutes: provides an agreed evidential record of what took place during the meeting

When designing a new process or system, it is worth considering what information you should be creating and why. For example, when documenting a new process, is it sufficient just to know that a stage in the process was completed or is it important to also know who completed it and when? This will depend on the nature of the process being undertaken, its importance and formality and a host of other factors. The important thing is that these issues are considered.

It is also important to ensure a consistency of approach to information creation across operations. It is less than ideal if a project being managed in one department creates a comprehensive set of documentation, whereas a similar project in another department creates next to nothing. Defining an agreed document 'set' to be used in such circumstances can be beneficial in terms of ensuring the right information is being created across all projects. The [JISC infoNet Project Management infoKit](#) includes an example of such a document set for project management and how it is used to ensure that each stage of a project is documented appropriately. Adopting this approach also has the benefit of making it much easier to compare information between projects.

It should be remembered that it is also possible to create too much information. This might not seem an important issue in an age where it is so easy and cheap to transmit and store information but it can easily lead to hidden costs and considerable user frustration. For example, it might be very useful for the administrator of a database or content management system to know when a change has been made to its content, but a system which generates an automatic email informing all users of every edit made will result in thousands of unnecessary emails being sent and considerable frustration on the part of staff who repeatedly receive worthless messages.

Creating Reliable Information

It is important that people are able to trust the information they are using. Often they will be relying on it as the single source of truth against which decisions will be made. It is not difficult to imagine the risks to the interests of individuals and the organisation as a whole if the figures in a budget spreadsheet are wrong, an employee's home address on the staff database incorrect, or the plans given to a builder out-of-date.

Human error is not always to blame when unreliable information is created. Simple things such as the wrong date and time settings being applied within a system can instantly lead to inaccurate information being created with documents or database entries seemingly being created on the 1st January 1900. Likewise it is common to find all documents apparently being created by a single person, simply because they were responsible for creating the original template, or happened to install the software on the network.

Data Quality

Where human error is to blame there are often ways in which the risk of it occurring could be reduced. Designing a system where the administrator is expected to type in each member of staff's name as a free text field is almost certain to lead to spelling mistakes, duplicate entries and other inconsistencies. These would largely be eliminated if the administrator was forced to select each name from a pre-defined pick-list. The margin for error is reduced still further if that pick-list could be provided direct from the main personnel database.

Data Currency

Users need to know that the information they are relying on is as current and up-to-date as possible. There is nothing more off-putting to a user than finding that the information they are relying on was last updated two years ago and has been superseded many times since. Who is to be responsible for updating content, how frequently this must be performed and what the 'triggers' to doing it will all need defining.

As ever, fitness for purpose is what counts. For example, many institutions will perform a proactive annual check of the personal information they hold on their staff to ensure it is all current (whilst also reacting to any changes they are informed about throughout the year).

Status And Version Control

However frequently your information is updated it is important that users can be confident that they are looking at the most recent version. Also that they know whether or not it is still information in draft and liable to further change, or whether it has been finalised.

Some examples of simple but effective measures for ensuring that version control is ensured can be found in our [Managing Information to Make Life Easier: A Guide for Administrators](#)

The Right People Creating Information

It's probably a safe bet that you wouldn't want the college nurse to be responsible for drafting your commercial contracts, nor for your procurement officer to be writing advice for students on meningitis. It is important that those responsible for creating information possess the authority and the ability to do so. This may require careful consideration of which functions and/or individuals within the organisation are trusted to create the information in question.

Proof Of Provenance

Once such responsibilities have been defined, it is usually easy to enforce them through system logins and the privileges attached to them (read only, read/write, system administration etc). However, it may be more difficult to automatically control at a systems level when it comes to creating documents, spreadsheets or emails. There is usually nothing to physically prevent anyone creating a document and calling it 'Annual Report 2007' or sending an email to the head of a neighbouring institution requesting a merger. These may be extreme examples that would be easy to spot as being bogus, but they serve to illustrate the point. The ability to prove its provenance and who was responsible for creating a particular piece of information could well be vital for protecting your institution's legal interests.

*"Provenance is the origin or source from which something comes, and the history of subsequent owners (also known in some fields as 'chain of custody')."*²

If this cannot be achieved by system functionality other possibilities may be to rely on the use of templates to which only relevant people have access, other types of authentication (such as biometric authentication), or the use of security controlled storage and publication facilities.

