4 Service Design processes

This chapter describes and explains the fundamentals of the key supporting Service Design processes. These processes are principally responsible for providing key information to the design of new or changed service solutions. There are five aspects of design that need to be considered:

- The design of the services, including all of the functional requirements, resources and capabilities needed and agreed
- The design of Service Management systems and tools, especially the Service Portfolio, for the management and control of services through their lifecycle
- The design of the technology architectures and management systems required to provide the services
- The design of the processes needed to design, transition, operate and improve the services, the architectures and the processes themselves
- The design of the measurement methods and metrics of the services, the architectures and their constituent components and the processes.

A results-driven approach should be adopted for each of the above five aspects. In each, the desired business outcomes and planned results should be defined so that what is delivered meets the expectations of the customers and users. Thus this structured approach should be adopted within each of the five aspects to deliver quality, repeatable consistency and continual improvement throughout the organization. There are no situations within IT service provision with either internal or external service providers where there are no processes in the Service Design area. All IT service provider organizations already have some elements of their approach to these five aspects in place, no matter how basic. Before starting on the implementation of the improvement of activities and processes, a review should be conducted of what elements are in place and working successfully. Many service provider organizations already have mature processes in place for designing IT services and solutions.

All designs and design activities need to be driven principally by the business needs and requirements of the organization. Within this context they must also reflect the needs of the strategies, plans and policies produced by Service Strategy processes, as illustrated in Figure 4.1.

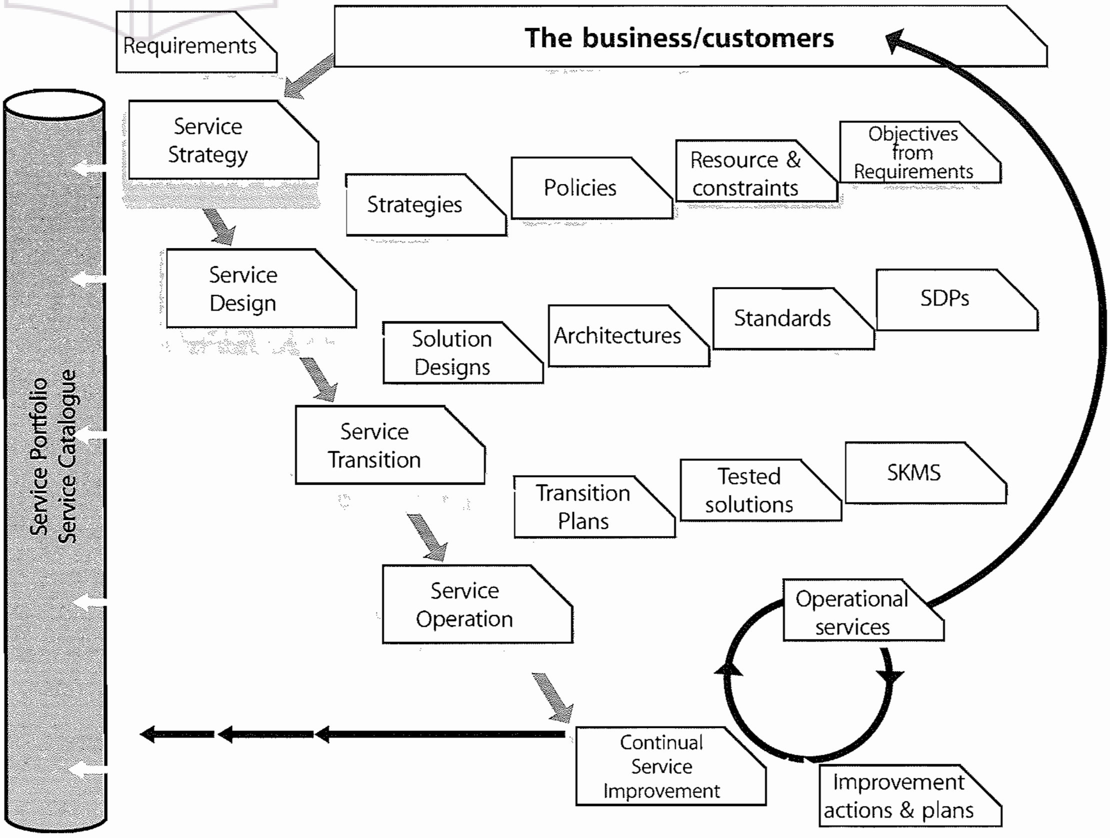


Figure 4.1 The key links, inputs and outputs of Service Design

Figure 4.1 gives a good overview of the links, inputs and outputs involved at each stage of the Service Lifecycle. It illustrates the key outputs produced by each stage, which are used as inputs by the subsequent stages. The Service Portfolio acts as 'the spine' of the Service Lifecycle. It is the single integrated source of information on the status of each service, together with other service details and the interfaces and dependencies between services. The information within the Service Portfolio is used by the activities within each stage of the Service Lifecycle.

The key output of the Service Design stage is the design of service solutions to meet the changing requirements of the business. However, when designing these solutions, input from many different areas needs to be considered within the various activities involved in designing the service solution, from identifying and analysing requirements, through to building a solution and SDP to hand over to Service Transition.

In order to develop effective and efficient service solutions that meet and continue to meet the requirements of the business and the needs of IT, it is essential that all the inputs and needs of all other areas and processes are reconsidered within each of the Service Design activities, as illustrated in Figure 4.2. This will ensure that all service solutions are consistent and compatible with existing solutions and will meet the expectations of the customers and users. This will most effectively be achieved by consolidating these facets of the key processes into all of these Service Design activities, so that all inputs are automatically referenced every time a new or changed service solution is produced.

4.1 SERVICE CATALOGUE MANAGEMENT

4.1.1 Purpose/goal/objective

The purpose of Service Catalogue Management is to provide a single source of consistent information on all of the agreed services, and ensure that it is widely available to those who are approved to access it.

The goal of the Service Catalogue Management process is to ensure that a Service Catalogue is produced and maintained, containing accurate information on all operational services and those being prepared to be run operationally.

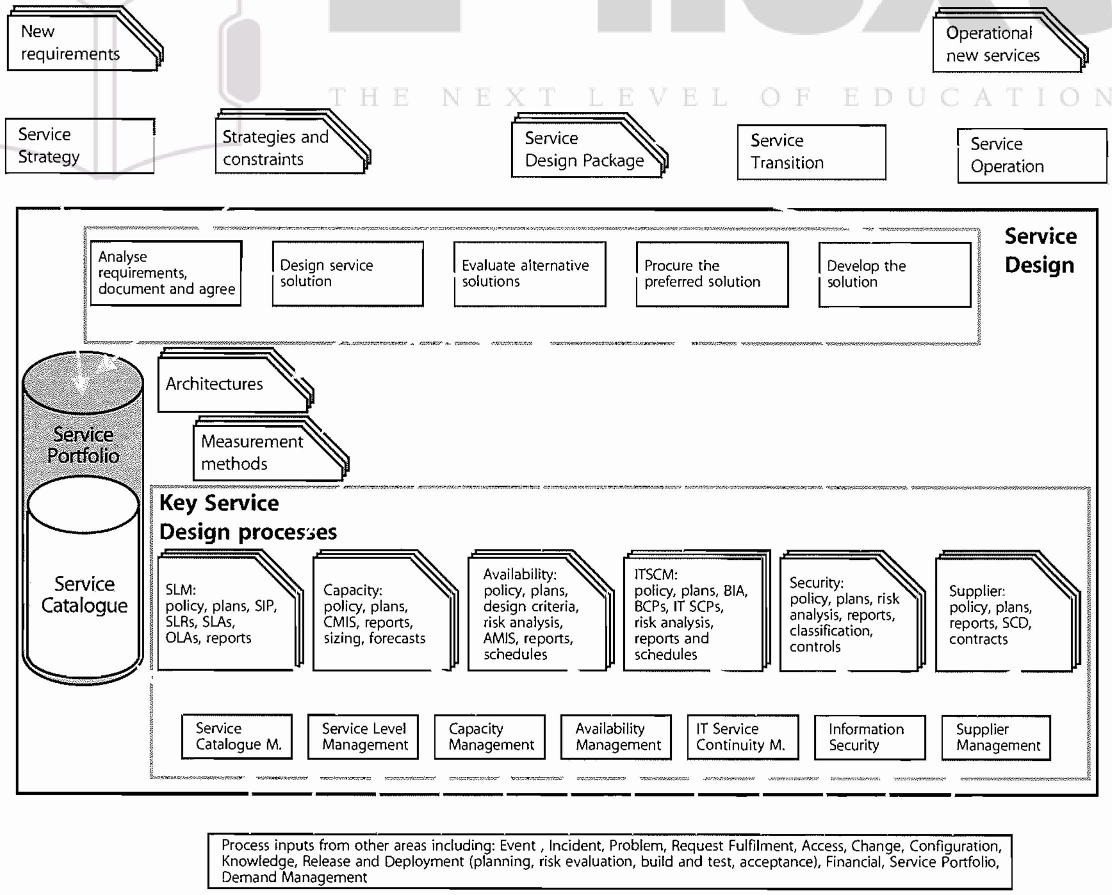

New requirements
Operational new services
Service Strategy
Strategies and constraints
Service Design Package
Service Transition
Service Operation
Service Design
Analyse requirements, document and agree
Design service solution
Evaluate alternative solutions
Procure the preferred solution
Develop the solution
Service Portfolio
Service Catalogue
Architectures
Measurement methods
Key Service Design processes
SLM: policy, plans, SIP, SLRs, SLAs, OLAs, reports
Capacity: policy, plans, CMIS, reports, sizing, forecasts
Availability: policy, plans, design criteria, risk analysis, AMIS, reports, schedules
ITSCM: policy, plans, BIA, BCPs, IT SCPs, risk analysis, reports and schedules
Security: policy, plans, risk analysis, reports, classification, controls
Supplier: policy, plans, reports, SCD, contracts
Service Catalogue M.
Service Level Management
Capacity Management
Availability Management
IT Service Continuity M.
Information Security
Supplier Management
Process inputs from other areas including: Event , Incident, Problem, Request Fulfilment, Access, Change, Configuration, Knowledge, Release and Deployment (planning, risk evaluation, build and test, acceptance), Financial, Service Portfolio, Demand Management

Figure 4.2 Service Design – the big picture

The objective of Service Catalogue Management is to manage the information contained within the Service Catalogue, and to ensure that it is accurate and reflects the current details, status, interfaces and dependencies of all services that are being run, or being prepared to run, in the live environment.

