

## Data Protection Basics

“You don’t know what you’ve got  
(until you lose it)” Ral Donner



A man in a dark suit and light shirt is smiling and leaning against a large, dark, textured backup device. The device has 'FUJITSU' at the top and 'ETERNUS CS' on the side. A speech bubble is positioned near the man, and a larger text box is below it.

FUJITSU

How can I achieve the most reliable data protection with a maximum of cost efficiency?

With ETERNUS CS. Fujitsu's virtual tape solution for the entire enterprise backup with intelligent data protection on disk and tape.

- Virtual tape support for mainframes and open systems
- Scalable for each class of data center environment
- High availability and disaster resiliency for protection against site outages
- Managed storage and managed backup

ETERNUS CS

FUJITSU

# **Data Protection Basics**

**“You don’t know what you’ve got  
(until you lose it)”** Ral Donner

**Terms of Use**

This document is Copyright 2009 Freeform Dynamics Ltd. It may be freely duplicated and distributed in its entirety on an individual one to one basis, either electronically or in hard copy form. It may not, however, be disassembled or modified in any way as part of the duplication process. Hosting of the document for download on the Web and/or mass distribution of the document by any means is prohibited unless express permission is obtained from Freeform Dynamics Ltd.

This document is provided for your general information and use only. Neither Freeform Dynamics Ltd nor any third parties provide any warranty or guarantee as to the suitability of the information provided within it for any particular purpose.

# Contents

<b>1. Data protection: Why aren't we there yet?</b>	<b>5</b>
<b>2. Why should you be thinking about data protection?</b>	<b>9</b>
The strategic value of information	9
Legal and compliance aspects of data protection	11
What are the risks?	12
What can I do about the risks?	14
Meanwhile, back in the real world...	15
<b>3. Mapping out the requirements</b>	<b>18</b>
The value of data	18
The business view	19
The user view	20
The regulatory view	21
The infrastructure view	23
Bringing it all together	24
Introducing 'RTO and RPO'	25
Common misconception #1: Backup is the same as archiving	28
Common misconception #2: High availability (HA) is the same as data protection	29

<b>4. Data protection technology</b>	<b>30</b>
Backup media	30
Methods and mechanisms	31
From technology to data protection	34
Delivery options	40
<b>5. Making the right data protection choices</b>	<b>42</b>
Scenario 1: Responding to inadequate data backup and restore	44
Scenario 2: Business continuity and disaster recovery	46
Scenario 3: Data protection improvements for compliance reasons	47
Scenario 4: Branch office/remote office data protection	48
When to use disk or tape?	50
<b>6. Taking things forward with data protection</b>	<b>52</b>
Getting the business case right	52
Gaining buy-in and support from relevant stakeholders	53
Evaluating and testing potential solutions	53
Defining appropriate policies and operational terms of reference	55
<b>7. Tips and tricks</b>	<b>56</b>
<b>References</b>	<b>59</b>

# 1. Data protection: Why aren't we there yet?

Some things in life, such as taxes and the music of the younger generation, have been around forever – and yet we are still not comfortable with them. The same can be said of certain areas of business and IT, with the need to protect an organisation's vital data using a variety of technical mechanisms – backup and recovery, archiving, and so on – springing readily to mind.



It is worth saying up-front (legal requirements aside), that the only valid reason for carrying out backup and related data protection operations at all is to be able to recover business information when it is lost.

There are plenty of ways to lose data, for example, following theft, fire or flood, through human error, computer viruses, hardware failures or security breaches. However, although IT staff have been dealing with data protection challenges for decades, research suggests that only a proportion – perhaps no more than a third – of corporate data is being backed up according to formal policies.

Certainly, nobody we have spoken to in the course of our research activities has seen data protection as unimportant, so why is this area not a 'done deal'? At the risk of stating the obvious, a significant factor is that protecting data can be difficult. Not least because of the well-known double whammy: with data volumes growing rapidly and business operations requiring access to information during ever-increasing hours, doing something as seemingly simple as taking backups quickly enough can be an extremely non-trivial challenge.

Also, complexities lie in the fragmented manner in which information is stored and in how it is changed. And although there may have been some significant advances in the mechanisms available for data protection, it can be difficult to get a clear idea of what approaches fit which requirements. Furthermore, operations staff tasked with keeping systems working in line with business demands rarely have

enough time to learn about new solutions, never mind to work out how things can be made better.

Today, the data protection capabilities in many organisations operate the way they do because ‘that’s the way they have always worked’. Frequently, backups are taken without the supporting systems having been tested or indeed, any real knowledge of whether the resulting copies can be recovered and used. Alas, in many businesses the real value of recovery systems is only appreciated when they are called upon in anger, or indeed grief. By which point, of course, it is too late to fix anything that is not working.

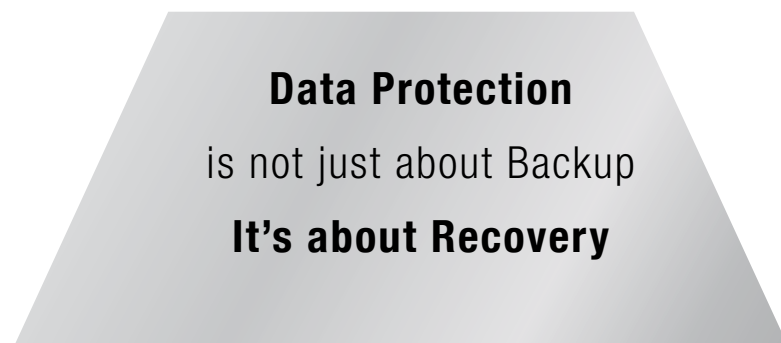


FIGURE 1: A gentle reminder

It is interesting to note that where organisations do look at data protection, attention is often focussed on the protection/backup process itself. The recovery process – and its speed – which may in fact be more critical, can be overlooked.

Many approaches and technologies are available to help businesses deal with data protection, including backup and recovery, archiving, replication and disaster recovery. With technological maturity comes affordability and workability, so now is as good a time as any to look with fresh eyes at this broad and deep area.



And so we come to the purpose of this guide; to help IT decision makers in all sizes of company make the right data protection choices for their own business needs. We have included the following sections to provide the means with which to do this:

**Why should you be thinking about data protection?** This covers the rationale behind protecting data – with the starting premise that not all data is created equal, which means (sadly) there can't be a one-size-fits-all approach.

**Mapping out the requirements.** Here we consider the elements which should form a data protection strategy. We present a number of requirements on protecting data, from business and user drivers, to compliance and infrastructure needs.

**Data protection technology.** This section covers the 'data protection tool chest' in a jargon-free manner. In a (very) brief history of storage, we review when and where it was first realised that information had to be in some way recovered – and we provide an overview of the technology which could be used to protect and recover data.

**Making the right data protection choices.** Here we offer pragmatic guidance on meeting data protection needs with the capabilities provided by current technology. Looking at a number of scenarios, we consider how to work with what is in place in your business already, all the while assuring future safety.

**Taking things forward with data protection.** Here we explore 'getting from where you are to where you want to be.' We look at how to achieve the benefits of data protection by working through the necessary steps of definition, deployment and operational management.

**Tips and tricks.** Finally, this section presents some nuggets of advice gained from the experiences of your peers around how to deliver efficient, effective data protection. By the end of this section you should have a clear understanding of the best way forward for your business.

Before we continue, we should remember that data protection is not an absolute science; rather, it is a balancing act that requires a hefty dose of pragmatism, which we have aimed to reflect in this guide. Whether you are looking to refresh existing facilities or adopt an evolutionary journey to better data protection, this guide is for you.

## 2. Why should you be thinking about data protection?

When Steven R. Covey was writing ‘The Seven Habits of Highly Effective People’, he reviewed as many self-help and management best practice books he could get his hands on, distilling down their lessons and guidance into a single volume. The timeless advice that resulted highlights two ‘habits’ above all else: namely, ‘Starting with the end in mind’ and ‘Knowing how to get there’.

The protection of business data is not an end in itself – there has to be a reason why people want to do it. That’s the theory anyway. In practice, organisations often have less of an idea why they are doing it, but do it anyway, with the inevitable result that some data is protected because it is straightforward to do so, rather than whether it is actually necessary.



So, why should you be protecting your data? Fundamentally, for two reasons:

- Because you want to, as the data has some intrinsic value to you or your business.
- Because you have to, as you will get into trouble if you don't.

### **The strategic value of information**

Few organisations today would deny their absolute dependency on both their computer systems and on the information they store. From a business perspective the ability to exploit information in a timely fashion has become critical. It is worth having a think at this point about what might be the impact on your business, should any of the following suddenly become unavailable:

- Customer and supplier information
- Data about business events, e.g. logistical information ('What's on what truck, and where?')
- Safety-critical information, e.g. relating to the status of a hazardous process, or material
- Research and analysis data
- Email and collaboration information
- Office documents, spreadsheets, plans and files
- Video and audio feeds.

Every business has its own needs – a news organisation might be able to cope without 'customers' for an hour or so, for example, whereas a bank would immediately start losing money if it wasn't able to trade due to a data loss. On unfortunate (but thankfully rare) occasions, organisations do go out of business due to their inability to manage information correctly.

On the upside, we can always do more with information. It is being created all the time, on servers in the data centre and at remote locations, on corporate laptops and netbooks, PDAs and smart phones. There is also a growing trend for non-corporate PCs and laptops being used for business-related work, frequently by staff working from home outside of office hours. Harnessing the power of information is a broad topic in itself and, while outside the scope of this guide, it is nonetheless critically important that data is available, accessible, protected and recoverable if organisations are going to make the most of it.

Ultimately, all this information has to end up somewhere so it should come as no surprise that the demand for storage does not seem to be

letting up. According to the Internet Innovation Alliance, “it took two centuries to fill the shelves of the Library of Congress with more than 57 million manuscripts, 29 million books and periodicals, 12 million photographs, and more. Now, the world generates an equivalent amount of digital information nearly 100 times each day.”

Indeed, the volume of information we are creating today is quite staggering and, given how easy this has become, it can be difficult to decide what is valuable and what is not. Gigabytes of raw video captured during a meeting may contain a single nugget of real value for example, or a simple spreadsheet may hold the master price list for an entire company.

Does all this sound familiar? Is it any wonder then that, in one study, 40% of organisations we researched said they had a ‘keep everything’ strategy for retaining data? However, there is a whole raft of reasons why we can’t just keep all of it. For a start, the discovery, management and, indeed, protection of such information can quickly become a nightmare, as ‘keep everything’ ends up meaning ‘manage and protect everything’. Second of all, it’s probably illegal. We look at this below.