*"In information technology, biometric authentication refers to technologies that measure and analyze human physical and behavioural characteristics for authentication purposes. Examples of physical (or physiological or biometric) characteristics include fingerprints, eye retinas and irises, facial patterns and hand measurements, while examples of mostly behavioural characteristics include signature, gait and typing patterns"*³

The Dangers Of Over-Restricting Access

Of course putting too many barriers in the way of creating information may well be counter-productive and lead to user frustration and disengagement. Mandating that only the academic who authored a paper can deposit it in your repository may help ensure the integrity of the system's contents but is also likely to reduce the number of submissions from busy staff with little time or appetite for such administrative tasks.

The Benefits Of Unfettered Access

Much of the new generation of technology that is emerging is founded on the principles of collaboration and sharing. Social software applications such as wikis work on the basis that anyone who wants to contribute content is free to do so. It is for others to judge the quality of the

² Taken from the Wikipedia entry for Provenance on 20 July 2007
<http://en.wikipedia.org/wiki/Provenance>

³ Taken from the Wikipedia entry for Biometric Authentication on 20 July 2007
http://en.wikipedia.org/wiki/Biometric_authentication

information they create and to either embrace or ignore it as appropriate. The online encyclopaedia, [wikipedia](#) is perhaps the best known example of this.

This approach may well be highly beneficial where the objective is to get as many people as possible to contribute their ideas and opinions - especially in situations where everyone's point of view is likely to be equally valid. The important thing to ensure is that those wishing to make use of this information are aware of its mixed origins and the fact that it may represent opinions (as opposed to facts) based on differing degrees of experience and knowledge.

Creating Information In Appropriate Formats

When designing a new process or system which will create information it is advisable to think about the format in which it would be most appropriate to create it. Issues regarding its expected longevity, security and access concerns, potential for future reuse and its possible evidential value may all have a bearing. The question of information format is often overlooked in favour of a concentration on its content which can prove a costly mistake. Once again it is about ensuring that what you decide is fit for purpose. It is the equivalent of selecting the right tool to complete a manual task: a small hand-pushed mower might be fine for a garden lawn, but virtually useless for preparing a football pitch.

The various purposes for which email is often used represent an all too common example of widespread misuse of a format for creating information. People often treat it as the equivalent of a phone conversation and say things or express views which they would never ordinarily commit to paper. People forget that an email is a written record which may come back to haunt them. At the other extreme, a significant amount of formal business activity is often conducted via email with contracts being agreed and projects signed off. Whilst this may represent a perfectly legitimate form of business communication, it is also often the case that organisations fail to manage email appropriately, leaving potentially vital business records to languish unmanaged in an individual's inbox.

The World Wide Web has revolutionised our ability to quickly and cheaply publish information. When compared with the cost and time involved in manually printing and distributing a publication, it is little wonder that more and more institutions are choosing to only publish online versions of publications. But what if decisions are being made against the content of that publication which means it suddenly becomes necessary to know exactly what it said at a particular point in time some months or years ago? Unless your web content management system has sufficient capability to track changes and 'roll back' to how it appeared at a certain date it may be impossible to prove what your Research Ethics Handbook, for example, stated at the time a particular research project commenced.

An ever-increasing number of users are beginning to make use of externally-hosted social software services in which to create and store content. This could be a wiki being used to develop ideas within a project team, an online photographic service to store learning materials or even a 'virtual world' such as [Second Life](#) to create complex 3D models. Such technologies offer marvellous potential, but what would happen should the company providing the service suddenly go bankrupt or otherwise withdraw the service? Would you still be able to access the content you have trusted to it? There may also be questions to address regarding who actually owns the rights to the intellectual property of an object created in Second Life. Is it the individual user, the institution or Linden Labs who run this virtual world?

Lastly, there is the question of longevity and preservation. Much of the information we create will have a short lifespan of less than 5 years making questions regarding their longevity largely irrelevant. But a small percentage may need to be kept for far longer, perhaps for decades or even centuries. It may seem a real advance to scan your entire collection of paper files and store them on a handful of CD ROMs, but what if some of those records will need to be accessed in 70 to 100 years' time? In a few short years, it is more than likely that a CD-ROM drive on a new PC

will be as rare a sight as a floppy disk drive is today. Unless adequate preventative measures are taken those invaluable records may be as good as lost.

Capturing The Right Metadata

We may not always recognise it but metadata plays an invaluable role in allowing most staff to find and use information on a daily basis. Properties as simple as the title of a Word document or the name of the sender of an email in your inbox are all examples of metadata that we take for granted during our working day but would struggle to operate effectively without.