4.1.2 Scope

The scope of the Service Catalogue Management process is to provide and maintain accurate information on all services that are being transitioned or have been transitioned to the live environment.

The Service Catalogue Management activities should include:

- Definition of the service
- Production and maintenance of an accurate Service Catalogue
- Interfaces, dependencies and consistency between the Service Catalogue and Service Portfolio
- Interfaces and dependencies between all services and supporting services within the Service Catalogue and the CMS
- Interfaces and dependencies between all services, and supporting components and Configuration Items (CIs) within the Service Catalogue and the CMS.

4.1.3 Value to the business

The Service Catalogue provides a central source of information on the IT services delivered by the service provider organization. This ensures that all areas of the business can view an accurate, consistent picture of the IT services, their details and their status. It contains a customer-facing view of the IT services in use, how they are intended to be used, the business processes they enable, and the levels and quality of service the customer can expect for each service.

4.1.4 Policies, principles and basic concepts

Over the years, organizations' IT infrastructures have grown and developed, and there may not be a clear picture of all the services currently being provided and the customers of each service. In order to establish an accurate picture, it is recommended that an IT Service Portfolio containing a Service Catalogue is produced and maintained to provide a central, accurate set of information on all services and to develop a service-focused culture.

The Service Portfolio should contain all the future requirements for services and the Service Catalogue should contain details of all services currently being provided or those being prepared for transition to the live environment, a summary of their characteristics, and details of the customers and maintainers of each. A degree of 'detective work' may be needed to compile this list and agree it with the customers (sifting through old documentation, searching program libraries, talking with IT staff and customers, looking at procurement records and talking with suppliers and contractors etc.). If a CMS or any sort of asset database exists, these may provide valuable sources of information, although they should be verified before inclusion within either the Service Portfolio or Service Catalogue. The Service Portfolio is produced as part of Service Strategy and should include participation by those involved in Service Design, Transition, Operation and Improvement. Once a service is 'chartered' (being developed for use by customers, Service Design produces the specifications for the service and it is at this point that the service should be added to the Service Catalogue.

Each organization should develop and maintain a policy with regard to both the Portfolio and the Catalogue, relating to the services recorded within them, what details are recorded and what statuses are recorded for each of the services. The policy should also contain details of responsibilities for each section of the overall Service Portfolio and the scope of each of the constituent sections.

The Service Catalogue Management process produces and maintains the Service Catalogue, ensuring that a central, accurate and consistent source of data is provided, recording the status of all operational services or services being transitioned to the live environment, together with appropriate details of each service.

What is a service? This question is not as easy to answer as it may first appear, and many organizations have failed to come up with a clear definition in an IT context. IT staff often confuse a 'service' as perceived by the customer with an IT system. In many cases one 'service' can be made up of other 'services' (and so on), which are themselves made up of one or more IT systems within an overall infrastructure including hardware, software, networks, together with environments, data and applications. A good starting point is often to ask customers which IT services they use and how those services map onto and support their business processes. Customers often have a greater clarity of what they believe a service to be. Each organization needs to develop a policy of what is a service and how it is defined and agreed within their own organization.

To avoid confusion, it may be a good idea to define a hierarchy of services within the Service Catalogue, by qualifying exactly what type of service is recorded, e.g. business service (that which is seen by the customer). Alternatively, supporting services, such as infrastructure services, network services, application services (all invisible to the customer, but essential to the delivery of IT services) will also need to be recorded. This often gives rise to a hierarchy of services incorporating customer services and other related services, including supporting services, shared services and commodity services, each with defined and agreed service levels.

When initially completed, the Service Catalogue may consist of a matrix, table or spreadsheet. Many organizations integrate and maintain their Service Portfolio and Service Catalogue as part of their CMS. By defining each service as a Configuration Item (CI) and, where appropriate, relating these to form a service hierarchy, the organization is able to relate events such as incidents and RFCs to the services affected, thus providing the basis for service monitoring and reporting using an integrated tool (e.g. 'list or give the number of incidents affecting this particular service'). It is therefore essential that changes within the Service Portfolio and Service Catalogue are subject to the Change Management process.

The Service Catalogue can also be used for other Service Management purposes (e.g. for performing a Business Impact Analysis (BIA) as part of IT Service Continuity Planning, or as a starting place for re-distributing workloads, as part of Capacity Management). The cost and effort of producing and maintaining the catalogue, with its relationships to the underpinning technology components, is therefore easily justifiable. If done in conjunction with prioritization of the BIA, then it is possible to ensure that the most important services are covered first. An example of a simple Service Catalogue that can be used as a starting point is given in Appendix G.

The Service Catalogue has two aspects:

- **The Business Service Catalogue:** containing details of all the IT services delivered to the customer, together with relationships to the business units and the business process that rely on the IT services. This is the customer view of the Service Catalogue.
- **The Technical Service Catalogue:** containing details of all the IT services delivered to the customer, together with relationships to the supporting services, shared services, components and CIs necessary to support the provision of the service to the business. This should underpin the Business Service Catalogue and not form part of the customer view.

The relationship between these two aspects is illustrated in Figure 4.3.

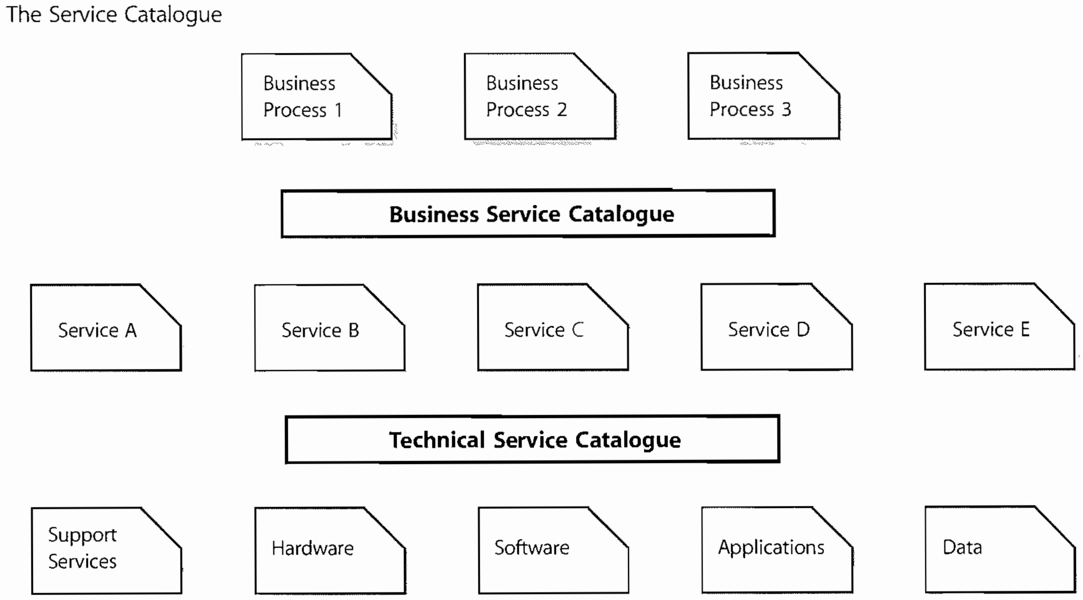


Figure 4.3 The Business Service Catalogue and the Technical Service Catalogue

Some organizations only maintain either a Business Service Catalogue or a Technical Service Catalogue. The preferred situation adopted by the more mature organizations maintains both aspects within a single Service Catalogue, which is part of a totally integrated Service Management activity and Service Portfolio. More information on the design and contents of a Service Catalogue is contained in Appendix G. The Business Service Catalogue facilitates the development of a much more proactive or even pre-emptive SLM process, allowing it to develop more into the field of Business Service Management. The Technical Service Catalogue is extremely beneficial when constructing the relationship between services, SLAs, OLAs and other underpinning agreements and components, as it will identify the technology required to support a service and the support group(s) that support the components. The combination of a Business Service Catalogue and a Technical Service Catalogue is invaluable for quickly assessing the impact of incidents and changes on the business. An example of relationships between the Business and Technical portions of a Service Catalogue is shown in Figure 4.4.