### **Legal and compliance aspects of data protection**

It’s taken a few pretty high profile court cases in various countries around the world to highlight just what organisations can and can’t do with the data they store. For example, we have the corporate scandals in the USA earlier this millennium involving a raft of organisations including Enron, WorldCom Tyco, Adelphia Communications, Xerox and Global Crossing, as well as various auditing firms, which hurried the creation of regulatory instruments such as Sarbanes Oxley. No doubt we shall see a number of new regulations created as a result of the credit crunch which surfaced in 2008. Meanwhile, in many countries around the world, an increasingly comprehensive set of directives exists around data protection itself, catalysed by the media storms around data leakage. The incidents in the same year, such as 21 million banking customer records and a further 17 million records

from a mobile phone operator appearing on the German black market, or the UK's Revenue and Customs (HMRC) leaks, are cases in point.



Thinking about how such regulations apply to data protection, your business may be obliged to do any of the following:

- Keep certain information (reliably and accessibly) for a certain length of time
- Destroy certain information (including all backup copies) after a certain length of time
- Find specific information on request.

It is not difficult to see how these requirements might have an impact on what data you need to protect, and how you go about it. However, while this may be simple to grasp in principle, in practice it is a minefield trying to balance the needs of industry specific regulations with those dictated by local/national governing bodies and those which apply across international boundaries. And of course, all these are influenced by where your organisation does business and with whom.

We will look at legalities of data protection in more detail, but for now, what issues do we face?

### **What are the risks?**

We have acknowledged the plentiful reasons why we might want to keep certain information. Which would be all very well still, if we were living in a world in which nothing ever went wrong: perhaps we could just salt it all away in some massive electronic warehouse. Unfortunately however, we don't live in a perfect world. A number of bad things can and sometimes do happen to our information assets.

Figure 2 shows our research findings<sup>[1]</sup> with respect to information-related risks and corporate sensitivity to them in a variety of sectors.

As you can see, ‘general data loss’ is seen as the biggest risk across all industries, closely followed by ‘systems failure’. Sadly, computer systems, the software that runs on them and indeed the humans that interact with them, do not always function as we might like.

Overall Ranking	Type of Business Risk	Public Sector	Financial Services	Comms Sector	Oil & Gas	General Industry
1	Loss of business critical information	Extreme	Extreme	Extreme	High	High
2	Downtime of key IT systems	Extreme	Extreme	Extreme	High	High
3	Illicit use of confidential information	Extreme	Extreme	Extreme	High	High
4	Legal exposure (inadvertent or otherwise)	High	High	High	High	High
5	Breach of building security (e.g. break-ins)	Extreme	High	High	High	High
6	Regulatory exposure (inadvertent or otherwise)	High	Extreme	High	High	High
7	Criminal activity (e.g. fraud)	High	High	High	High	High
8	Malicious damage by disgruntled employees	High	High	High	High	High
9	Accidental damage (e.g. fire)	High	Significant	Significant	High	High
10	Political instability	High	High	Significant	Significant	Significant
11	Natural disaster	Significant	Low	Low	Significant	Low
12	Public health emergency	High	Low	Low	Low	Low
13	Poor performance of financial markets	Low	High	Significant	Significant	Low
14	Terrorist activity	Significant	Low	Low	Significant	Low

FIGURE 2: Types of business risk and industry sensitivity to them

The consequences of data loss are why risk managers distinguish between ‘probability’ and ‘impact’ when it comes to risk assessment. Loss of business critical information and downtime of IT systems may be the most likely but might not be as damaging as a building fire, for example. Such things as criminal activity and accidental damage may appear further down the list, but when they do happen their consequences can be far reaching. In one incident, a Brazilian bank lost both its IT systems and backup tapes to flood damage. Fortunately, a specialist data recovery company was able to recover the majority of the mission-critical data by manually cleaning the waterlogged tapes in a special chemical solution.

So, just as ‘keep everything’ is not a practical strategy, neither is ‘avoid all risks’. By measuring impact as well as probability it becomes possible to take a more sanguine approach to risk prioritisation and management. Where both are high in relation to data, then it becomes significantly more important to mitigate the risks using some kind of data protection mechanism.

### **What can I do about the risks?**

It makes sense to try to deal with some of these risks head on, for example by installing more reliable systems, ensuring users are correctly trained, or building data centres and back-up sites away from flood-prone areas. However, it is when such pre-emptive efforts fail or are simply not practical that data protection mechanisms really show their worth. The most important elements of data protection in all its forms are encapsulated in a few simple statements:

**‘Take a copy’.** Wow! Is it that easy? In truth, for organisations and individuals alike, much of data protection really can be as simple as that: whatever happens to the data, at least you will know that you can get it back. Taking a copy, and storing it somewhere safe, is the principle at the heart of many data protection mechanisms, including the backup, snapshot, replication and continuous data protection (CDP) solutions that we shall be looking at later.



Not all the risks are about loss or damage however, which brings us to a second aspect of data protection:

**‘Prevent unauthorised access’.** A number of mechanisms, from encryption to authentication, have the same goal: stop the wrong people accessing our information.

Finally, we need to think about the third element we require from data protection, which, initially, may seem counter-intuitive. However there may be times when we need to prove we can do this:

**‘Assure destruction’.** There we have it – if only we could take copies of required data when necessary and store them somewhere safe, ensure that individuals can access only the information they are allowed, and provide a guarantee that data will, and can, be deleted when we no longer require it (or are no longer permitted to keep it), then our data protection requirements will be met.

If only it were as simple as this!

### **Meanwhile, back in the real world...**

A number of challenges can easily get in the way of implementing the right data protection mechanisms. Not least is the fragmentation of data itself. What do we mean by this? Data is now routinely held in many locations, by no means all of them sitting centrally inside a data centre. So even finding and identifying the data that needs protecting, deciding its value to the business and agreeing the right level of protection it needs is a task that can be extremely difficult to undertake.

This is especially true when we remember that such activities can never be ‘finished’ because things change so rapidly. Business, and information technology as a whole does not stand still, and this affects the kinds of data we need to protect and the mechanisms available to protect it. As examples of this constant change, let’s consider the impact of recent high profile movements in the technology and business services space:

**Server virtualisation.** This is the ability to run multiple ‘virtual machines’ on a single, physical server. While this has a number of benefits, it can potentially create a number of new workloads that need to be backed up, particularly if provisioning is uncontrolled. In addition, the virtual machines themselves may need to be protected, which can be a challenge – not least in capturing the running state of a virtual machine.

**Software as a service (SaaS).** There is growing interest in the use of various SaaS offerings by business groups, if not by the organisation as a whole. SaaS is an application delivery approach where the application is hosted by a third party and delivered over the Internet (for example Salesforce.com). While most SaaS providers offer services to backup the data they store, they might not automatically comply with customers’ specific requirements so it is essential to ascertain exactly what data protection services are on offer and what service levels are guaranteed for data recovery operations. In addition, the storage of certain data in countries outside the ones the customers operate in might be in breach of local regulations.

It should also be noted that data protection may itself be offered via this ‘as-a-service’ model. Later in the book we shall look at the potential use of data protection and recovery as a service over the Internet as well as the possible deployment of managed services to look after the daily operations of data protection and recovery services.

Such complications can result in spending substantial time and effort to understand just what an organisation’s data protection obligations, and therefore requirements, are. Ideally, all data protection conversations should start with the business itself, but this can also be problematic, particularly if the business finds it difficult to say what it wants. An important factor is that data protection is, quite frankly, seen as a ‘boring topic’. This has real (negative) consequences when it comes to establishing business needs for data protection, and when seeking funds to put the right mechanisms in place.

Establishing a solid approach to data protection for your organisation is likely to be an uphill struggle unless there is a specific commitment from the very top, or a sufficiently compelling reason to sort things out.

To start things off in the right direction, we first need to consider the pressures on the organisation that drive the need for data to be protected, and come up with a workable, achievable set of requirements that consider the risks to, and the value of the data to be protected.



The important point here is to recognise that different data sets will need different levels of backup and recovery protection.

Let's take a look at how we might go about assessing requirements.

## 3. Mapping out the requirements

To establish the requirements for data protection, you will need to draw on information from a number of different sources, each of which will have its own priorities when it comes to the data types and how they should be protected. These include:

- The value of data to your business and how it changes over time
- The business view: considering drivers from the top down
- The user view: meeting the needs of individual data ‘customers’
- The regulatory view: incorporating legal and compliance aspects
- The infrastructure view: ensuring that mechanisms can work alongside existing IT.

There will undoubtedly be some conflicts between the different perspectives, so the question then becomes “how do we bring all of these together?” At the end of this section you should at least know where to look to establish the right data protection requirements, and how to integrate the different views to establish a single set of appropriate policies.

### **The value of data**

The value of data to the business changes significantly over time. During the first few weeks of its life, it may be accessed and changed regularly. Thereafter, it generally enters a more settled period, where it may be accessed or changed less often. Further along, the number of times the data is accessed or changed at all may drop to practically zero, depending on whether it is used for historical analysis.

The rate of change of the data is a good indication of its value to the business at a given point in time. However, age and change are not the only metrics which define business value, there is of course the data itself! An easy way to think about all these things at once is in terms of a ‘data lifecycle model’. One of the important things a lifecycle view can

help us appreciate is that data will have different data protection and recovery requirements at different points throughout its lifecycle.

**The business view**

A ‘top-down’ view of data protection combines the need to balance the risks to data with the requirement to grow and evolve the business. Looking at risks first, we know from our research<sup>[2]</sup> that organisations are most concerned with legal exposure, compliance, customer trust and the commercial impact of all three (figure 3).