When designing a new process or system which will result in the creation of information, it is important to think through what metadata it may be required to create, not only to allow people to find it but for a host of other purposes. Whilst not all of the following types of metadata will always be relevant, it is useful to consider all possible aspects before deciding which are necessary and which can be discounted.

Metadata Types

According to the [MANDATE Toolkit](#) produced for JISC by John Wheatley College, the following main types of metadata can be identified:

- Bibliographic (e.g. "a picture of the Eiffel tower taken by Bob")
- Administrative (e.g. "taken on 19/04/2005")
- Legal (e.g. "all rights reserved")
- Preservation (e.g. "requirement to view as jpeg")
- Technical (e.g. "jpeg format" ; "85.8kb file")
- Educational (e.g. "an illustration of construction using cast iron"; "UK Education Level 11")
- Structural (e.g. "single file")

Whilst not a definitive list it does serve to illustrate the breadth of types of metadata to be considered. As ever, it is important to be selective, particularly if you are requiring users to input metadata manually. Wherever possible, systems should be designed to extract metadata automatically, thus ensuring its consistency and accuracy. If you are relying on users to add their own metadata, consider the level of burden this may place on the user and look for ways to ease it where possible, such as through the use of pick-lists rather than free-text fields.

Equally important is the need to ensure the consistency of metadata being applied to information shared across systems and processes. In an ideal world approved keywords would be taken from a single 'master' source and applied where necessary (for example, the list of budget codes being provided by the finance system, and staff names from the personnel database). If this isn't possible, it is important to at least achieve a degree of consistency regarding what metadata is captured for a particular type of information and in what format it should be expressed (all dates to be DD/MM/YYYY etc).

Active Use

Organisations are awash with information. Although the topics discussed in the previous section should help place some management controls around the creation of information, it is also necessary to think about how that information can and should be used. This is not to try to prevent innovation or stifle creativity, but to balance these positive aspects with the potential risks to the interests of your organisation which could occur as a result of the unfettered and uncontrolled use of information.

The active use phase of the lifecycle can best be described as the period during which information is in constant or regular use, primarily as part of the purpose for which it was originally created. There are no hard and fast rules for defining exactly what constitutes active use, but information which matches some or all of the following characteristics is likely to fall into this category:

- information which relates to a project or subject that is live and ongoing
- information required to perform a current function
- information still in draft status
- information that has been accessed and/or edited within the past three months, and
- the most current version of information that is routinely updated on a scheduled basis (for example the most recent version of an asset register which is only updated every five years)

Addressing the following questions will help ensure that your information is managed appropriately whilst in active use:

- Have you defined the purpose(s) for which your information can be used?
- Can your information be easily located and accessed?
- Do you know who needs access to your information?
- Is access to your information controlled?
- Is your information safe?

Define The Purposes For Which Your Information Can Be Used

The 2nd principle of the [Data Protection Act](#) states:

“Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.”

Failure to adhere to this principle risks separating the content of the information in question with the management and governance rules which have been agreed as being suitable for it. When dealing with personal data this could well damage the interests of the data subject (an identifiable living individual who is the subject of personal data) and in turn the reputation and legal standing of the organisation responsible. Without clarity regarding why that information was obtained, what it is to be used for and by whom, it is all too easy for it to be inadvertently reused in a way in which the data subject may not be comfortable with or which damages their interests. For example, an individual who gave information in confidence as part of a research project suddenly finding themselves publicly associated with those views.

But it is not just in relation to personal data that such issues need to be considered. The almost boundless potential to reuse information has been one of the defining characteristics of information technology in recent years. In the vast majority of cases the outcomes of such reuse will be positive and are to be encouraged. This could be the embedding of a digital photograph within the university website, the aggregation of existing statistics to undertake fresh analysis or making use of the content from a previous project proposal as part of a new initiative.

What is important is for there to be active and deliberate consideration as to whether any of the management controls associated with the original information either prevents or limits its potential reuse, or needs to be transferred to the new resource. For example, do the intellectual property rights associated with the digital photograph allow for it to be reused on your website? Does it contain images of individuals whose consent might be required? Will the credit for the picture be correctly associated to the original photographer or mistakenly attributed to the person responsible for scanning and manipulating it?

In order to achieve this, it is necessary not only to have defined what management controls are required for the information being created for its original purpose, but also to identify whether there are likely to be any issues associated with its reuse in the future. If not, great; but if so it is better to consider these issues in advance - or at least to flag that there may be issues to be addressed if and when this occurs in the future.