4.1.5 Process activities, methods and techniques

The key activities within the Service Catalogue Management process should include:

- Agreeing and documenting a service definition with all relevant parties
- Interfacing with Service Portfolio Management to agree the contents of the Service Portfolio and Service Catalogue
- Producing and maintaining a Service Catalogue and its contents, in conjunction with the Service Portfolio
- Interfacing with the business and IT Service Continuity Management on the dependencies of business units and their business processes with the supporting IT services, contained within the Business Service Catalogue
- Interfacing with support teams, suppliers and Configuration Management on interfaces and dependencies between IT services and the supporting services, components and CIs contained within the Technical Service Catalogue

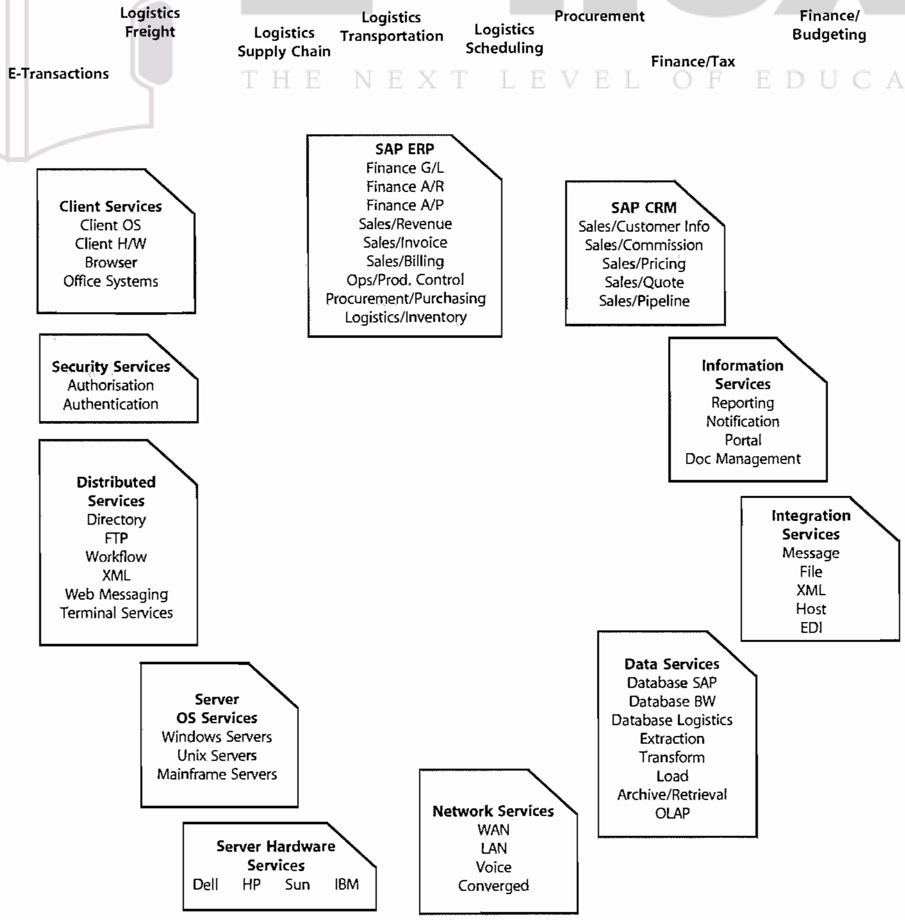


Figure 4.4 Example Service Catalogue

- Interfacing with Business Relationship Management and Service Level Management to ensure that the information is aligned to the business and business process.

4.1.6 Triggers, inputs, outputs and interfaces

There are a number of sources of information that are relevant to the Service Catalogue Management process. These should include:

- Business information from the organization's business and IT strategy, plans and financial plans, and information on their current and future requirements from the Service Portfolio
- Business Impact Analysis, providing information on the impact, priority and risk associated with each service or changes to service requirements
- Business requirements: details of any agreed, new or changed business requirements from the Service Portfolio
- The Service Portfolio
- The CMS
- Feedback from all other processes.

The triggers for the Service Catalogue Management process are changes in the business requirements and services, and therefore one of the main triggers is Request For Changes (RFCs) and the Change Management process. This will include new services, changes to existing services or services being retired.

The process outputs of SCM are:

- The documentation and agreement of a 'definition of the service'
- Updates to the Service Portfolio: should contain the current status of all services and requirements for services
- The Service Catalogue: should contain the details and the current status of every live service provided by the service provider or service being transitioned into the live environment, together with the interfaces and dependencies. An example of a Service Catalogue is contained in Appendix G.

4.1.7 Information management

The key information within the Service Catalogue Management process is that contained within the Service Catalogue. The main input for this information comes from the Service Portfolio and the business via either the Business Relationship Management (BRM) or Service Level Management (SLM) processes. This information needs to be verified for accuracy before being recorded within the Service Catalogue. The information and the Service Catalogue itself need to be maintained using the Change Management process.

4.1.8 Key Performance Indicators

The two main Key Performance Indicators (KPIs) associated with the Service Catalogue and its management are:

- The number of services recorded and managed within the Service Catalogue as a percentage of those being delivered and transitioned in the live environment
- The number of variances detected between the information contained within the Service Catalogue and the 'real-world' situation.

Other measurements and KPIs that could be used are:

- Business users' awareness of the services being provided, i.e. percentage increase in completeness of the Business Service Catalogue against operational services
- IT staff awareness of the technology supporting the services:
  - Percentage increase in completeness of the Technical Service Catalogue against IT components that support the services
  - Service Desk having access to information to support all live services, measured by the percentage of incidents without the appropriate service-related information.

4.1.9 Challenges, Critical Success Factors and risks

The major challenge facing the Service Catalogue Management process is that of maintaining an accurate Service Catalogue as part of a Service Portfolio, incorporating both the Business Service Catalogue and the Technical Service Catalogue as part of an overall CMS and SKMS. This is best approached by developing stand-alone spreadsheets or databases before trying to integrate the Service Catalogue and Service Portfolio within the CMS or SKMS. In order to achieve this, the culture of the organization needs to accept that the Catalogue and Portfolio are essential sources of information that everyone within the IT organization needs to use and help maintain. This will often assist in the standardization of the Service Catalogue and the Service Portfolio and enable increase in cost performance through economies of scale.

The main Critical Success Factors for the Service Catalogue Management process are:

- An accurate Service Catalogue
- Business users' awareness of the services being provided
- IT staff awareness of the technology supporting the services.

The risks associated with the provision of an accurate Service Catalogue are:

- Inaccuracy of the data in the catalogue and it not being under rigorous Change control
- Poor acceptance of the Service Catalogue and its usage in all operational processes. The more active the catalogue is, the more likely it is to be accurate in its content
- Inaccuracy of information received from the business, IT and the Service Portfolio, with regard to service information
- The tools and resources required to maintain the information
- Poor access to accurate Change Management information and processes
- Poor access to and support of appropriate and up-to-date CMS and SKMS
- Circumvention of the use of the Service Portfolio and Service Catalogue
- The information is either too detailed to maintain accurately or at too high a level to be of any value. It should be consistent with the level of detail within the CMS and the SKMS.

4.2 SERVICE LEVEL MANAGEMENT

Service Level Management (SLM) negotiates, agrees and documents appropriate IT service targets with representatives of the business, and then monitors and produces reports on the service provider's ability to deliver the agreed level of service. SLM is a vital process for every IT service provider organization in that it is responsible for agreeing and documenting service level targets and responsibilities within SLAs and SLRs, for every activity within IT. If these targets are appropriate and accurately reflect the requirements of the business, then the service delivered by the service providers will align with business requirements and meet the expectations of the customers and users in terms of service quality. If the targets are not aligned with business needs, then service provider activities and service levels will not be aligned with business expectations and problems will develop. The SLA is effectively a level of assurance or warranty with regard to the level of service quality delivered by the service provider for each of the services delivered to the business.

The success of SLM is very dependent on the quality of the Service Portfolio and the Service Catalogue and their contents, because they provide the necessary information on the services to be managed within the SLM process.

4.2.1 Purpose/goal/objective

The goal of the Service Level Management process is to ensure that an agreed level of IT service is provided for all current IT services, and that future services are delivered to agreed achievable targets. Proactive measures are also taken to seek and implement improvements to the level of service delivered.

The purpose of the SLM process is to ensure that all operational services and their performance are measured in a consistent, professional manner throughout the IT organization, and that the services and the reports produced meet the needs of the business and customers.

The objectives of SLM are to:

- Define, document, agree, monitor, measure, report and review the level of IT services provided
- Provide and improve the relationship and communication with the business and customers
- Ensure that specific and measurable targets are developed for all IT services
- Monitor and improve customer satisfaction with the quality of service delivered
- Ensure that IT and the customers have a clear and unambiguous expectation of the level of service to be delivered
- Ensure that proactive measures to improve the levels of service delivered are implemented wherever it is cost-justifiable to do so.

4.2.2 Scope

SLM should provide a point of regular contact and communication to the customers and business managers of an organization. It should represent the IT service provider to the business, and the business to the IT service provider. This activity should encompass both the use of existing services and the potential future requirements for new or changed services. SLM needs to manage the expectation and perception of the business, customers and users and ensure that the quality of service delivered by the service provider is matched to those expectations and needs. In order to do this effectively, SLM should establish and maintain SLAs for all current live services and manage the level of service provided to meet the targets and quality measurements contained within the SLAs. SLM should also produce and agree SLRs for all planned new or changed services.