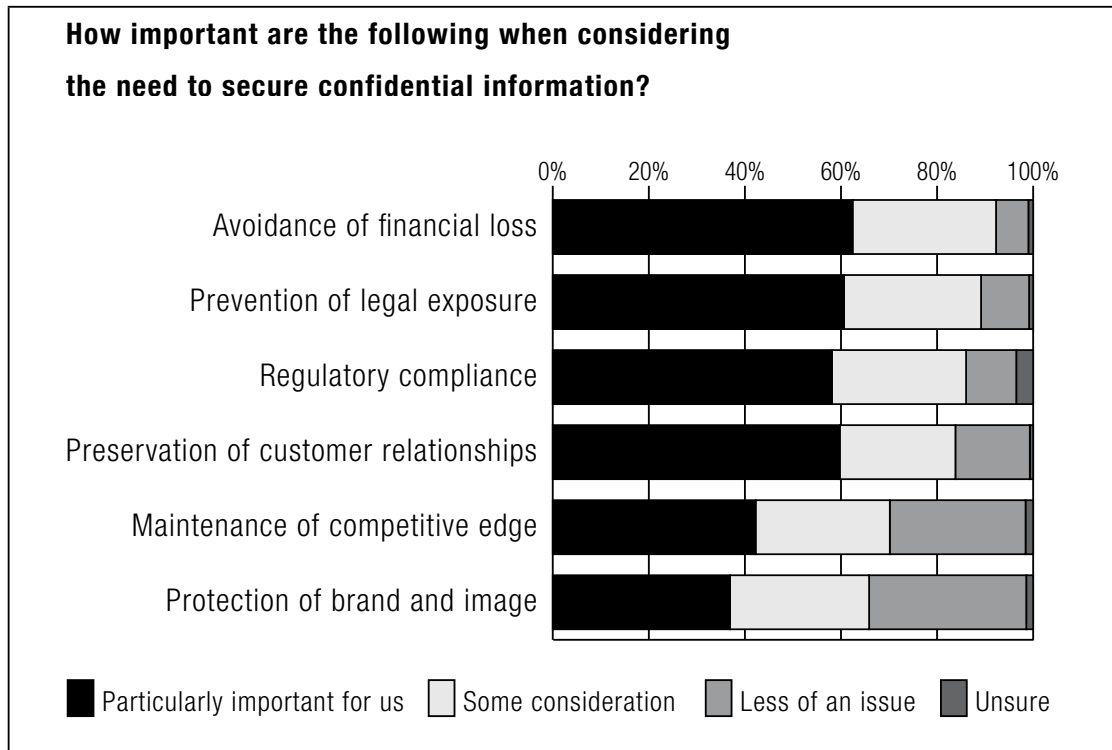


FIGURE 3: Considerations when securing confidential information

Ultimately, the bottom line for commercial organisations is around money and brand reputation, but even the latter has a financial motivation. Since 2003, most US States, and Japan, have obliged businesses to inform customers when they have mislaid or leaked certain information. The EU has also made clear statements of

its intention to put this law in place by 2012. There are significant negative implications for failing to protect customer data. The embarrassment factor is an obvious disadvantage and the possible loss of business due to losing the confidence of customers could be critical. For public sector organisations, things are measured more in terms of service delivery, so the financial impact can be thought of in terms of cost-of-service rather than profit. But the same point stands.

### The user view

As we all know, whenever users, or 'customers' as they must now be called, are asked to specify how important their data or applications are, they are certain to respond in positive terms. In the eyes of the customer their own data is of paramount importance, and needs to be easily accessible over its entire lifetime, which may well run into years. Customers will also request the best possible 'protection' in terms of backup points and recovery speeds.

Common sense dictates that these expectations can only be justified and met for certain types of information.



A recent study by the University of California at Santa Cruz showed that 90% of data stored on the network was never accessed again, and another 7% of the data was accessed only once.

While keeping all data at hand on high speed disks might seem ideal, in reality to do so could be prohibitively expensive in terms of acquiring and operating the hardware and finding the physical space in which to house it.

So, what we need to think about is how different data has different protection requirements. For example – data that changes very slowly, or not at all, doesn't need backing up as often as data which changes rapidly. To get to grips with all this, we need to have tools and processes in place to provide the means of assessing data protection and recovery requirements.

The place to start is first to 'discover' just what data is out there in the business and then 'categorise' it in terms of its importance and by the policies that should protect it. Metrics that can be used to assess protection and recovery requirements may include:

- Direct access frequency by the user(s) concerned/date of last access
- Frequency of indirect use of historical data, e.g. for analytics, or management reporting.

While this may sound simple, it requires time and communication with the users concerned. But, with such metrics to hand, it then becomes possible to define policies that describe just how and on which platforms different classes of data should be held, how they should be protected and the most appropriate recovery capabilities. Again, thinking in terms of a 'data lifecycle' may help but put this in context: the use and value of data changes over time, and so data protection and recovery requirements change with them.

### **The regulatory view**

Given that legal exposure is a significant consideration when it comes to data protection from a business standpoint, how do you make sure you understand your legal obligations? The starting point is to understand the jurisdictions in which you operate. This can be easier than it sounds, given the diverse nature of many modern organisations. To make things a little simpler, you need only to consider where data is being held or processed, which could be influenced by:

- The locations of your offices
- Use of outsourcing and hosting of data processing (including Software as a Service)
- Off-shoring of customer service, development and administrative activities.

From a regulatory perspective, you need to be concerned with a number of requirements, based on the information you hold about your customers, the industry you are working in, the organisations you work with (particularly governments) and the individuals you employ. For example:

**Data protection laws.** These tend to revolve around the information you hold about other people. EU countries are all bound by Directive 95/46/EC (which local laws then build upon) and, where such a framework does not exist, data protection elements will always be enshrined in country law. Data protection law tends to concentrate on the principle of ‘fair and lawful processing’ of data. Words such as ‘fair’ are open to interpretation of course. In this context it boils down to having a reason to keep the information you hold about an individual, and ensuring that it is handled with due care, which, of course brings us to the mechanisms we use to protect data. Such laws also guard against keeping ‘excessive’ data, which can translate into an obligation to destroy all copies of information when it is no longer necessary to keep them.

**Industry regulation.** Individual industries, notably the financial sector, healthcare and utilities, have their own laws and regulatory frameworks within which participating businesses need to operate. These laws can be in apparent conflict with data protection regulation, as well as with each other – some healthcare regulations require information to be kept for the lifetime of a patient for example, which may appear excessive (and perhaps expensive in storage!); meanwhile, the Payment Card Industry regulations in the financial services sector, stipulate “no longer than is absolutely necessary” for card-related information. At the extreme, building regulations may require plans to be kept for the lifetime of a building, which could run into hundreds of years.

**Government secrets.** Unsurprisingly, most governments won’t want you to work for them unless you sign confidentiality agreements when performing work. The UK’s Official Secrets Act for example, applies to



individuals and subcontracting organisations for many years after work is complete – 50 years and above in some cases.

**Employee laws and agreements.** While data protection law applies to the people you employ, equal rights legislation and worker agreements also come into play. In Human Resources some personnel records need to be retained up to the age of 75, for example. Meanwhile, individual countries and companies can reach agreements with employees that will be legally binding. In Germany, for example, it is illegal to delete email without an employee's consent, even if it is unsolicited (i.e. spam). Clearly, this can add additional burden to storage and backup requirements.

**Regulatory bodies.** As regulations tend to be aimed at business and not IT, your organisation's lawyers are the best place to start helping you understand what you should be retaining (or not). Speaking of lawyers, it is worth introducing the concept of eDiscovery – that is, collating information to respond to a legal information request. eDiscovery is a term that's front of mind for many US organisations (for which litigation is a frequent issue) and is also a growing concern in Europe, where data protection laws add complexity to legal requests for access to data, given that member states will have implemented the overarching EU data protection directive in slightly different ways. Several examples of major corporations getting into hot water in this area over the last few years include cases involving Intel/AMD, Qualcomm/Broadcom and, as far back as 2005, Europe's (then) biggest bank, UBS AG, was forced to pay an ex-employee \$29m in a discrimination case, part of which hinged on UBS breaking its own data retention policies and deleting and destroying email evidence.

### **The infrastructure view**

As we know only too well, when it comes to building any IT system it is very, very rarely that one has the luxury to start with a completely clean sheet of paper. In nearly every case some components of the existing IT infrastructure will need to be considered as integral to a new system. In other words, when looking at implementing or improving your data

protection and recovery capabilities, the state and capability of the existing infrastructure needs to be the first port of call.

First and foremost we need to understand where data is held, both physically and logically (in terms of what part of what application/database for example) and on which type of hardware. Information is likely to be stored on a wide range of hardware platforms, on applications running at multiple locations, so an accurate picture of 'what is running where' is essential to gain a clear picture of requirements.

For many organisations, a number of other infrastructure considerations may also come in to play if you need to change something. These could include possible restrictions on space and availability of electrical power, or limited additional cooling capabilities in your data centre.

### **Bringing it all together**

With so many different angles to consider on the data you need to protect, you will want to collate all of the requirements you have gathered and consider them as a whole. You will therefore need to conduct some kind of data protection needs analysis exercise, where you identify the different kinds of information in use inside your organisation, and decide what level of protection is appropriate for each. In any exercise such as this, it is essential to get the level of detail right.

You may find you have several types of data in your organisation and it is important to group them according to how important they are to the business. Try to keep the number of groups as low as possible. Note that in the main we are concerned with how the business sees information, rather than how IT sees it (we're not talking about 'content management repositories', 'databases' and so on): this is because we want to set policies according to business requirements, not IT capabilities.



The scope of a needs analysis exercise will be dictated by the size and nature of your business, but the goal is the same – to generate a list of types of data that reflects an accurate picture of the data in your business. Once you have this, you can rank the list according to the relative importance of each type. The following three ‘protection rankings’ can cover the majority of cases:

- ‘Mandatory protection’ – because of value or legal requirement
- ‘Highly desirable’ – protect due to value or convenience
- ‘Less important’ – little or no damage if lost

As mentioned, certain constituencies in the business will say ‘their’ data is all of high-value. They may be right; for example, ‘computer logs’ may seem unimportant but they could be vital when trying to resolve a security problem. But once people understand that there may be a cost associated with protecting their data at the highest levels possible, their priorities may well change. Indeed, the idea of charging for storage resource usage is one that is being considered by organisations to try to place some form of brake on the demand for ‘top tier’ storage. By the time you have finished a data protection needs analysis exercise, you should be able to summarise the different kinds of information in use, together with the characteristics of, and the constraints – such as legal and regulatory obligations – on that information.

### **Introducing ‘RTO and RPO’**

These acronyms may sound like cute androids from a certain popular 1970s space opera but they are in fact two measures which have been defined by the IT industry. We can use them in tandem with the insight you have gained from carrying out an assessment of the different types of data in your organisation.

RTO and RPO relate to the acknowledgement we have already made about not all data needing the same level of protection. Where your assessment has identified the value of different types of information to

your business, RTO and RPO can help you assess how to treat your data:

- **RTO – Recovery Time Objective:** how quickly do you need to get the data back/what is the time it takes to restart a failed application or service?
- **RPO – Recovery Point Objective:** how recent must the last complete data set be to minimise the impact to your business of a data loss event? Minutes, hours or a few days?

These measures are central to defining workable data protection policies and implementing the most appropriate data protection mechanisms.