The issues raised in this section help demonstrate the intrinsic link between information and the processes which both create and rely upon it which is a recurrent theme throughout this infoKit.

Locating And Accessing Information

Being able to locate and access the information being created within your organisation in as quick and easy a manner as possible is clearly a vital issue. Failure to achieve this risks user frustration, wasted resources and potential difficulties in complying with legal requirements. It would seem logical that such factors are most likely to surface during the *active use* phase of the lifecycle for the obvious reason that this is the phase during which information will be most regularly accessed. However, it is also important that a longer term perspective is taken which also considers how access requirements may change over time. This is because information which is no longer being actively used on a daily or weekly basis is less likely to be found through proximity (i.e. appearing in your *recent documents*) and intuition and will be more reliant on the quality and logic of the measures you have put in place to describe it.

Unfortunately information description is not an area where most users excel. Information creators often resent having to manually add metadata to their resources and this needs to be borne in mind when designing or implementing any information system that is heavily reliant on user-generated metadata for content retrieval. Even if the completion of metadata fields is made mandatory there is no guarantee of the quality of data being provided with abbreviations, spelling mistakes and format inconsistency (ie how dates are expressed) dramatically reducing the quality and usefulness of the metadata being gathered.

In an ideal world systems should be designed to automatically generate as much resource discovery metadata as possible, either from information already known to the system (e.g. the user's name from the system login) or from the particular part of the process they are performing. Allied to this it is preferable wherever possible for any additional metadata required to be 'fed' from other systems which contain the definitive version of such information (for example providing staff names direct from the HR database). As well as ensuring the accuracy of the metadata such cross-system integration should also help facilitate cross-system searching.

Should this level of system integration not be possible there are still measures which can be taken to improve the ease with which information can be located and accessed. The [Managing Information to Make Life Easier Guide](#) has some good practical advice which you may find of use in this regard.

At the very least when creating a new system or process it is important to consider who else needs to know of its existence within the organisation and how they can obtain access to the information it produces. Individual 'silos' of information whose contents are known to and accessible only by a single member of staff is seldom helpful. Not only does it severely limit the usefulness of that information but may also make it difficult to provide an accurate and complete answer to a Freedom of Information request or other legal/regulatory discovery exercise. Perhaps the most overlooked example of such a silo is the average user's email inbox whose (considerable) contents remain accessible only by the individual user.

Who Needs Access To The Information?

Knowing who needs access to information and on what basis is of course closely related to discussions regarding how to locate and access it; but here we are dealing with the who and the why rather than the how.

Information delivered at the right time, in the right format to the right people is likely to be of far more use and impact than content which needs to be tracked down and liberated every time it is needed. This requires an accurate knowledge of your business processes and a clear understanding of how your information system is integrated within it. All too often the act of creating information is divorced from the business activity(s) for which it is required. This quickly creates monolithic silos of information with no apparent relevance to the functioning of the organisation.

The alternative is also true. One of the most common complaints heard of modern office life is that people are struggling to cope with ever-increasing volumes of information. Simply exposing staff to yet more information of limited or no value to their role is likely to be a futile and wasteful endeavour. Moreover, information overload is one of the main contributors to a lack of appropriate information management. Email is a classic example of this where, due to the excessive volume of messages received, users struggle to deal appropriately with messages as and when they arrive. So instead of either acting on the email and deleting it or filing it within the appropriate corporate space, messages tend to accumulate unmanaged within the user's inbox.

Restricting Access

It is also necessary to control access to information for reasons other than the convenience of the user. It is also required to protect the intellectual assets of the organisation, the personal data of stakeholders and the interests of third parties. Access to commercially-sensitive information and that given in confidence must be carefully controlled, as must personal information which is covered by the seventh principle of the Data Protection Act which states:

“Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.”

In practice it can be difficult to determine exactly who does require access to information, especially as this need can often be only temporary to fulfil a specific task. Once again a clear understanding of the business processes for which the information is to be used will be instrumental in ensuring an appropriate level of access control is maintained.

The results of such analysis will be required when formulating the access controls of a file plan within a shared document storage facility. Careful consideration is required to ensure that an appropriate balance is struck between open access and security to ensure that only the right people have access to the right information.