This will enable SLM to ensure that all the services and components are designed and delivered to meet their targets in terms of business needs. The SLM processes should include the:

- Development of relationships with the business
- Negotiation and agreement of current requirements and targets, and the documentation and management of SLAs for all operational services
- Negotiation and agreement of future requirements and targets, and the documentation and management of SLRs for all proposed new or changed services
- Development and management of appropriate Operational Level Agreements (OLAs) to ensure that targets are aligned with SLA targets
- Review of all underpinning supplier contracts and agreements with Supplier Management to ensure that targets are aligned with SLA targets
- Proactive prevention of service failures, reduction of service risks and improvement in the quality of service, in conjunction with all other processes
- Reporting and management of all services and review of all SLA breaches and weaknesses
- Instigation and coordination of a Service Improvement Plan (SIP) for the management, planning and implementation of all service and process improvements.

4.2.3 Value to the business

SLM provides a consistent interface to the business for all service-related issues. It provides the business with the agreed service targets and the required management information to ensure that those targets have been met. Where targets are breached, SLM should provide feedback on the cause of the breach and details of the actions taken to prevent the breach from recurring. Thus SLM provides a reliable communication channel and a trusted relationship with the appropriate customers and business representatives.

4.2.4 Policies/principles/basic concepts

SLM is the name given to the processes of planning, coordinating, drafting, agreeing, monitoring and reporting of SLAs, and the ongoing review of service achievements to ensure that the required and cost-justifiable service quality is maintained and gradually improved. However, SLM is not only concerned with ensuring that current services and SLAs are managed, but it is also involved in ensuring that new requirements are captured and that new or changed services and SLAs are developed to match the business needs and expectations. SLAs provide the basis for managing the relationship between the service provider and the customer, and SLM provides that central point of focus for a group of customers, business units or lines of business.

An SLA is a written agreement between an IT service provider and the IT customer(s), defining the key service targets and responsibilities of both parties. The emphasis must be on agreement, and SLAs should not be used as a way of holding one side or the other to ransom. A true partnership should be developed between the IT service provider and the customer, so that a mutually beneficial agreement is reached – otherwise the SLA could quickly fall into disrepute and a 'blame culture' could develop that would prevent any true service quality improvements from taking place.

SLM is also responsible for ensuring that all targets and measures agreed in SLAs with the business are supported by appropriate underpinning OLAs or contracts, with internal support units and external partners and suppliers. This is illustrated in Figure 4.5.

Figure 4.5 shows the relationship between the business and its processes and the services, and the associated technology, supporting services, teams and suppliers required to meet their needs. It demonstrates how important the SLAs, OLAs and contracts are in defining and achieving the level of service required by the business.

An OLA is an agreement between an IT service provider and another part of the same organization that assists with the provision of services – for instance, a facilities department that maintains the air conditioning, or network support team that supports the network service. An OLA should contain targets that underpin those within an SLA to ensure that targets will not be breached by failure of the supporting activity.

4.2.5 Process activities, methods and techniques

The key activities within the SLM process should include:

- Determine, negotiate, document and agree requirements for new or changed services in SLRs, and manage and review them through the Service Lifecycle into SLAs for operational services
- Monitor and measure service performance achievements of all operational services against targets within SLAs
- Collate, measure and improve customer satisfaction
- Produce service reports

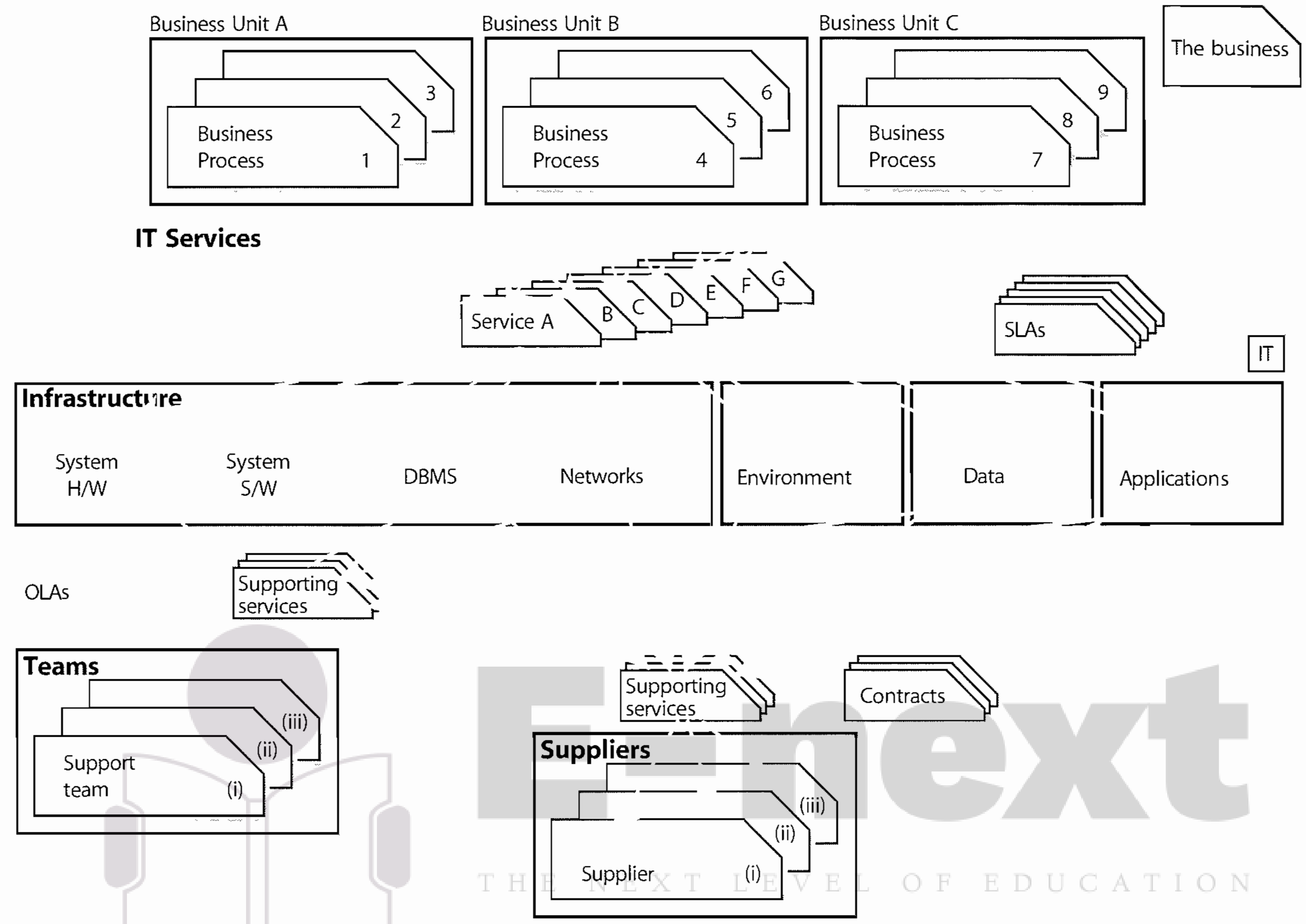


Figure 4.5 Service Level Management

- Conduct service review and instigate improvements within an overall Service Improvement Plan (SIP)
- Review and revise SLAs, service scope OLAs, contracts, and any other underpinning agreements
- Develop and document contacts and relationships with the business, customers and stakeholders
- Develop, maintain and operate procedures for logging, actioning and resolving all complaints, and for logging and distributing compliments
- Log and manage all complaints and compliments
- Provide the appropriate management information to aid performance management and demonstrate service achievement
- Make available and maintain up-to-date SLM document templates and standards.

The interfaces between the main activities are illustrated in Figure 4.6.

Although Figure 4.6 illustrates all the main activities of SLM as separate activities, they should be implemented as one integrated SLM process that can be consistently applied to all areas of the businesses and to all customers. These activities are described in the following sections.

4.2.5.1 Designing SLA frameworks

Using the Service Catalogue as an aid, SLM must design the most appropriate SLA structure to ensure that all services and all customers are covered in a manner best suited to the organization's needs. There are a number of potential options, including the following.

Service-based SLA

This is where an SLA covers one service, for all the customers of that service – for example, an SLA may be established for an organization's e-mail service – covering all the customers of that service. This may appear fairly straightforward. However, difficulties may arise if the specific requirements of different customers vary for the same service, or if characteristics of the infrastructure mean that different service levels are inevitable (e.g. head office staff may be connected via a high-speed LAN, while local offices may have to use a lower-speed WAN line). In such cases, separate targets may be needed within the