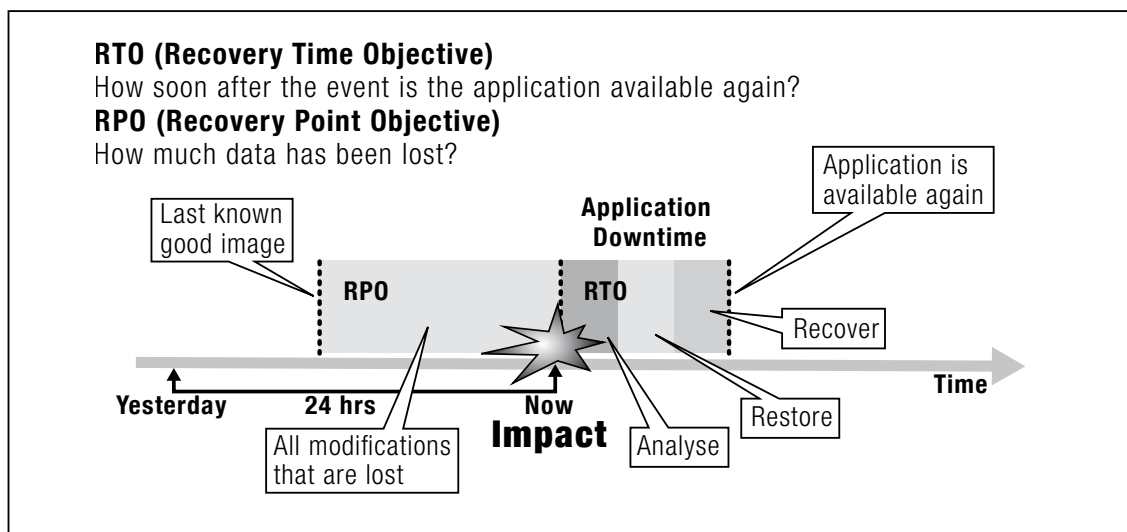


FIGURE 4: Introducing RTO and RPO

Based on the knowledge you have gained, it should be possible to start defining the RPO and RTO for the data in your organisation. Remember, RPO relates to how recent the data is, and RTO is the delay in recovering it. For example, consider your email service. You may view email as being of high-value to the organisation, and therefore mandatory to protect. You may also be legally obliged to

keep email for a number of years. However, having reviewed things, you may also decide that you could survive for a few hours without email (or indeed, you may not). While you want to be sure you have kept everything (short RPO), you could get away with a longer RTO.

It is important to consider some practical limitations of these metrics. The most important one is to remember that it is not possible to achieve an RTO of zero. Firstly it's pretty much impossible to do technically – there will always be a time involved in recovering information from wherever it is stored. Secondly you may need to analyse what went wrong in order to determine what exactly needs recovering, and this will take time as well. Where power or hardware problems may make themselves known quite obviously, data corruption could go un-noticed for a long time, hence finding the 'last known good image' may require some investigation. You may find a table such as the following useful to collate your own requirements:

DATA TYPE	VALUE	CONSTRAINTS	RTO	RPO

TABLE 1

By this point, then, and without 'boiling the ocean' you should be able to identify the following:

- The most important areas of data to the organisation
- What 'protection ranking' they should be given
- What levels of performance you need from the protection mechanisms you use for each type of data (RPO, RTO).

You've probably already noticed that there are a lot of angles to cover here. Before we look at the choices we have to address data protection capabilities, let's address some of the common misconceptions we encounter in this area.

One of the things you will need when you start talking about data protection in your own business is to be sure everyone knows what you are talking about and that expectations are aligned appropriately.

### **Common misconception #1: Backup is the same as archiving**

Bluntly put, an archive does not protect your data. If your archive is lost and it is not backed up, you lose your data. Hence, it is not a data protection mechanism per se. Here are some of the main differences for your reference:

<b>BACKUP</b>	<b>ARCHIVING</b>
Backup takes a copy of the data.	An archive contains your data. It is not a copy, and may need to be backed up.
A backup is used when we want to recover data that has been lost.	An archive is a data store you can retrieve (typically) older data from when you need it.
Backups typically take copies of data regularly.	An archive could contain data which is months, years or decades old.
Data is usually overwritten on a regular basis (daily or weekly).	Data is preserved for analysis or compliance reasons, for as long as is required.

TABLE 2

### Common misconception #2: High availability (HA) is the same as data protection

A high availability system offers (you guessed it) a system with high availability. But what if your data was corrupted? A high availability system without data protection would simply keep running with the corrupted data. HA and data protection are complementary to each other. Again, here are some of the main differences:

DATA PROTECTION	HIGH AVAILABILITY
Target – long term data recovery.	Target – rapid service restoration.
Takes a copy of primary data for recovery upon demand.	Takes a copy of IT infrastructure and primary data available for fast recovery of systems.
Multiple stores held going back over significant periods of time. (Weeks, months, years) (Minutes, hours, days)	
Data is copied regularly. (Daily, weekly, monthly, yearly)	Employed for rapid recovery, data kept usually over short periods of time. Data is replicated regularly. (Real-time, hourly, 3 hourly, daily)
Suitable for handling recovery from data corruption events.	May not work in cases of data corruption where a corruption event remains undetected for time period exceeding roll back capacity.
Low cost compared to infrastructure hosting live data.	Usually requires high cost infrastructure component replication.
Restoration speeds may be slow.	Recovery time is designed to be as short as required (but cannot be reduced to zero).

TABLE 3

## 4. Data protection technology

Like the rest of the technology industry, mechanisms for data protection have been evolving rapidly. Early computers employed disk drives the size of washing machines (which promised to hold 5 megabytes of capacity!) and tape reels as large as bicycle wheels. We can probably state that the first backups were made in the early 1950s, when copies of punched cards were made of the instruction codes of the first commercial computers.<sup>[3]</sup> Cards were replaced by magnetic tape in the 1960s, and given that one tape could store the equivalent of 10,000 punched cards, it's not hard to see the attraction!

### Backup media

The storage technologies at the heart of data protection continue to evolve and diversify to keep pace with data growth and the varying demands of business. Here's an indication of the relative pros and cons of different storage media available:

	ADVANTAGES	DISADVANTAGES
<b>Tape</b>	<ul style="list-style-type: none"> <li>High-speed tape burst read and writes.</li> <li>Security features including hardware-based data encryption.</li> <li>Very low long-term costs.</li> <li>Shelf life – data stored on modern tape systems can last for considerably longer than disk- based alternatives.</li> <li>Easy to move offsite.</li> </ul>	<ul style="list-style-type: none"> <li>Performance is not as fast as disk and SSD systems for either random backup or recovery operations.</li> </ul>
<b>Disk</b>	<ul style="list-style-type: none"> <li>Cost versus SSD – more suitable for 'online' use where data needs to be accessed directly.</li> <li>Speed of random access/low latency.</li> <li>Can be used with replication software to provide high availability.</li> </ul>	<ul style="list-style-type: none"> <li>Expensive (compared to tape).</li> <li>Speed compared with SSD.</li> <li>Relatively energy inefficient</li> <li>Difficult to move 'offsite' unless replication technology is employed and duplicate disk systems are used.</li> <li>Burst access speed low compared to tape.</li> </ul>



	ADVANTAGES	DISADVANTAGES
<b>Solid State Disk (SSD)</b>	<p>Very high performance.</p> <p>Mechanical reliability – no moving parts.</p> <p>Read latency (i.e. the time it takes to read data from storage).</p> <p>Low electricity consumption.</p> <p>Low heat generation.</p>	<p>Write latency when compared to spinning disk.</p> <p>Long-term stability if frequent writes are made to same area of the drive.</p> <p>Low capacity.</p> <p>Expensive compared to disk and tape.</p>

TABLE 4

We have seen no evidence that organisations are looking to dispense with tape, and neither should they: decisions on storage always need to be made in the context of the business opportunity they satisfy and the cost of providing the service. In other words, it's 'horses for courses'.

### Methods and mechanisms

Disk, tape and SSD give us the raw capacity to deal with the 'taking a copy' element of data protection. Over the years, the use of disk-based storage has increased in data protection roles, one that was formerly purely the domain of tape-based solutions. A number of number of data protection mechanisms have emerged, for example:

- Individual backup drives – these allow backups to be taken to individual media, the most common being tape and disk. However, you can even consider a flash-based 'thumb' drive (or 'USB stick') as a single backup device.
- Tape libraries – here, devices can act as a 'juke box' containing multiple tapes. Such a unit can store much more than an individual drive, though of course individual storage media (or the whole lot) can be swapped out.

- **Disk to Disk (D2D)** – this configuration uses a Redundant Array of Inexpensive Disks (RAID) as a backup device, enabling more direct access to backup copies of data.
- **Virtual Tape Libraries (VTL)** – backup software has traditionally been written assuming that backups are tape-based. This configuration makes the disk storage appear as a series of virtual tapes, delivering the advantage of disk but with the ‘appearance’ of tape.
- **Disk to Disk to Tape (D2D2T)** – see where we are going with this? In this configuration, backups first use (faster access) disk, supplemented with (lower cost/capacity) tape.

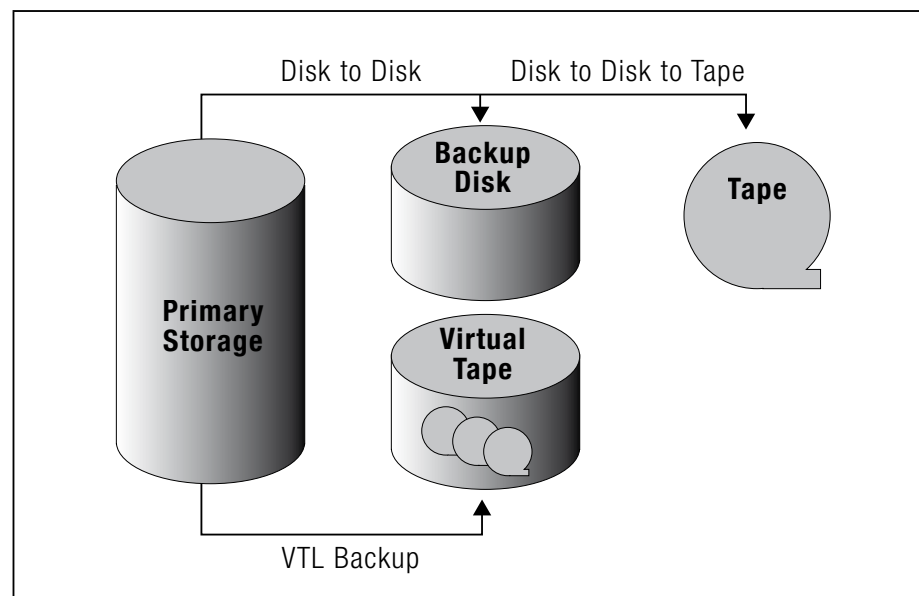


FIGURE 5 Different data protection mechanisms

Given the combinations available, the challenge for most organizations is to understand the characteristics of the wide range of data protection mechanisms available today and where each offers benefits. Table 5 gives some idea of how the offerings can be distinguished.