Semi-Active Use

Information is often only in regular active use for a comparatively short space of time. In due course its level of use will decline, perhaps because the information has been superseded or simply because the initiative to which it relates has itself ended. But this does not mean that its use, or indeed its usefulness have disappeared altogether. The information may still have residual value and be referred to on occasion for reference purposes, or may no longer have any informational value but still be required due to its possible evidential importance.

In many respects this represents the most problematic and potentially dangerous phase of all in the information lifecycle. This is because when information is in almost continual use this in itself often dictates its own management: we keep it close at hand and know instinctively where to find it, we know who we are working with and therefore who needs access to it and we know it has value because of its relevance to the task in hand. But when we have moved on to the next task and that peak of use begins to decline things become less clear-cut - a problem compounded by the fact that our interest and efforts have now been transferred to tackling the next task in front of us.

This is often the period during which people forget where the information came from, what its place in the process was or indeed why it was created at all. Such uncertainty often leads to the accidental breakdown of whatever carefully constructed management controls were put in place during the creation and active use phases. This can often lead to quite different and contradictory patterns of behaviour depending on the nature of the individual user. Where one member of the project team may come across project files from a long completed project and delete them as obsolete, five others may find the same files and each retain copies 'just in case'.

The following sections are designed to try to help you to steer through this grey period in the lifecycle and impose some structure on what can otherwise become the least well defined and managed part of the model:

- Do you know what information is being held and why?
- Do you know how long your information must/should be kept for?
- Can you assess the value of the information to the organisation?
- What are the most appropriate and cost effective means of storing information?
- Is your information safe?

What Information Is Held And Why?

It is important that your information management measures extend to semi-active and inactive information as well as covering information being used on a regular, active basis. As well as still having potential value this information, often sizable in volume, is still the responsibility of the institution and could be a potential liability if not managed appropriately.

There is a balance to be struck between sensibly relegating older information down the priority list whilst simultaneously avoiding the dangers of adopting an 'out of sight out of mind' mentality. Clearly you do not want to force users to wade through masses of less current information in order to find the information they need to perform the active task in hand. But equally it is important to retain contextual links between current initiatives and what has gone before and for users to know that the knowledge and accumulated wisdom obtained through previous endeavours may be relevant to their current work and is available to be learned from.

Defining The Journey From Active, To Semi-Active Use

It can be useful to clearly define the event-driven stage points which govern when different types of information move from an active to semi-active phase. For example, for project records 'project closure' may represent this trigger, or for student records it may be one year after the student has graduated. Once defined, such points should mark the formal dividing point between phases of the lifecycle and the trigger for a review of the management controls governing it.

The first such consideration should obviously be whether the information is still likely to be required now that its immediate *raison d'être* is at an end. The specific factors to consider in this process will be addressed in the following sections, so for now it is sufficient to highlight the importance of weeding information at this stage to ensure that only information required for a defined reason is being kept.

This review point is also the opportunity to consider the other management controls currently in place for the information and whether they are still appropriate. For example, do the security and access controls need to be adjusted, could the information now be stored elsewhere in less expensive 'offline' facilities, does the format the information was created in need to be changed to ensure it will be accessible in the future?

Retaining The Context

It is also worth considering what links may need to be established between this older content and any current or future work to which it relates. Will the way in which users search for information routinely cover both active and semi-active material? If so, does there need to be any kind of indicator to flag the different status of the information found?

Continued Maintenance

Although the kind of events we have discussed as marking the transition from active to semi-active use are an important stage-post, it is also important to bear in mind the need for continued review and maintenance of information throughout its semi-active life. Circumstances and requirements change over time. The sensitivity of information may decrease (or potentially increase), the rationale for continuing to keep it may change, or measures need to be taken to ensure its longevity. As a result, regular review of information held throughout the semi-current stage of the lifecycle is recommended.

How Long Should Information Be Kept For?

In an age where storage costs are relatively cheap (at least for electronic information), it can often be tempting to assume that the best route is to keep everything. However, this approach does lead to significant costs, even if many of them - such as decreased efficiency thanks to the amount of time taken to find the information required - are often hidden.

There are also considerable risks associated with keeping everything. All content held by (or on behalf of) an institution is potentially disclosable under the Freedom Of Information Act or as part of any other form of legal discovery exercise. Many organisations have found themselves in damaging or embarrassing situations thanks to the disclosure of information they still held, but had forgotten all about...

By using a consistent and objective set of criteria for determining how long your information should be retained, not only do you decrease the chances of important information being mistakenly deleted too soon, but you are also in a position to defend your inability to produce information requested under FOI.