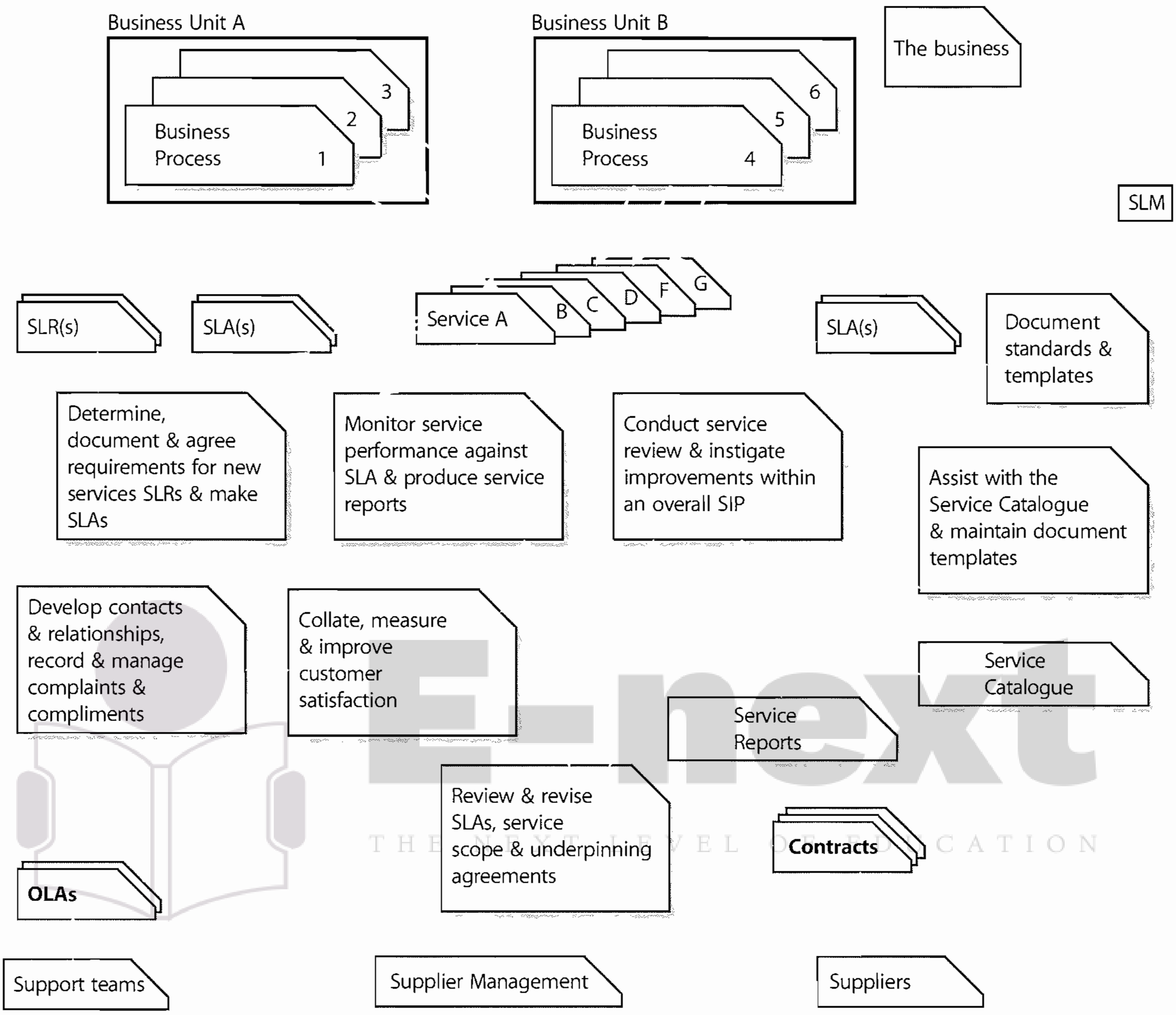

Business Unit A
Business Process 1
2
3
Business Unit B
Business Process 4
5
6
The business
SLM
SLR(s)
SLA(s)
Service A
B
C
D
F
G
SLA(s)
Document standards & templates
Determine, document & agree requirements for new services SLRs & make SLAs
Monitor service performance against SLA & produce service reports
Conduct service review & instigate improvements within an overall SIP
Assist with the Service Catalogue & maintain document templates
Develop contacts & relationships, record & manage complaints & compliments
Collate, measure & improve customer satisfaction
Service Catalogue
Service Reports
Review & revise SLAs, service scope & underpinning agreements
Contracts
OLAs
Support teams
Supplier Management
Suppliers

Figure 4.6 The Service Level Management process

one agreement. Difficulties may also arise in determining who should be the signatories to such an agreement. However, where common levels of service are provided across all areas of the business, e.g. e-mail or telephony, the service-based SLA can be an efficient approach to use. Multiple classes of service, e.g. gold, silver and bronze, can also be used to increase the effectiveness of service-based SLAs.

Customer-based SLA

This is an agreement with an individual customer group, covering all the services they use. For example, agreements may be reached with an organization's finance department covering, say, the finance system, the accounting system, the payroll system, the billing system, the procurement system, and any other IT systems that they use. Customers often prefer such an agreement, as all of their requirements are covered in a single document. Only one signatory is normally required, which simplifies this issue.

Hints and tips

A combination of either of these structures might be appropriate, providing all services and customers are covered, with no overlap or duplication.

Multi-level SLAs

Some organizations have chosen to adopt a multi-level SLA structure. For example, a three-layer structure as follows:

- **Corporate level:** covering all the generic SLM issues appropriate to every customer throughout the organization. These issues are likely to be less volatile, so updates are less frequently required

- **Customer level:** covering all SLM issues relevant to the particular customer group or business unit, regardless of the service being used
- **Service level:** covering all SLM issues relevant to the specific service, in relation to a specific customer group (one for each service covered by the SLA).

As shown in Figure 4.7, such a structure allows SLAs to be kept to a manageable size, avoids unnecessary duplication, and reduces the need for frequent updates. However, it does mean that extra effort is required to maintain the necessary relationships and links within the Service Catalogue and the CMS.

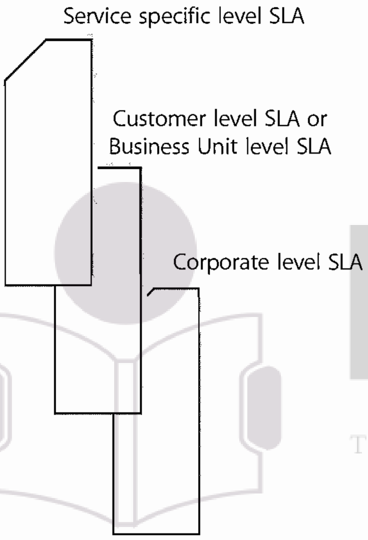


Figure 4.7 Multi-level SLAs

Many organizations have found it valuable to produce standards and a set of proformas or templates that can be used as a starting point for all SLAs, SLRs and OLAs. The proforma can often be developed alongside the draft SLA. Guidance on the items to be included in an SLA is given in Appendix F. Developing standards and templates will ensure that all agreements are developed in a consistent manner, and this will ease their subsequent use, operation and management.

Hints and tips

Make roles and responsibilities a part of the SLA. Consider three perspectives – the IT provider, the IT customer and the actual users.

The wording of SLAs should be clear and concise and leave no room for ambiguity. There is normally no need for agreements to be written in legal terminology, and plain language aids a common understanding. It is often helpful to have an independent person, who has not been involved with the drafting, to do a final read-through. This often throws up potential ambiguities and difficulties that can then be addressed and clarified. For this reason alone, it is recommended that all SLAs contain a glossary, defining any terms and providing clarity for any areas of ambiguity.

It is also worth remembering that SLAs may have to cover services offered internationally. In such cases the SLA may have to be translated into several languages. Remember also that an SLA drafted in a single language may have to be reviewed for suitability in several different parts of the world (i.e. a version drafted in Australia may have to be reviewed for suitability in the USA or the UK – and differences in terminology, style and culture must be taken into account).

Where the IT services are provided to another organization by an external service provider, sometimes the service targets are contained within a contract and at other times they are contained within an SLA or schedule attached to the contract. Whatever document is used, it is essential that the targets documented and agreed are clear, specific and unambiguous, as they will provide the basis of the relationship and the quality of service delivered.

4.2.5.2 Determine, document and agree requirements for new services and produce SLRs

This is one of the earliest activities within the Service Design stage of the Service Lifecycle. Once the Service Catalogue has been produced and the SLA structure has been agreed, a first SLR must be drafted. It is advisable to involve customers from the outset, but rather than going along with a blank sheet to start with, it may be better to produce a first outline draft of the performance targets and the management and operational requirements, as a starting point for more detailed and in-depth discussion. Be careful, though, not to go too far and appear to be presenting the customer with a 'fait accompli'.

It cannot be over-stressed how difficult this activity of determining the initial targets for inclusion with an SLR or SLA is. All of the other processes need to be consulted for their opinion on what are realistic targets that can be achieved, such as Incident Management on incident targets. The Capacity and Availability Management processes will be of particular value in determining appropriate service availability and performance targets. If there is any doubt, provisional targets should be included within a pilot SLA that is monitored and adjusted through a service warranty period, as illustrated in Figure 3.5.

While many organizations have to give initial priority to introducing SLAs for existing services, it is also important to establish procedures for agreeing Service Level Requirements (SLRs) for new services being developed or procured.

The SLRs should be an integral part of the Service Design criteria, of which the functional specification is a part. They should, from the very start, form part of the testing/trialling criteria as the service progresses through the stages of design and development or procurement. This SLR will gradually be refined as the service progresses through the stages of its lifecycle, until it eventually becomes a pilot SLA during the early life support period. This pilot or draft SLA should be developed alongside the service itself, and should be signed and formalized before the service is introduced into live use.

It can be difficult to draw out requirements, as the business may not know what they want – especially if not asked previously – and they may need help in understanding and defining their needs, particularly in terms of capacity, security, availability and IT service continuity. Be aware that the requirements initially expressed may not be those ultimately agreed. Several iterations of negotiations may be required before an affordable balance is struck between what is sought and what is achievable and affordable. This process may involve a redesign of the service solution each time.

If new services are to be introduced in a seamless way into the live environment, another area that requires attention is the planning and formalization of the support arrangements for the service and its components. Advice should be sought from Change Management and Configuration Management to ensure the planning is comprehensive and covers the implementation, deployment and support of the service and its components. Specific responsibilities need to be defined and added to existing contracts/OLAs, or new ones need to be agreed. The support arrangements and all escalation routes also need adding to the CMS, including the Service Catalogue where appropriate, so that the Service Desk and other support staff are aware of them. Where appropriate, initial training and familiarization for the Service Desk and other support groups and knowledge transfer should be completed before live support is needed.