PLATFORM	CHARACTERISTICS	POTENTIAL USE SCENARIOS
Individual Tape and Disk Drives	<p>Simple to acquire/install.</p> <p>Tapes can be removed for offsite storage.</p> <p>Low acquisition costs.</p>	<p>Good for scenarios where the data to be protected can be written to a single tape/ cartridge of tapes or single disk array.</p> <p>Long term storage/archive of data that does not require very fast access times.</p> <p>Small and medium sized businesses.</p>
Tape Libraries	<p>Low TCO (total cost of ownership).</p> <p>Optimal power efficiency.</p> <p>Long life.</p> <p>Data portability.</p> <p>High capacity.</p> <p>Fast streaming for batch data (e.g. large files that are read end-to-end).</p> <p>Automated storage (Libraries).</p>	<p>Cost conscious environments.</p> <p>Long-term archiving.</p> <p>Storing infrequently accessed data.</p> <p>Off-site data storage.</p> <p>Automated storage.</p>
Disk to Disk (D2D)	<p>Fast backup and file restore.</p> <p>Ease of use.</p> <p>Data deduplication.</p> <p>More frequent RPO.</p>	<p>Short backup windows, frequent RPO.</p> <p>Fast recovery of frequently accessed data.</p> <p>Fast data recovery times for short RTO.</p>
Virtual Tape Libraries (VTL)	<p>Faster data availability and accessibility.</p> <p>Automation and consolidation of data.</p>	<p>Meets requirement for rapid restore times.</p> <p>Shorter backup windows.</p> <p>Automated storage.</p> <p>Flexibility of configuration.</p>
Disk to Disk to Tape (D2D2T)	<p>Combines best characteristics of both tape and disk solutions.</p>	<p>Covers all data protection and archiving scenarios.</p>

TABLE 5

The choice of storage platform and data protection solution depends principally upon the protection, recovery and retention requirements of the business, matched with the cost of the various options available. For situations that demand speed for example, virtual tape libraries may be appropriate. In scenarios where performance and cost might be an issue, archiving solutions, discussed in more detail in the next section, can reduce the volume of data needing regular backup by removing inactive data. Other situations will be dictated by the nature of the information. Some (e.g. patient records) needs to be quickly accessible at all times, which precludes using 'offline' storage mechanisms such as tape. Where offline access is acceptable, tape can be a cost-effective medium. In some cases, a good halfway house can be reached using optical disk. Different regulations may specify certain media, for example the WORM ('write once, read many') nature of optical disk makes it appropriate when records are not to be altered.

### **From technology to data protection**

So, we now have lots of options for data protection mechanisms, but we do not yet have data protection: we get this with the addition of a software layer and appropriate management processes to make the best use of the mechanisms we have discussed. There are a lot of options here too, so let's take the plunge and look at each of them.

**Backup and recovery.** This is the traditional 'entry level' data protection mechanism, where a copy of data is made to a backup system on a regular (daily, weekly or monthly) basis. Having identified the data to be protected, you can configure the software to make copies of data to (traditionally) tape systems, although more recently options may also include writing the protected data to tape, disk or virtual tape. Backups can be either 'full', i.e. taking a complete copy of all data to be protected in one session, or 'incremental' whereby only changes made to data since the last backup run are copied.

A large number of vendors supply such solutions and the functionality of the tools can be extensive, so it is important to choose a suitable tool based on the functional protection and infrastructure requirements

and capabilities of the business. Backup processes are traditionally scheduled to take place when users are not actively changing data to help in both ensuring data consistency for recovery purposes and to minimise the overhead on the IT and network infrastructure. Such runs have, therefore, usually taken place overnight. Recovery processes can only restore data to the last occasion when a backup run took place, which is set by the selected protection schedule.

**Snapshots.** This is where a point-in time copy is made of a set of data, which is particularly useful when data needs to be consistent across several applications. Typical scenarios involve taking snapshots to disk-based stores to enable fast copies to be made, thereby minimising any service interruption. Until recently such capabilities were largely confined to mid-range and high-end data arrays but entry-level systems now also have such capabilities.

**Continuous Data Protection (CDP).** This is where all changes to data are continually streamed to an alternative location, meaning that recovery can be made right up to the point of data loss. This development adds capabilities to the 'traditional' backup and recovery systems identified above. CDP can be useful if data sets change rapidly and the value of the changes is significant. It can also avoid the need for significant manual effort to be invested in updating data changes made after the last backup process ran, if these are periodic in nature.

A disadvantage of CDP can be the need for high volumes of data storage. As of today, this type of protection is confined to file-based systems rather than databases and complex data stores.

**Shredding.** This is not about the actual, physical shredding of disks, but rather the permanent erasing of information from them. This is a matter that needs to be addressed in many organisations as external regulations and some elements of best practice will necessitate the permanent removal of certain data resources at prescribed dates. In addition the growing use of encryption systems offers another means to achieve the same ends.

In addition to the technical solutions already covered, a number of other functional mechanisms exist to support various data storage, data protection and data recovery solutions. These include:

**Deduplication.** This is where repeated files, or indeed patterns within files, are removed to enable faster data transfers and reduce backup storage capacity requirements. There are two modes of operation; 'source deduplication' is carried out on the platform where the data resides, and 'target deduplication' takes place on the backup and recovery servers themselves. The former is appropriate for branch office/ remote locations as it reduces the volume of data over the network. Today, deduplication is a hot topic among IT vendors and there are signs that the end user community is now beginning to look seriously at tools delivering such functionality. The reasons for such interest are clear and are focussed primarily around 'cost' and 'speed'.

A quick look around any user's file system or the storage arrays of nearly every system will show that a significant volume of data consists of copies of other material. This may be in the form of entire copies of files, perhaps word processing documents, spreadsheet or presentations, as well as including multiple generations of documents built on the shoulders of earlier copies, much of which is substantially identical. The idea behind deduplication systems is to find patterns at either a file or component level within data sets. A single master copy of the data is kept and all users are linked to the master copy.

Users of data which has been 'treated' by a deduplication solution should detect no change to the data they access. Any 'repeated' data appears to be available exactly as always but to the system only a single copy is held in storage with very small 'tags' linking all other copies. This may save anything from 10% to, in extreme cases, over 50% of disk storage with the benefit of reducing the need to procure more storage capacity. Furthermore, as there is physically less data to hold, the associated protection and recovery processes can work much faster.

**Replication.** This is where a copy of information is made on a continuous basis but kept online, for example for access by a remote or branch office. If bandwidth is at a premium, it can make sense to create a replica at the remote site when use is low (e.g. overnight) which can then be used during the day. Replication systems operate in two modes: synchronous and asynchronous. The former requires that the source and target disks are kept synchronised in real time and so need low latency network connections. With asynchronous data replication the local disk passes data to a server which in turn relays it to the target disk. The main difference is that changes to the source disk may not be reflected in the target disk for some time. Replication is also an important component of HA systems.

**Compression.** This works on a file transfer basis to enable better use of available network bandwidth. Compression is also now being introduced into some storage systems to reduce the volume of physical data to be held. It looks for patterns in data sources that can be codified in such a way that the net volume of information held to reproduce the data is reduced. There are benefits here for data storage itself as well as in scenarios where data needs to be transmitted over a network, thus possibly reducing overall bandwidth requirements.

**Encryption.** Allows the confidentiality of data to be preserved, whether at rest (i.e. on disks) or in transit (across networks). The principle of protecting data from being accessed and read by anyone other than those deemed appropriate is widely appreciated, but the encryption of data without adversely inhibiting performance is quite a recent development. Encryption generally relies on the existence of keys – the codes used to encrypt or decrypt the data. A spin-off benefit of encryption is how it can be used to render data inaccessible: if storage media are encrypted to a suitably high level then destroying the encryption keys can block access to the associated data – at least until decryption tools become sufficiently powerful to crack the codes!

**Remote management of distributed data.** This is really about data synchronisation processes. Currently, this approach is mainly targeted at 'personal data', where there are few if any occasions when multiple parties would attempt to access the same information. It is coming into vogue for VDI (virtual desktop infrastructure, aka 'virtual desktops') systems where users 'check out' their virtual machine and data files to a portable device, use the device whilst travelling or working from home and then 'check in' the virtual machine and all data files at the end of the journey. At this stage any data files that have been updated are synchronised with central systems to ensure that alterations and additions are captured and can be protected appropriately.

**Archiving.** This is where inactive data is moved to another (generally less accessible and cheaper) location while the end user is left unaware of the data relocation. Archiving has been used in business scenarios where there is an incentive to track information and to ensure that it can be readily searched and recovered when needed. The ability to do this is enhanced if the archiving solution has an indexing feature. It is likely that the archiving of non-sensitive content will become more pervasive as archiving products mature and as the rise in data volumes increase the cost pressure on both the primary data storage facility (the main disks) and traditional backup and recovery systems.

**Hierarchical Storage Management (HSM).** HSM has long existed in mainframe operations and is now being offered across distributed systems too. HSM principles involve physically moving data which is not often used or whose business value is no longer high enough to justify it remaining on top tier storage platforms, and placing it on lower cost platforms. The increasing sophistication of data protection mechanisms is also making it possible for similar principles to be applied when considering backup and recovery options. For example, data that is rapidly changing is also likely to require the most frequent backups and to demand the most rapid recovery when required. Data held on lower tiers of the storage hierarchy are likely to require less frequent backups and frequency of access to it will almost certainly be satisfied with lower recovery times.



Note that data protection mechanisms are not perfect or immune from disaster either! It is not always possible to prevent destruction, accidental or otherwise. Throwing a bucket of water onto a storage array might have an untoward effect for example! However, these scenarios, as unlikely as they may seem, do need to be taken into account as part of an approach to data protection. Depending on your circumstances, a combination of data protection technologies and high availability infrastructure solutions may be seen as most appropriate. Lessons can be drawn from the fields of information management and security, for example, in terms of keeping accurate records of what is stored and security mechanisms such as encryption and digital signatures for enterprise DRM (digital rights management).