The Drivers Governing Information Retention

There are a number of factors which can influence how long your information should be retained for. In broad terms, they can be divided into two main camps: internal and external factors.

Internal factors will primarily be determined by operational considerations, for example how long the information is likely to be needed both to fulfil the purpose for which it was originally created, but also for any secondary purposes. However, it is also important to consider the longer term historical perspective and whether the information in question is likely to be of interest to future generations as part of the documentary record of the development of the institution. The [Guidance on the appraisal of archival records](#) provides further information on making this decision.

External factors will be largely governed by legal and regulatory requirements. Many pieces of legislation will have statutes of limitations stated within them which helps define the minimum amount of time information covered by that legislation should be kept to ensure any subsequent legal challenge can be resolved.

"Statute of Limitation: 'A statute of limitations is a statute in a common law legal system that sets forth the maximum period of time, after certain events, that legal proceedings based on those events may be initiated. In civil law systems, similar provisions are usually part of the civil code or

criminal code and are often known collectively as 'periods of prescription' or "prescriptive periods"⁴

Although sometimes less clear cut, it is often possible to determine similar periods implicit within sector-specific regulation which can also be used as a guide.

Risk Management

It is perhaps worth remembering that all decisions regarding the retention of information is based on the management of risk. Even where limitation periods are clearly defined this only represents the minimum amount of time the information should be retained for - there may still be other factors which mean a longer period may be justified. Conversely, a shorter period could be considered where the volume of information covered is high and the perceived likelihood of it being required very low.

Assessing The Value of Information To the Organisation

As we have seen in the previous section, there will sometimes be clear and compelling external reasons for retaining information. However, where these do not exist it can sometimes prove difficult to assess the value of information to the institution with any degree of objectivity. Without a reasonably accurate estimate of its worth, it can then prove difficult to make any informed risk-based decisions regarding retention. After all, all storage of information costs money whether it is the fees associated with commercial storage of paper records, or the considerable overheads associated with ensuring continued access to electronic data. It is therefore important that you are sure that the value of the information to the institution justifies the cost of continued retention.

Defining Value

Unfortunately it is not always easy to quantify the value of information. If you decide to dispose of a text book you know how much it will cost to buy a replacement, but this is seldom true of internal information as its value is rarely defined in strictly monetary terms. The time required to recreate the information might be one measure, but in itself this does not help to define the likelihood of this scenario which will itself be dependent on the perceived value of its contents to the institution, its staff and stakeholders.

Thankfully there are some methodologies which are beginning to emerge which appreciate the intangible nature of the value of information and provide ways of quantifying their worth. The [espida project](#) at Glasgow University represents one such approach which may well prove useful for larger scale initiatives in this area.

A more basic 'ready reckoner' for deciding whether information is still worthy of retention is to ask yourself the following questions:

- Does it contain useful information that I or others will need to perform a specific and known task or role?
- Have you or a colleague referred to this information in the last 6 months?
- Is this the only place where such information is available?
- Is it likely that an auditor would wish to see this information?
- Are there legal or regulatory reasons for keeping this information?
- Is it likely that future generations are likely to be interested in this information as a historical record?

⁴ Taken from the Wikipedia entry for Statute of Limitations on 20 July 2007
http://en.wikipedia.org/wiki/Statute_of_limitations

If the answer to any of the above is 'yes' this may indicate that it is information that is worth keeping - at least for the time being. If the answer to all the above is 'no' it is unlikely that it is required and consideration should be given to removing it. Of course this can only ever be a rough guide and should be applied within the context of a risk management-based approach.

Appropriate And Cost-Effective Means Of Storing Information

As with all aspects included within the information lifecycle model it is important that the storage of information is fit for purpose. When in active use the prime determinant of fitness for purpose is ease of access: people want to be able to get hold of the information they require as quickly and easily as possible. However, when the frequency of use begins to decline so other factors also need to be considered.

Access will inevitably continue to be a factor - after all, there is seldom much point in continuing to store information if nobody is aware of its existence or knows how to find it. However, this consideration will inevitably be tempered by cost considerations. Near-line storage costs money, whether it is the costs associated with keeping paper records in prime office space, or delays caused by people sifting through unnecessary information to find the content they urgently require. This raises the question of what the alternatives may be and when it is appropriate to consider them.

Storage Options

For physical information, it may be worth considering moving semi-active content into off-line storage facilities which are usually located in cheaper, lower-grade accommodation and able to take advantage of economies of scale. Alternatively commercial storage providers can take away the need for onsite storage altogether - though at a cost and at a trade off in terms of speed and ease of access.