It should be noted that additional support resources (i.e. more staff) may be needed to support new services. There is often an expectation that an already overworked support group can magically cope with the additional effort imposed by a new service.

Using the draft agreement as a basis, negotiations must be held with the customer(s), or customer representatives to finalize the contents of the SLA and the initial service level targets, and with the service providers to ensure that these are achievable.

4.2.5.3 Monitor service performance against SLA

Nothing should be included in an SLA unless it can be effectively monitored and measured at a commonly agreed point. The importance of this cannot be overstressed, as inclusion of items that cannot be effectively monitored almost always results in disputes and eventual loss of faith in the SLM process. A lot of organizations have discovered this the hard way and as a result have absorbed heavy costs, both in a financial sense as well as in terms of negative impacts on their credibility.

Anecdote

A global network provider agreed availability targets for the provision of a managed network service. These availability targets were agreed at the point where the service entered the customer's premises. However, the global network provider could only monitor and measure availability at the point the connection left its premises. The network links were provided by a number of different national telecommunications service providers, with widely varying availability levels. The result was a complete mismatch between the availability figures produced by the network provider and the customer, with correspondingly prolonged and heated debate and argument.

Existing monitoring capabilities should be reviewed and upgraded as necessary. Ideally this should be done ahead of, or in parallel with, the drafting of SLAs, so that monitoring can be in place to assist with the validation of proposed targets.

It is essential that monitoring matches the customer's true perception of the service. Unfortunately this is often very difficult to achieve. For example, monitoring of individual components, such as the network or server, does not guarantee that the service will be available so far as the customer is concerned. Customer perception is often that although a failure might affect more than one service, all they are bothered about is the service they cannot access at the time of the reported incident – though this is not always true, so caution is needed. Without monitoring all components in the end-to-end service (which may be very difficult and costly to achieve) a true picture cannot be gained. Similarly, users must be aware that they should report incidents immediately to aid diagnostics, especially if they are performance-related, so that the service provider is aware that service targets are being breached.

A considerable number of organizations use their Service Desk, linked to a comprehensive CMS, to monitor the customer's perception of availability. This may involve making specific changes to incident/problem logging screens and may require stringent compliance with incident logging procedures. All of this needs discussion and agreement with the Availability Management process.

The Service Desk is also used to monitor incident response times and resolution times, but once again the logging screen may need amendment to accommodate data capture, and call-logging procedures may need tightening and must be strictly followed. If support is being provided by a third party, this monitoring may also underpin Supplier Management.

It is essential to ensure that any incident/problem-handling targets included in SLAs are the same as those included in Service Desk tools and used for escalation and monitoring purposes. Where organizations have failed to recognize this, and perhaps used defaults provided by the tool supplier, they have ended up in a situation where they are monitoring something different from that which has been agreed in the SLAs, and are therefore unable to say whether SLA targets have been met, without considerable effort to manipulate the data. Some amendments may be needed to support tools, to include the necessary fields so that relevant data can be captured.

Another notoriously difficult area to monitor is transaction response times (the time between sending a screen and receiving a response). Often end-to-end response times are technically very difficult to monitor. In such cases it may be appropriate to deal with this as follows:

- Include a statement in the SLA along the following lines: 'The services covered by the SLA are designed for high-speed response and no significant delays should be encountered. If a response time delay of more than x seconds is experienced for more than y minutes, this should be reported immediately to the Service Desk'.
- Agree and include in the SLA an acceptable target for the number of such incidents that can be tolerated in the reporting period.
- Create an incident category of 'poor response' (or similar) and ensure that any such incidents are logged accurately and that they are related to the appropriate service.
- Produce regular reports of occasions where SLA transaction response time targets have been breached, and instigate investigations via Problem Management to correct the situation.

This approach not only overcomes the technical difficulties of monitoring, but also ensures that incidents of poor response are reported at the time they occur. This is very important, as poor response is often caused by a number of transient interacting events that can only be detected if they are investigated immediately.

The preferred method, however, is to implement some form of automated client/server response time monitoring in close consultation with the Service Operation. Wherever possible, implement sampling or 'robot' tools and techniques to give indications of slow or poor performance. These tools provide the ability to measure or sample actual or very similar response times to those being experienced by a variety of users, and are becoming increasingly available and increasingly more cost-effective to use.

Hints and tips

Some organizations have found that, in reality, 'poor response' is sometimes a problem of user perception. The user, having become used to a particular level of response over a period of time, starts complaining as soon as this is slower. Take the view that 'if the user thinks the service is slow, then it is'.

If the SLA includes targets for assessing and implementing Requests for Change (RFCs), the monitoring of targets relating to Change Management should ideally be carried out using whatever Change Management tool is in use (preferably part of an integrated Service Management support tool) and change logging screens and escalation processes should support this.

4.2.5.4 Collate, measure and improve customer satisfaction

There are a number of important 'soft' issues that cannot be monitored by mechanistic or procedural means, such as customers' overall feelings (these need not necessarily match the 'hard' monitoring). For example, even when there have been a number of reported service failures, the customers may still feel positive about things, because they may feel satisfied that appropriate actions are being taken to improve things. Of course, the opposite may apply, and customers may feel dissatisfied with some issues (e.g. the manner of some staff on the Service Desk) when few or no SLA targets have been broken.

From the outset, it is wise to try and manage customers' expectations. This means setting proper expectations and appropriate targets in the first place, and putting a systematic process in place to manage expectations going forward, as satisfaction = perception – expectation (where a zero or positive score indicates a satisfied customer).

SLAs are just documents, and in themselves do not materially alter the quality of service being provided (though they may affect behaviour and help engender an appropriate service culture, which can have an immediate beneficial effect, and make longer-term improvements possible). A degree of patience is therefore needed and should be built into expectations.

Where charges are being made for the services provided, this should modify customer demands. (Customers can have whatever they can cost-justify – providing it fits within agreed corporate strategy – and have authorized budget for, but no more.) Where direct charges are not made, the support of senior business managers should be enlisted to ensure that excessive or unrealistic demands are not placed on the IT provider by any individual customer group.

It is therefore recommended that attempts be made to monitor customer perception on these soft issues. Methods of doing this include:

- Periodic questionnaires and customer surveys
- Customer feedback from service review meetings
- Feedback from Post Implementation Reviews (PIRs) conducted as part of the Change Management process on major changes, releases, new or changed services, etc.
- Telephone perception surveys (perhaps at random on the Service Desk, or using regular customer liaison representatives)
- Satisfaction survey handouts (left with customers following installations, service visits, etc.)
- User group or forum meetings
- Analysis of complaints and compliments.

Where possible, targets should be set for these and monitored as part of the SLA (e.g. an average score of 3.5 should be achieved by the service provider on results given, based on a scoring system of 1 to 5, where 1 is poor performance and 5 is excellent). Ensure that if users provide feedback they receive some return, and demonstrate to them that their comments have been incorporated in an action plan, perhaps a SIP. All customer satisfaction measurements should be reviewed, and where variations are identified, they should be analysed with action taken to rectify the variation.

4.2.5.5 Review and revise underpinning agreements and service scope

IT service providers are dependent to some extent on their own internal technical support teams or on external partners or suppliers. They cannot commit to meeting SLA targets unless their own support team's and suppliers' performances underpin these targets. Contracts with external suppliers are mandatory, but many organizations have also identified the benefits of having simple agreements with internal support groups, usually referred to as OLAs. 'Underpinning agreements' is a term used to refer to all underpinning OLAs, SLAs and contracts.

Often these agreements are referred to as 'back-to-back' agreements. This is to reflect the need to ensure that all targets within underpinning or 'back-to-back' agreements are aligned with, and support, targets agreed with the business in SLAs or OLAs. There may be several layers of these underpinning or 'back-to-back' agreements with aligned targets. It is essential that the targets at each layer are aligned with, and support, the targets contained within the higher levels (i.e. those closest to the business targets).

OLAs need not be very complicated, but should set out specific back-to-back targets for support groups that underpin the targets included in SLAs. For example, if the SLA includes overall time to respond and fix targets for incidents (varying on the priority levels), then the OLAs should include targets for each of the elements in the support chain. It must be understood, however, that the incident resolution targets included in SLAs should not normally match the same targets included in contracts or OLAs with suppliers. This is because the SLA targets must include an element for all stages in the support cycle (e.g. detection time, Service Desk logging time, escalation time, referral time between groups etc, Service Desk review and closure time – as well as the actual time fixing the failure).

The SLA target should cover the time taken to answer calls, escalate incidents to technical support staff, and the time taken to start to investigate and to resolve incidents assigned to them. In addition, overall support hours should be stipulated for all groups that underpin the required service availability times in the SLA. If special procedures exist for contacted staff (e.g. out-of-hours telephone support) these must also be documented.

OLAs should be monitored against OLA and SLA targets, and reports on achievements provided as feedback to the appropriate managers of each support team. This highlights potential problem areas, which may need to be addressed internally or by a further review of the SLA or OLA. Serious consideration should be given to introducing formal OLAs for all internal support teams, which contribute to the support of operational services.

Before committing to new or revised SLAs, it is therefore important that existing contractual arrangements are investigated and, where necessary, upgraded. This is likely to incur additional costs, which must either be absorbed by IT or passed on to the customer. In the latter case, the customer must agree to this, or the more relaxed targets in existing contracts should be agreed for inclusion in SLAs. This activity needs to be completed in close consultation with the Supplier Management process, to ensure not only that SLM requirements are met, but also that all other process requirements are considered, particularly supplier and contractual policies and standards.