Every organisation is likely to adopt a range of approaches to data protection and recovery, especially where requirements for data sets vary throughout their lifetimes. Indeed, even in extreme cases where an organisation has data with similar protection and recovery requirements, it is likely that a range of methods as part of an overall approach to protection would be appropriate. There is no single 'golden bullet' solution that is capable of meeting all the needs of a customer. The important factor is to centralise the management of the protection and recovery processes, to ensure service levels are met and to reduce the cost of service whilst reducing any possible risk exposures.

## Delivery options

In addition to the many available mechanisms, a number of ways to source them exist too, including:

- **In-house.** A traditional purchase, installation and operation of a solution in-house using internal resources or contractors. Such solutions may consist of an individual tool and its associated hardware, or they may be composed from a range of individual point products or may involve an all-encompassing suite. The advantage of using internal resources is that the knowledge of the business should be high, and there can be a greater freedom to match the businesses requirements as closely as possible. Given that such solutions may require significant manual resources to implement, it is quite usual for external skilled resources to be employed during the implementation and integration processes.
- **Managed services.** Specialist data protection companies can offer a number of the mechanisms above as a 'managed service' – that is, while the equipment may exist on-site, they look after it and ensure everything is ship-shape. Managed services tend to be customer-specific and locked down, which is no bad thing for smaller companies, but may prove too inflexible for larger ones.
- **Fully outsourced.** An external company can take full charge of data protection in particular and, indeed, storage in general. This would usually be an option for a larger company with complex data protection needs (possibly requiring consulting time to get things going), and legacy data protection equipment to be taken under new management.

- **Software as a Service.** In this arrangement, specialist providers offer particular capabilities (such as backup) over the Internet. The benefits are highlighted as reduced cost of acquisition and cheaper ongoing costs, yet the long-term cost benefit is still to be proven. At this stage such services are relatively nascent, but we expect to see this grow. SaaS could be attractive to smaller organisations, as it could offer disaster recovery capabilities that may not have been affordable in the past.

As both data protection mechanisms and delivery options (particularly SaaS) mature, we're likely to see major opportunities for specialist providers of new solutions that help address a broader range of data protection challenges. For now, your task is to match the requirements you have identified for yourself with the best tool for the job. We look at this in the next section.

## 5. Making the right data protection choices

So far we have examined the data protection options available, and given an indication of where they might be most suitable. But how should an organisation go about deciding the most appropriate mix of options for their own needs?

The first step towards a data storage strategy is to consider the value of the different types of data that the business must store. As we have already discussed, not all data is created equal and, moreover, the value of data varies over time.

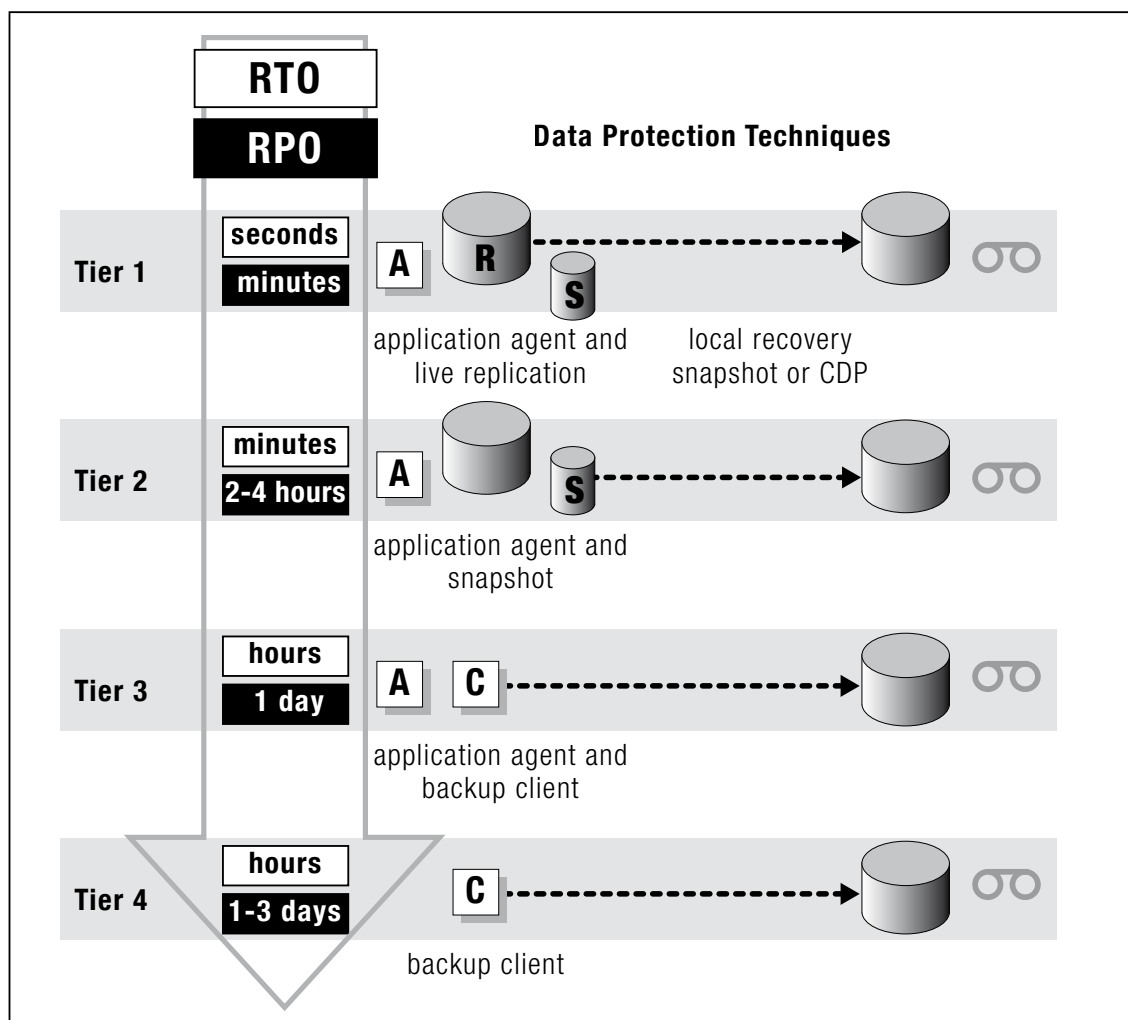


FIGURE 6: Data with different needs = different data protection approaches

Now we get down to the nitty-gritty of meeting your needs. We don't want to risk this turning into a project management textbook – so we won't dwell on best practice for defining the business case, managing stakeholders, and so on. It is likely that your business already has 'legacy' solutions in place, so a good first step is to derive one or more 'ideal world' views of how you would like your data protection mechanism to look.



Developing an 'ideal world' view gives an opportunity to 'work back from the answer', rather than trying to patch up what may be inappropriate products and approaches. Once you have a clear view of where you are trying to get, you can undertake an assessment of what capabilities you already have in place. The resulting gap analysis can help you establish:

- The priorities in terms of what needs to be done first
- An action plan to set out shorter and longer term tasks to resolve any issues
- The risks to the business of not having an appropriate mechanism in place
- Cost models for acquisition and ongoing operations.

It may be obvious to you from the outset that your data protection mechanisms require a complete overhaul. If you have the cash, of course this can make things much easier, but few organisations are in this enviable position. More likely and for reasons we have already stated, you will be working to a limited budget. Therefore it is important to be pragmatic and make the best use of existing resources.

To illustrate how to approach data protection, we consider a number of example scenarios and how data protection capabilities might be employed to meet them. The scenarios are:

- Responding to inadequate data backup and restore
- Business continuity and disaster recovery
- Data protection improvements for compliance reasons
- Branch office/remote office data protection.

Let's consider each of them in turn. Please note that these are not exhaustive examples and do not indicate best practice in each area, but they should give an indication of the thought processes and approaches needed for each.

### **Scenario 1: Responding to inadequate data backup and restore**

In this scenario we consider a large services organisation that has grown through acquisition and expansion into new markets and is now operating across multiple sites in a number of countries. Its backup and restore capabilities are inconsistent. In certain areas, backups are conducted according to established policies, but elsewhere in the business backup is carried out in an inconsistent manner. In addition, while policies exist they are generally not up-to-date. In particular, they fail to cover more recent working practices such as home and mobile working.

Given the nature of the business, a review of data types has established the following categories of data that require protection:

- Business information such as sales and accounting data. The majority of this is accessed via business applications and is stored in databases.
- 'Paperwork' associated with client and supplier management, including tender documents, responses, contracts, insurances and so on. Most of this is electronic; the rest is paper-based.

- Documents and files associated with projects, including plans, spreadsheets and other office documents.
- Marketing materials, brochures and web site content.
- Email, which is used as the primary communications mechanism by everyone in the company.

A data protection needs analysis has established the following:

DATA TYPE	VALUE	CONSTRAINTS	RTO	RPO
Business information	Very high	Needs to be retained for legal reasons. Confidentiality is essential.	Minutes	Minutes
'Paperwork'	High	While referenced only occasionally, the company is legally obliged to retain copies and show them when required.	Days (in general other parties will also keep copies)	Days (unless legal request)
Project files	High	Modifications are often made when consultants offline, e.g. on laptops. Confidentiality policy may vary depending on the client and project.	As short as possible depending on online status of consultants.	As short as possible.
Marketing materials	Medium	May contain NDA information prior to launch.	Day	Day
Email	Medium	Needs special consideration in some countries due to local laws.	Day	Day

TABLE 6

In addition, the organisation has determined that (following a consolidation exercise a couple of years before) the core IT systems are relatively stable, therefore the main risks come from accidental data loss, targeted attacks by hackers, or damage due to fire/flood.

Having analysed all this, the plan is to:

- Centralise backup of all data with the exception of email, which will be dealt with by local teams in the countries concerned.
- Encrypt backups of business information (by default) and of specific projects (by exception). All other data will be backed up 'in clear', i.e. unencrypted.
- Employ disk-based backup for business information, and tape for all other data.

Attention then turns to the protection mechanisms required. A total of 50 terabytes of existing information will require backing up, which is currently the maximum that the existing backup capabilities can handle. However, as much of the information is unstructured content (i.e. documents and files) relating to projects, the decision is taken to incorporate deduplication technology into the solution. This reduces the backup requirement to 20 terabytes, leaving 30 terabytes free and meaning no new hardware is required.