Crossing the boundary into the semi-current phase of the lifecycle might also mark the time to consider changing the format of information. Scanning can be an efficient way of reducing the physical footprint of your information but is not the quick and foolproof solution it can appear. Scanning is an expensive process and you want to be sure you are only scanning information that deserves the expenditure. There are also significant questions to be resolved regarding the management of the scanned copies and the longevity of the electronic, scanned versions.

For information that is born digital decisions regarding appropriate storage are less pressing, but do still exist. Some institutions employ email archives which automatically capture email and store emails within user accounts which are over a certain age. In such situations it is obviously important to consider carefully what this age should be and what the implications may be for information retrieval. Alternatively it is not unusual for users to create their own annual .PST files containing all of the emails they have sent or received during the year as a personal reference archive. It would be wise to consider the relative merits and potential risks involved in users adopting this approach and deciding whether to condone or condemn this practice.

Out Of Sight Shouldn't Mean Out Of Mind

It is worth remembering that the institution is responsible, and potentially liable, for all the information it holds. It is therefore important that wherever information is being stored, appropriate management controls are extended to it. Moulding boxes and unlabelled back-up tapes stored in a broom cupboard are of no use to anyone, but are a potential risk to everyone.

Is Your Information Safe?

The seventh principle of the Data Protection Act states:

“Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.”

This should act as a reminder that not only is it in the obvious interests of the institution to appropriately protect its information assets, but also potentially to a much wider set of stakeholders whose rights in this regard are enshrined in law. Clearly this is a pertinent issue throughout all phases of the information lifecycle, but is often perhaps most pressing an issue when the immediate use of the information diminishes and interest in it wanes. It is during this phase, as with so many other elements, that agreed management controls can inadvertently slip and good practice be forgotten.

The Overlooked Importance Of 'Good Housekeeping'

It is often the simple things which are overlooked but which can make the biggest difference. Although few of us ever manage to leave for home each night having left a completely 'clear desk' behind, users should be encouraged to at least ensure that confidential or sensitive material is locked away when not in use. Likewise, ensuring password protected screen savers are used, especially on PCs kept in open environments, is another simple yet effective preventative measure. Even such simple steps as ensuring that monitors being used in the processing of sensitive data cannot be read through a window or passing corridor can help.

Users should also be given guidance on what constitutes a good password (i.e. one which avoids proper names and includes numbers and symbols). When designing systems, consider whether to enforce the regular changing of user passwords - but beware that one unintended side affect of doing this might well be to encourage users to write their new password in obvious places rather than commit it to memory!

External Service Providers

There is an increasing trend for institutions and individual users to rely on online services provided by external companies to create and store information. This could be anything ranging from so called 'social software' services providing wikis, blogs and other online systems, to external hosts of online email and office applications.

It is worth giving careful consideration to the nature of the content users are entrusting to these services, and the guarantees you have regarding continued service delivery. Many of these companies are small with products in a perpetual state of beta development. It is inevitable in a competitive commercial environment that some services will be withdrawn and companies dissolved, potentially with little or no notice.

As such it would be unwise to rely on such services as the sole means of storing information considered to be of any real significance or value.

Final Outcome

The final outcome phase is, as the name suggests, the last phase of the lifecycle. As such it covers two different and diametrically opposed processes which together encapsulate the fate of all information: either its long term preservation or its deletion.

The whole lifecycle concept is founded on the importance of taking a comprehensive and proactive view about information management which seeks to look ahead, as well as dealing with the immediate challenges posed within any given phase. The issues surrounding the preservation of digital information illustrate the importance of this concept perhaps more than any other area due to the speed of technical change and the immense, sometimes unrecoverable, problems which can arise through attempting to deal with them retrospectively.

It is for this reason that there have been specific references to the challenges presented by digital preservation throughout the earlier sections of the lifecycle. However, it is in this last, final

outcome phase that we will explicitly address some of the main threats to ensuring access to electronic information in the longer term and the measures which can be taken to address these risks.

The two topics included in this phase cover both of the outcomes possible during the final outcome phase and include:

- How will you ensure continued access to the information?
- How is the disposal of information controlled?

Ensuring Continued Access To Information

Unlike paper, information stored in electronic media is inherently unstable and vulnerable to a wide range of risks, any one of which can threaten access to the content unless appropriate preventative measures are taken.