4.2.5.6 *Produce service reports*

Immediately after the SLA is agreed and accepted, monitoring must be instigated, and service achievement reports must be produced. Operational reports must be produced frequently (weekly – perhaps even more frequently) and, where possible, exception reports should be produced whenever an SLA has been broken (or threatened, if appropriate thresholds have been set to give an 'early warning'). Sometimes difficulties are encountered in meeting the targets of new services during the early life support period because of the high volume of RFCs. Limiting the number of RFCs processed during the early life support period can limit the impact of changes.

The SLA reporting mechanisms, intervals and report formats must be defined and agreed with the customers. The frequency and format of Service Review Meetings must also be agreed with the customers. Regular intervals are recommended, with periodic reports synchronized with the reviewing cycle.

Periodic reports must be produced and circulated to customers (or their representatives) and appropriate IT managers a few days in advance of service level reviews, so that any queries or disagreements can be resolved ahead of the review meeting. The meeting is not then diverted by such issues.

The periodic reports should incorporate details of performance against all SLA targets, together with details of any trends or specific actions being undertaken to improve service quality. A useful technique is to include a SLA Monitoring (SLAM) chart at the front of a service report to give an 'at-a-glance' overview of how achievements have measured up against targets. These are most effective if colour coded (Red, Amber, Green, and sometimes referred to as RAG charts as a result). Other interim reports may be required by IT management for OLA or internal performance reviews and/or supplier or contract management. This is likely to be an evolving process – a first effort is unlikely to be the final outcome.

The resources required to produce and verify reports should not be underestimated. It can be extremely time-consuming, and if reports do not reflect the customer's own perception of service quality accurately, they can make the situation worse. It is essential that accurate information from all areas and all processes (e.g. Incident Management, Problem Management, Availability Management, Capacity Management, Change and Configuration Management) is analysed and collated into a concise and comprehensive report on service performance, as measured against agreed business targets.

SLM should identify the specific reporting needs and automate production of these reports, as far as possible. The extent, accuracy and ease with which automated reports can be produced should form part of the selection criteria for integrated support tools. These service reports should not only include details of current performance against targets, but should also provide historic information on past performance and trends, so that the impact of improvement actions can be measured and predicted.

4.2.5.7 *Conduct service reviews and instigate improvements within an overall SIP*

Periodic review meetings must be held on a regular basis with customers (or their representatives) to review the service achievement in the last period and to preview any issues for the coming period. It is normal to hold such meetings monthly or, as a minimum, quarterly.

Actions must be placed on the customer and provider as appropriate to improve weak areas where targets are not being met. All actions must be minuted, and progress should be reviewed at the next meeting to ensure that action items are being followed up and properly implemented.

Particular attention should be focused on each breach of service level to determine exactly what caused the loss of service and what can be done to prevent any recurrence. If it is decided that the service level was, or has become, unachievable, it may be necessary to review, renegotiate, review-agree different service targets. If the service break has been caused by a failure of a third-party or internal support group, it may also be necessary to review the underpinning agreement or OLA. Analysis of the cost and impact of service breaches provides valuable input and justification of SIP activities and actions. The constant need for improvement needs to be balanced and focused on those areas most likely to give the greatest business benefit.

Reports should also be produced on the progress and success of the SIP, such as the number of SIP actions that were completed and the number of actions that delivered their expected benefit.

Hints and tips

'A spy in both camps' – Service Level Managers can be viewed with a certain amount of suspicion by both the IT service provider staff and the customer representatives. This is due to the dual nature of the job, where they are acting as an unofficial customer representative when talking to IT staff, and as an IT provider representative when talking to the customers. This is usually aggravated when having to represent the 'opposition's' point of view in any meeting etc. To avoid this the Service Level Manager should be as open and helpful as possible (within the bounds of any commercial propriety) when dealing with both sides, although colleagues should never be openly criticized.

4.2.5.8 Review and revise SLAs, service scope and underpinning agreements

All agreements and underpinning agreements, including SLAs, underpinning contracts and OLAs, must be kept up-to-date. They should be brought under Change and Configuration Management control and reviewed periodically, at least annually, to ensure that they are still current and comprehensive, and are still aligned to business needs and strategy.

These reviews should ensure that the services covered and the targets for each are still relevant – and that nothing significant has changed that invalidates the agreement in any way (this should include infrastructure changes, business changes, supplier changes, etc.). Where changes are made, the agreements must be updated under Change Management control to reflect the new situation. If all agreements are recorded as CIs within the CMS, it is easier to assess the impact and implement the changes in a controlled manner.

These reviews should also include the overall strategy documents, to ensure that all services and service agreements are kept in line with business and IT strategies and policies.

4.2.5.9 Develop contacts and relationships

It is very important that SLM develops trust and respect with the business, especially with the key business contacts. Using the Service Catalogue, especially the Business Service Catalogue element of it, enables SLM to be much more proactive. The Service Catalogue provides the information that enables SLM to understand the relationships between the services and the business units and business process that depend on those services. It should also provide the information on all the key business and IT contacts relating to the services, their use and their importance. In order to ensure that this is done in a consistent manner, SLM should perform the following activities:

- Confirm stakeholders, customers and key business managers and service users.
- Assist with maintaining accurate information within the Service Portfolio and Service Catalogue.
- Be flexible and responsive to the needs of the business, customers and users, and understand current and planned new business processes and their requirements for new or changed services, documenting and communicating these requirements to all other processes as well as facilitating and innovating change wherever there is business benefit.
- Develop a full understanding of business, customer and user strategies, plans, business needs and objectives, ensuring that IT are working in partnership with the business, customers and users, developing long-term relationships.
- Regularly take the customer journey and sample the customer experience, providing feedback on customer issues to IT. (This applies to both IT customers and also the external business customers in their use of IT services).
- Ensure that the correct relationship processes are in place to achieve objectives and that they are subjected to continuous improvement.
- Conduct and complete customer surveys, assist with the analysis of the completed surveys and ensure that actions are taken on the results.
- Act as an IT representative on organizing and attending user groups.
- Proactively market and exploit the Service Portfolio and Service Catalogue and the use of the services within all areas of the business.
- Work with the business, customers and users to ensure that IT provides the most appropriate levels of service to meet business needs currently and in the future.
- Promote service awareness and understanding.
- Raise the awareness of the business benefits to be gained from the exploitation of new technology.
- Facilitate the development and negotiation of appropriate, achievable and realistic SLRs and SLAs between the business and IT.
- Ensure the business, customers and users understand their responsibilities/commitments to IT (i.e. IT dependencies).
- Assist with the maintenance of a register of all outstanding improvements and enhancements.

4.2.5.10 Complaints and compliments

The SLM process should also include activities and procedures for the logging and management of all complaints and compliments. The logging procedures are often performed by the Service Desk as they are similar to those of Incident Management and Request Fulfilment. The definition of a complaint and compliment should be agreed with the customers, together with agreed contact points and procedures for their management and analysis. All complaints and compliments should be recorded and communicated to the relevant parties. All complaints should also be actioned and resolved to the satisfaction of the originator. If not, there should be an escalation contact and procedure for all complaints that are not actioned and resolved within an appropriate timescale. All outstanding complaints should be reviewed and escalated to senior management where appropriate. Reports should also be produced on the numbers and types of complaints, the trends identified and actions taken to reduce the numbers received. Similar reports should also be produced for compliments.

4.2.6 Triggers, inputs, outputs and interfaces

There are many triggers that instigate SLM activity. These include:

- Changes in the Service Portfolio, such as new or changed business requirements or new or changed services
- New or changed agreements, SLRs, SLAs, OLAs or contracts
- Service review meetings and actions
- Service breaches or threatened breaches
- Compliments and complaints
- Periodic activities such as reviewing, reporting and customer satisfaction surveys
- Changes in strategy or policy.

4.2.6.1 SLM process inputs

There are a number of sources of information that are relevant to the Service Level Management process. These should include:

- Business information: from the organization's business strategy, plans, and financial plans and information on their current and future requirements
- Business Impact Analysis: providing information on the impact, priority, risk and number of users associated with each service
- Business requirements: details of any agreed, new or changed business requirements
- The strategies, policies and constraints from Service Strategy
- The Service Portfolio and Service Catalogue
- Change information: from the Change Management process with a forward schedule of changes and a need to assess all changes for their impact on all services
- CMS: containing information on the relationships between the business services, the supporting services and the technology
- Customer and user feedback, complaints and compliments
- Other inputs: including advice, information and input from any of the other processes (e.g. Incident Management, Capacity Management and Availability Management), together with the existing SLAs, SLRs, and OLAs and past service reports on the quality of service delivered.

4.2.6.2 SLM process outputs

The outputs of Service Level Management should include:

- Service reports: providing details of the service levels achieved in relation to the targets contained within SLAs. These reports should contain details of all aspects of the service and its delivery, including current and historical performance, breaches and weaknesses, major events, changes planned, current and predicted workloads, customer feedback, and improvement plans and activities
- Service Improvement Plan (SIP): an overall programme or plan of prioritized improvement actions, encompassing all services and all processes, together with associated impacts and risks
- The Service Quality Plan: documenting and planning the overall improvement of service quality
- Document templates: standard document templates, format and content for SLAs, SLRs and OLAs, aligned with corporate standards
- Service Level Agreements (SLAs): a set of targets and responsibilities should be documented and agreed within an SLA for each operational service
- Service Level Requirements (SLRs): a set of targets and responsibilities should be documented and agreed within an SLR for each proposed new or changed service
- Operational Level Agreements (OLAs): a set of targets and responsibilities should be documented and agreed within an OLA for each internal support team
- Reports on OLAs and underpinning contracts

- Service review meeting minutes and actions: all meetings should be scheduled on a regular basis, with planned agendas and their discussions and actions recorded and progressed
- SLA review and service scope review meeting minutes: summarizing agreed actions and revisions to SLAs and service scope
- Revised contracts: changes to SLAs or new SLRs may require existing underpinning contracts to be changed, or new contracts to be negotiated and agreed.