### **Scenario 2: Business continuity and disaster recovery**

In this scenario a petrochemical company is feeling vulnerable due to having its main data centres on the same site as one of its main refineries – a contributing factor was the recent news reports about a competitor's refinery fire (which resulted in significant commercial damage). The board has authorised a review of existing disaster recovery (DR) capabilities, and promised adequate funding to ensure the company doesn't go the same way.

Once again, a needs analysis has identified several kinds of information that are being stored on site, two of which stand out – customer metering data and site management information.



DATA TYPE	VALUE	CONSTRAINTS	RTO	RPO
Customer metering records	Very high	Legacy system which cannot be moved.  Large quantities of historical information with frequent updates.	Minutes	Minutes
Site management information	Very high	Doesn't change much but essential if problems on site.	Days (in general other parties will also keep copies)	Days (unless legal request)

TABLE 7

An ideal solution would incorporate both backup/recovery and replication, in order to protect against the most likely risks of systems failure or data loss. Owing to the large quantities of mission-critical data involved, the proposed solution is to use replication and continuous data protection in parallel to an offsite location, backing up the replica to tape as a basic, but effective solution. As an additional measure, network compression technology is used to ensure that the link between the data centre and the offsite location has sufficient capacity.

While DR capabilities do exist for other types of data being stored elsewhere, the organisation decides to invest in the above technologies in such a way that it can incorporate other data at a later point.

### **Scenario 3: Data protection improvements for compliance reasons**

In this scenario, a financial institution is working to meet some changes in the law. These include: the need to keep customer transaction data for far longer than before (10 years rather than 3 years); to demonstrate, if requested, how customer data is being managed and protected; to enable customers to access their own historical records on request; and to report to all affected customers in the case of data breaches.

Following a needs analysis, a D2D2T approach is chosen, which enables certain information to be migrated onto tape after a certain length of time, while other data can be kept online and backed up to tape as necessary. While existing hardware can be used to support this, it does require upgraded data protection software to enable policies to be set and modified as necessary.

Without such forethought, meeting the set requirements could have been very costly in terms of new hardware and software. As it stands, the company has made best use of existing capabilities with minimal additional cost. It should also be noted that this is a prime scenario where data protection and recovery solutions could well be combined with deduplication technologies and an archiving solution to fulfil the legal and business requirements.

#### **Scenario 4: Branch office/remote office data protection**

Here, we consider an organisation that has a number of branch offices in remote locations. These focus on customer service delivery and have limited personnel available to manage non-customer facing issues. With traditional data protection and recovery systems this scenario has frequently led to situations where each branch office may have its own small servers on site. Backing up such servers poses problems as there are no skilled IT personnel available in the offices to run backup and restore programs, or to ensure that everything is managed on schedule. The potential for 'mishap' is high.

A data protection needs analysis has established the following requirements:

DATA TYPE	VALUE	CONSTRAINTS	RTO	RPO
Business information	Very high	Needs to be retained for legal reasons. Confidentiality is essential.	Minutes	Minutes
Personal customer data	Medium	Needs special consideration in some countries due to local laws.	Day	Day

TABLE 8

This scenario is ideally suited to an approach whereby data protection and management processes are administered in the organisation's data centre. This reduces the risk of mishap due to the lack of available skills at the remote sites. Deduplication could be used in the branch offices, (known as 'at source') and compression and encryption can make the transfer of data to be backed up over the network to the central systems more efficient and secure. Automation and scheduling software ensure that operations are performed according to policy.

Clearly, no single solution provides all the protection required in the simple example scenarios given, but they do show how different approaches to data protection can combine to resolve the challenges highlighted. You may find that your data protection needs are more, or less, complex. However we would still propose you work through the steps outlined, to ensure you have covered all the bases.

### When to use disk or tape?

Each of the scenarios we have described needed to make a trade-off between disk and tape. We summarise the key constraints and considerations in table 9:

SCENARIO	CONSTRAINTS	CONSIDERATIONS	USE TAPE WHEN:	USE DISK WHEN:
Data backup and restore	Quantity of data to be backed up. RPO and RTO. Length of time to store backups. Availability of staff to handle media.	The higher the backup capacity, the more the cost of media becomes a factor. How quickly do you need to get your data back? If backups to be kept for very long periods, the solution may need to have low cost media with a long shelf-life. Centralised management, especially in branch/remote offices, to remove the need for local staff resources.	Budgets are constrained. Speed of access is less of an issue. Longer term storage is required.	Access times are paramount and cost is less of an issue. Disk-based backups are only retained for a short period (potentially before archiving to tape).
Business continuity	The allowable downtime for the business – RTO and RPO. Cost involved.	If business downtime is critical to success, then systems and applications will need to be replicated on a secondary site, potentially in real-time.	Budgets are constrained. A recent point-in-time copy of information would be sufficient for disaster recovery.	Downtime needs to be kept to an absolute minimum and real-time replication is required.

SCENARIO	CONSTRAINTS	CONSIDERATIONS	USE TAPE WHEN:	USE DISK WHEN:
Data retention for compliance reasons	eDiscovery time. Specific media requirements. Security.	Discovery can be very expensive in terms of legal fees if information is not readily accessible. Indexing and archiving can assist in discovery scenarios. If data is to be retained for very long periods, the total cost of re-copying from older media (disk or tape) becomes a factor. Some data retention laws stipulate the type of media to be used, for example concerning non-immutable media.	Lowest cost is required.  A reliable long-term shelf life is needed to hold data over extended periods.  WORM or tape-based encryption is desirable.	Fast time to eDiscovery is an overriding factor.
Branch office/ remote office data protection	Availability of staff to handle media.	Centralised management, especially in branch/remote offices, to remove the need for local staff resources.	Budgets are constrained.  Speed of access is less of an issue.  Longer term storage is required.	Access times are paramount and cost is less of an issue.  Disk-based backups are only retained for a short period (potentially before archiving to tape).

TABLE 9

By now you should have a clearer idea of how and where to start protecting your core data assets. We hope you can ‘see the wood for the trees’ and have an idea also of what solutions might be appropriate to your needs. Next we consider how to meet those needs in practice.

## 6. Taking things forward with data protection

In the course of this guide we have been looking at what needs to be done to protect data. What we haven't covered yet is how to approach the deployment of data protection solutions. This is not really the place for a treatise on project management best practice; however, we can offer some pointers when it comes to the specifics. These boil down to:

- Putting in place a comprehensive yet appropriate business case
- Gaining buy-in and support from relevant stakeholders
- Evaluating and testing potential solutions
- Defining appropriate policies and operational terms of reference.

We look at these below.

### **Getting the business case right**

Documenting the business case for data protection can be difficult because, as we have already discussed, data protection mechanisms do not add any particular 'value' to the business. It is unsurprising, perhaps, that data protection vendors sometimes rely on spreading fear, uncertainty, and doubt in their marketing literature in an attempt to drive a need for their products.

Ultimately, we can think about data protection in the same way that we think about insurance. It is worth taking a lead from the insurance sales handbook, about how to sell products 'that people don't really want'. The essential technique is to present two scenarios: one in which insurance mechanisms are in place, and in the other where they are not.

It is easy to see how this might apply when we put together a business case for data protection, which should incorporate the following:

- A map of the value of information to the business – the same information we recommend pulling together during a data protection needs analysis. It will be nearly impossible to set an absolute value on this, so we suggest you don't try.
- The risks. These should have been derived from the business, so nobody should be in any doubt that they exist. As we have discussed, the impact of the risks will have an associated and demonstrable cost.
- The cost of measures to be put in place to mitigate these risks.

However, you need to remember that data protection is a trade-off and, as a subset of business risk management, is ultimately a business decision. If your data centres have been built on a flood plain, and the CEO is aware of the dangers and still doesn't want to do anything about them, then you cannot do a great deal apart from clearly stating the risks and potential impacts.

### **Gaining buy-in and support from relevant stakeholders**

As we have already discussed, the challenge to data protection can be to build any interest at all, never mind gaining agreements on funding. As part of this exercise, it is worth identifying data owners: these are people inside the organisation who have responsibility for certain data. It is they who will care the most should anything untoward happen to it.

In a situation where management does not want to buy into data protection, regulatory compliance can be your friend. The easiest approach may be to tell them that they have no choice. Again, however, it may be within the organisation's rights not to protect anything, subject to compliance laws. However, it is far better that this choice is made actively rather than passively.

### **Evaluating and testing potential solutions**

You will eventually reach the stage where you have identified a number of potential options in terms of the data that you want to protect, and the mechanisms available to protect it. When comparing options you

can expect to end up with a range: some will make the most of what you already have in place and some that involve replacing it all completely. Each option will have an associated shortlist of vendors, but do not feel you need to rush into a decision, or only work with existing suppliers.

Some of the ways of assessing data protection requirements that we discussed earlier can now help you in the decision-making process. For example, 'mandatory' requirements will enable you to determine which providers should be on the shortlist, then 'highly desirable' requirements will enable you to decide between them.



Before making your final decision on the provider, it is very important to evaluate the proposed solutions and ensure that they will meet your needs. Clearly, it can be difficult to test data protection in your own real-world scenario: you cannot mock up a flood for example. However, examples of what you can do are:

- Work through a recovery situation on paper (or indeed, ask the vendor to do so), and assure yourself that all the pieces are in place.
- Talk to reference customers who have resolved similar needs to your own.
- Run a pilot, at least to assure yourselves that information flows and operational constraints can be met.
- Visit a vendor's laboratories or reference installations and talk to the experts who are running them.

Above all, make good use of the expertise of the vendor or service provider. After all, you are going to be spending a lot of money with them potentially, so make them work for it!



## **Defining appropriate policies and operational terms of reference**

Data cannot be protected by technology alone, and IT solutions do not make you compliant with specific laws simply by installing them. To be truly effective, data protection requires a combination of people, process and technology to work together. People can only act appropriately if they know what is expected of them. Hence, user awareness training driven by business requirements is important. For example, you may request (or mandate) that mobile staff synchronise their data onto the network on a regular basis so that it can be backed up, and lay down rules regarding when and how data is copied and removed from a system or user's device. There may be other criteria, for example ensuring that the manager signs off any new computer systems to ensure that they are subject to the corporate data protection policies you have defined. Implementing all kinds of clever protection can be rendered academic if there is no way of extending their reach whenever a new business process or IT investment is rolled out.

You will also need to review how things stand from an operational perspective. For example, do current management processes need to change to take new data protection mechanisms into account? Will there be any requirement for training or new skills to be brought into the organisation? What about management reporting? Particularly with respect to compliance, you may now be required to report any losses of the data that has been protected, and when it happened.