The major threats to the long term accessibility of digital information include hardware obsolescence, the incompatibility of software versions and media decay. Though it is impossible to define a specific time-span after which preservation issues are likely to become a pressing issue, it may be useful to consider 5 years as a general rule of thumb. For example, information that you know will only be required for less than 5 years is less likely to be affected by some of the preservation threats outlined above. Conversely, if you know the information being created will be of value for longer than 5 years, it is more important that these issues are at least considered. The table below is based on the assumption that the information in question will be required for longer than 5 years.

Although appropriate actions to counter these threats are likely to require large scale institution-wide measures, there are steps which can be taken whenever a new information system is proposed to help meet these challenges.

Measure	Purpose
Avoid proprietary data formats & systems	If the information being created is expected to be required for longer than 5 years careful thought should be given as to whether it is appropriate to rely on proprietary formats which may 'lock in' your data. This can make it difficult to perform future preservation actions and increase your reliance on external agencies over which you have no control.
Avoid storage on removable media	Removable storage media tend to age quickly, usually being replaced within a matter of years by new media with greater capacity, thus increasing the risk of hardware obsolescence. Removable media are also more likely to be stored inappropriately increasing the likelihood of loss or damage.
Consider when to upgrade software	Whilst it is not necessary to upgrade when every new software version is released, it is important to be aware of when your particular version will no longer be supported or will no longer be readable by subsequent versions and to take action accordingly.

Consider existing information when introducing technical changes	It is easy for decisions to be made on purely technical grounds without due consideration of its impact on existing information. For example, a decision to replace all PCs without an appreciation of the fact that a large volume of information still exists on floppy disks, the drives for which are not present on the new machines.
Consider the physical care of electronic storage media	Electronic storage media are often fragile and sensitive to fluctuations in temperature and humidity or to magnetic fields. Consideration should be given to where and how such media are being stored and they should be checked periodically to detect the first signs of deterioration before data is lost.

Controlling The Disposal of Information

Disposing of information may seem as straightforward a process as hitting the 'delete' key or finding the nearest waste paper bin. Unfortunately it may not always be a simple as that. It can prove more difficult than might be imagined to irrevocably remove electronic information to the degree required by the law. It is also a process which should now be as controlled and auditable as every other aspect of information management in order to protect the interests of the institution.

When Has Deleted Information Really Gone?

According to [guidance from the Information Commissioner](#) accompanying the FOI Act,

“Information located in desktop recycle bins is clearly subject to the FOIA as this continues to be held and is easily accessible. Once deleted from the recycle bin the information will also continue to be held unless the electronic record is completely erased from the computer system. Information in a deleted file or in a back-up, whether a server, disc or tape, may be regarded as being held by a public authority for the purposes of the FOIA depending on the particular circumstances of the individual case.”

It is therefore important that your deletion procedures are comprehensive enough to ensure that information you rightfully and lawfully wish to remove from your possession has actually been purged to the satisfaction of the above criteria.

It is possible to draw analogous conclusions for paper records - highlighting the importance of regular emptying of waste bins (particularly recycling bins) and the wisdom of providing confidential waste facilities and/or shredders where required.

An Auditable Process

To ensure appropriate levels of transparency and accountability it is considered good practice to document the disposal process and its outcomes. For example, to record what information has been destroyed, by what criteria it has been assessed as requiring destruction, on whose authority this has been carried out and to confirm the outcomes of the process.

Clearly any measures introduced in this regard should be proportionate and will require an analysis of risk. Obviously there is no need for the deletion of every email to be documented to this degree, but it may be wise to introduce a general policy statement which defines the types of emails which users can routinely destroy (spam, ephemera etc) and which should be subject to formal retention and appraisal procedures based on the significance of their content.

Disclaimer

We aim to provide accurate and current information on this website. However, we accept no liability for errors or omissions, or for loss or damage arising from using this information.

The statements made and views expressed in publications are those of the authors and do not represent in any way the views of the Service.

The JISC infoNet Service offers general guidance only on issues relevant to the planning and implementation of information systems. Such guidance does not constitute definitive or legal advice and should not be regarded as a substitute therefor. The JISC infoNet Service does not accept any liability for any loss suffered by persons who consult the Service whether or not such loss is suffered directly or indirectly as a result of reliance placed on guidance given by the Service.

The reader is reminded that changes may have taken place since issue, particularly in rapidly changing areas such as internet addressing, and consequently URLs and email addresses should be used with caution. We are not responsible for the content of other websites linked to this site.

This material is licensed under the [Creative Commons License](#) - 2007