Perhaps the biggest challenge is one of sustainability. Whatever you are putting in place now will need to work today. However, over the months and years the business will change and so will data protection needs. Each new office opened, department reorganised or business application decommissioned will have an impact on data protection requirements. You need to be sure that whatever approach you have put in place is subject to regular review so that you can be confident that the level of protection is still appropriate.

## 7. Tips and tricks

The art of selecting a data protection and recovery solution is ensuring that it delivers the capabilities required at levels of service appropriate to the environment in which it has to operate. Moreover, the chosen solution needs to fit into the existing IT infrastructure and its management processes while simultaneously attempting to be future proof. It goes without saying that all of this should be achieved at the least possible cost and with minimum risk!

So what can be done to smooth out a data protection project or initiative? The steps below have shown themselves time and again to help such projects deliver the desired value:

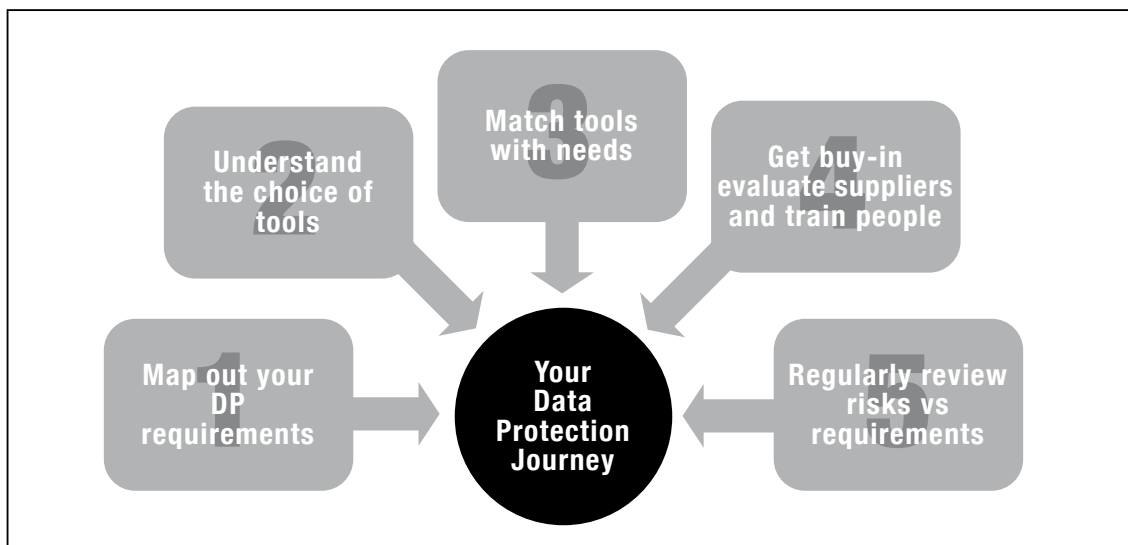


FIGURE 7: Summing it all up

**Gain high level buy-in.** The importance of getting senior management involved in any data protection project should not be underestimated. As matters are likely to arise which require operational process change and a widespread acceptance of different levels of data recovery services, it is essential to have them on board to explain why things are changing and to gain their own acknowledgement and acceptance of the new situation.

**Manage customer/stakeholder expectations.** Once the project has high level support, it is important to communicate the impact of any backup and recovery process changes that have been implemented, especially if different data sets or operational groups will now be subject to different recovery service metrics. It is vital that everyone understands the level of service they will receive rather than have some groups expecting higher service levels that may be available to other, more business critical resources.

**Educate users.** This should be a major feature when it comes to managing expectations and also to ensuring that everyone knows of any new or changed processes they need to follow to ensure that all business data is protected appropriately.

**Regularly review risks and needs.** As we all know, nothing in IT stands still for long. And as the ability to exploit data effectively is at the heart of the success of most businesses, it is essential to ensure that any operational changes in business practice are reflected in data protection and recovery services.

**Talk to a lawyer.** If your industry is governed by any external regulations – and frankly that is just about everyone under the sun – then it may be worthwhile getting an opinion on how data should be protected in its broadest context, along with how long it should be kept available for inspection and any appropriate guidance on how it should be destroyed at the end of its working life.

**Listen to the business.** Business users will have some idea of how valuable their information is in the grand scheme of things and their requirements for data protection and recovery services. However, they may need some encouragement to tell you the real story rather than that ‘everything is absolutely essential and of the highest importance’!



**Test, test, test!** The biggest problem in most data protection and recovery scenarios is that the testing of recovery processes and procedures does not take place regularly over time. Without testing as things change, it will be impossible to ensure that data can be recovered in a timely manner. Do not allow regular testing to fall off the 'To Do' list.

**Centralise management of data protection and recovery.** The most effective way to protect and recover data is to utilise a centralised management and administration solution. However, if all data protection and recovery requirements are designed to work in a single system, it is absolutely essential to ensure that the system is suitably robust and will be operational come what may. The alternative may be to split workloads between complementary systems that have some ability to offer operational cover.

**Don't try to boil the ocean.** Do not try and do everything at once. Start with achievable sub-projects that test out systems, processes and skills and that offer the chance to modify things as your understanding of systems and business needs refine over time. This approach also provides a means to report back to business on successes that are achieved rapidly, and within a single budget cycle. Trying to change everything at once is resource intensive and operationally risky.

In conclusion, while data protection may be an art, by understanding your own needs and what risks you face, you can get yourself in a position to do something about them. The ways we create and use data in our daily lives are changing as rapidly as the platforms we use to store and indeed protect it. Successful data protection lies in staying 'on the ball', considering and dealing with changes as they happen and managing the careful balancing act between cost and risk.

We hope that you have found this guide useful, and we welcome any feedback you may have.

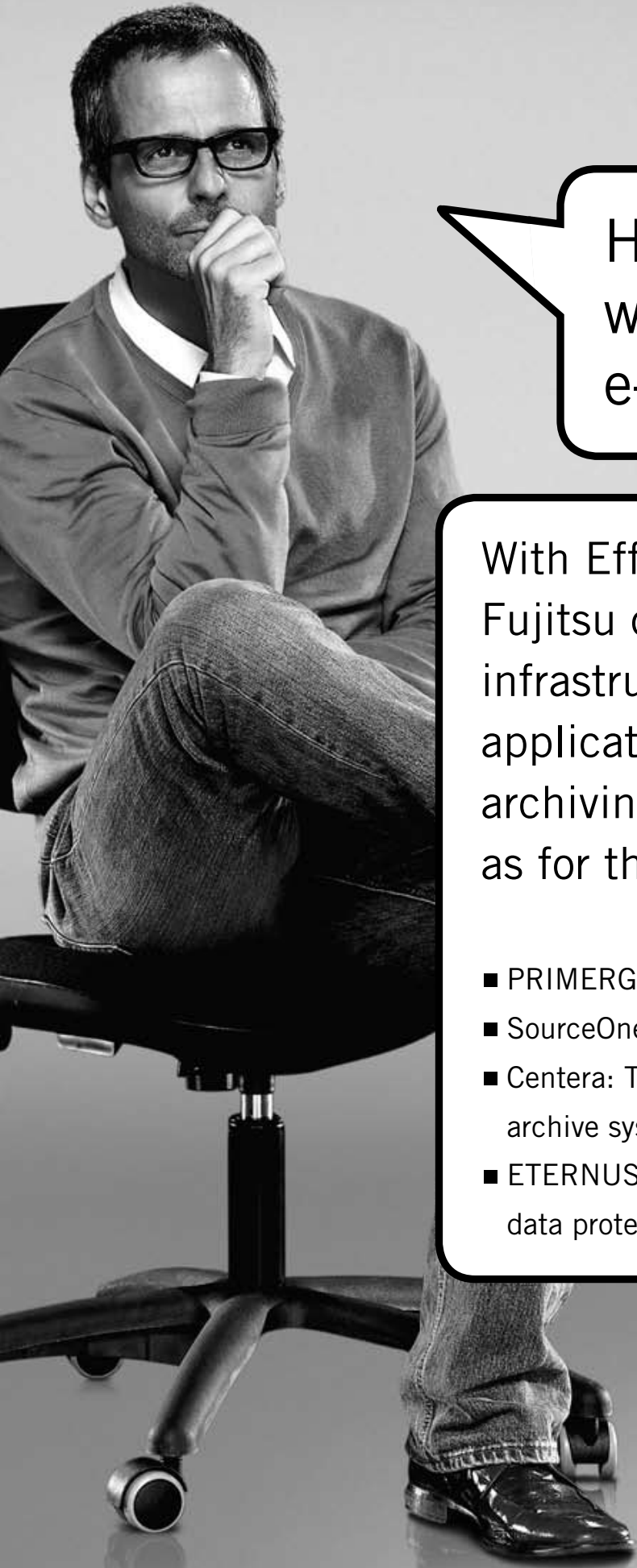
# References

[1, 2] IT Risk in Context – Towards a more integrated approach  
Freeform Dynamics

<http://www.freeformdynamics.com/fullarticle.asp?aid=74>

[3] Punched cards – a brief illustrated technical history  
Douglas W Jones

<http://www.cs.uiowa.edu/~jones/cards/history.html>




How can I stop us  
wasting time on  
e-mail management?

With Efficient E-Mail.

Fujitsu offers a comprehensive infrastructure for your e-mail application; for operational and archiving environments as well as for their backup.

- PRIMERGY X86 server: The better alternative
- SourceOne: The intelligent archiving software
- Centera: The simple, affordable & secure archive system
- ETERNUS CS: The virtual tape for intelligent data protection

  
**FUJITSU**

# Technology needs context

Freeform Dynamics is a research and analysis firm. We track and report on the business impact of developments in the IT and communications sectors.

An innovative research methodology allows us to gather feedback directly from those involved in IT strategy, planning, procurement and implementation.

Our output is grounded in real-world practicality for use by mainstream business and IT professionals.

For further information or to subscribe to the Freeform Dynamics research service, visit our website or contact us at **[info@freeformdynamics.com](mailto:info@freeformdynamics.com)**



**[freeformdynamics.com](http://freeformdynamics.com)**

Published by

## Fujitsu Technology Solutions

Mies-van-der-Rohe-Strasse 8, 80807 Munich,  
Germany

Copyright: ©Freeform Dynamics 2009

Printed in the UK

Order no.: 10734-3-1009-en

Contact: [ts.fujitsu.com/contact](http://ts.fujitsu.com/contact)

All rights reserved, including rights created by patent grant or registration of a utility model. all designations used in this document may be trademarks, the use of which by third parties for their own purposes could violate the rights of their owners. We reserve the right to change delivery options or make technical modifications.