4.2.7 Key Performance Indicators

Key Performance Indicators (KPIs) and metrics can be used to judge the efficiency and effectiveness of the SLM activities and the progress of the SIP. These metrics should be developed from the service, customer and business perspective and should cover both subjective and objective measurements such as the following.

Objective:

- Number or percentage of service targets being met
- Number and severity of service breaches
- Number of services with up-to-date SLAs
- Number of services with timely reports and active service reviews.

Subjective:

- Improvements in customer satisfaction.

More information on KPIs, measurements and improvements can be found in the following section and in the Continuous Service Improvement publication.

Hints and tips

Don't fall into the trap of using percentages as the only metric. It is easy to get caught out when there is a small system with limited measurement points (i.e. a single failure in a population of 100 is only 1%; a single failure in a population of 50 is 2% – if the target is 98.5%, then the SLA is already breached). Always go for number of incidents rather than a percentage on populations of less than 100, and be careful when targets are accepted. This is something organizations have learned the hard way.

The SLM process often generates a good starting point for a SIP – and the service review process may drive this, but all processes and all areas of the service provider organization should be involved in the SIP.

Where an underlying difficulty has been identified that is adversely impacting on service quality, SLM must, in conjunction with Problem Management and Availability Management, instigate a SIP to identify and implement whatever actions are necessary to overcome the difficulties and restore service quality. SIP initiatives may also focus on such issues as user training, service and system testing and documentation. In these cases, the relevant people need to be involved and adequate feedback given to make improvements for the future. At any time, a number of separate initiatives that form part of the SIP may be running in parallel to address difficulties with a number of services.

Some organizations have established an up-front annual budget held by SLM from which SIP initiatives can be funded. This means that action can be undertaken quickly and that SLM is demonstrably effective. This practice should be encouraged and expanded to enable SLM to become increasingly proactive and predictive. The SIP needs to be owned and managed, with all improvement actions being assessed for risk and impact on services, customers and the business, and then prioritized, scheduled and implemented.

If an organization is outsourcing its Service Delivery to a third party, the issue of service improvement should be discussed at the outset and covered (and budgeted for) in the contract, otherwise there is no incentive during the lifetime of the contract for the supplier to improve service targets if they are already meeting contractual obligations and additional expenditure is needed to make the improvements.

4.2.7.1 KPIs

Manage the overall quality of IT service needed, both in the number and level of services provided and managed:

- Percentage reduction in SLA targets missed
- Percentage reduction in SLA targets threatened
- Percentage increase in customer perception and satisfaction of SLA achievements, via service reviews and Customer Satisfaction Survey responses
- Percentage reduction in SLA breaches caused because of third-party support contracts (underpinning contracts)
- Percentage reduction in SLA breaches caused because of internal Operational Level Agreements (OLAs).

Deliver service as previously agreed at affordable costs:

- Total number and percentage increase in fully documented SLAs in place
- Percentage increase in SLAs agreed against operational services being run
- Percentage reduction in the costs associated with service provision

- Percentage reduction in the cost of monitoring and reporting of SLAs
- Percentage increase in the speed and of developing and agreeing appropriate SLAs
- Frequency of service review meetings.

Manage business interface:

- Increased percentage of services covered by SLAs
- Documented and agreed SLM processes and procedures are in place
- Reduction in the time taken to respond to and implement SLA requests
- Increased percentage of SLA reviews completed on time
- Reduction in the percentage of outstanding SLAs for annual renegotiation
- Reduction in the percentage of SLAs requiring corrective changes (for example, targets not attainable; changes in usage levels). Care needs to be taken when using this KPI
- Percentage increase in the coverage of OLAs and third-party contracts in place, whilst possibly reducing the actual number of agreements (consolidation and centralization)
- Documentary evidence that issues raised at service and SLA reviews are being followed up and resolved
- Reduction in the number and severity of SLA breaches
- Effective review and follow-up of all SLA, OLA and underpinning contract breaches.

4.2.8 Information Management

SLM provides key information on all operational services, their expected targets and the service achievements and breaches for all operational services. It assists Service Catalogue Management with the management of the Service Catalogue and also provides the information and trends on customer satisfaction, including complaints and compliments.

SLM is crucial in providing information on the quality of IT service provided to the customer, and information on the customer's expectation and perception of that quality of service. This information should be widely available to all areas of the service provider organization.

4.2.9 Challenges, Critical Success Factors and risks

One challenge faced by SLM is that of identifying suitable customer representatives with whom to negotiate. Who 'owns' the service? In some cases, this may be obvious, and a single customer manager is willing to act as the signatory to the agreement. In other cases, it may take quite a bit of negotiating or cajoling to find a representative 'volunteer' (beware that volunteers often want to express their own personal view rather than represent a general consensus), or it may be necessary to get all customers to sign.

If customer representatives exist who are able genuinely to represent the views of the customer community, because they frequently meet with a wide selection of customers, this is ideal. Unfortunately, all too often representatives are head-office based and seldom come into contact with genuine service customers. In the worst case, SLM may have to perform his/her own programme of discussions and meetings with customers to ensure true requirements are identified.

Anecdote

On negotiating the current and support hours for a large service, an organization found a discrepancy in the required time of usage between Head Office and the field office's customers. Head Office (with a limited user population) wanted service hours covering 8.00 to 18.00, whereas the field office (with at least 20 times the user population) stated that starting an hour earlier would be better – but all offices closed to the public by 16.00 at the latest, and so wouldn't require a service much beyond this. Head Office won the 'political' argument, and so the 8.00 to 18.00 band was set. When the service came to be used (and hence monitored) it was found that service extensions were usually asked for by the field office to cover the extra hour in the morning, and actual usage figures showed that the service had not been accessed after 17.00, except on very rare occasions. The Service Level Manager was blamed by the IT staff for having to cover a late shift, and by the customer representative for charging for a service that was not used (i.e. staff and running costs).

Hints and tips

Care should be taken when opening discussions on service levels for the first time, as it is likely that 'current issues' (the failure that occurred yesterday) or long-standing grievances (that old printer that we have been trying to get replaced for ages) are likely to be aired at the outset. Important though these may be, they must not be allowed to get in the way of establishing the longer-term requirements. Be aware, however, that it may be necessary to address any issues raised at the outset before gaining any credibility to progress further.

If there has been no previous experience of SLM, then it is advisable to start with a draft SLA. A decision should be made on which service or customers are to be used for the draft. It is helpful if the selected customer is enthusiastic and wishes to participate – perhaps because they are anxious to see improvements in service quality. The results of an initial customer perception survey may give pointers to a suitable initial draft SLA.

> **Hints and tips**
>
> Don't pick an area where large problems exist as the first SLA. Try to pick an area that is likely to show some quick benefits and develop the SLM process. Nothing encourages acceptance of a new idea quicker than success.

One difficulty sometimes encountered is that staff at different levels within the customer community may have different objectives and perceptions. For example, a senior manager may rarely use a service and may be more interested in issues such as value for money and output, whereas a junior member of staff may use the service throughout the day, and may be more interested in issues such as responsiveness, usability and reliability. It is important that all of the appropriate and relevant customer requirements, at all levels, are identified and incorporated in SLAs.

Some organizations have formed focus groups from different levels from within the customer community to help ensure that all issues have been correctly addressed. This takes additional resources, but can be well worth the effort.

The other group of people that has to be consulted during the whole of this process is the appropriate representatives from within the IT provider side (whether internal or from an external supplier or partner). They need to agree that targets are realistic, achievable and affordable. If they are not, further negotiations are needed until a compromise acceptable to all parties is agreed. The views of suppliers should also be sought, and any contractual implications should be taken into account during the negotiation stages.

Where no past monitored data is available, it is advisable to leave the agreement in draft format for an initial period, until monitoring can confirm that initial targets are achievable. Targets may have to be re-negotiated in some cases. Many organizations negotiate an agreed timeframe for IT to negotiate and create a baseline for establishing realistic service targets. When targets and timeframes have been confirmed, the SLAs must be signed.

Once the initial SLA has been completed, and any early difficulties overcome, then move on and gradually introduce SLAs for other services/customers. If it is decided from the outset to go for a multi-level structure, it is likely that the corporate-level issues have to be covered for all customers at the time of the initial SLA. It is also worth trialling the corporate issues during this initial phase.

> **Hints and tips**
>
> Don't go for easy targets at the corporate level. They may be easy to achieve, but have no value in improving service quality or credibility. Also, if the targets are set at a sufficiently high level, the corporate SLA can be used as the standard that all new services should reach.

One point to ensure is that at the end of the drafting and negotiating process, the SLA is actually signed by the appropriate managers on the customer and IT service provider sides to the agreement. This gives a firm commitment by both parties that every attempt will be made to meet the agreement. Generally speaking, the more senior the signatories are within their respective organizations, the stronger the message of commitment. Once an SLA is agreed, wide publicity needs to be used to ensure that customers, users and IT staff alike are aware of its existence and of the key targets.

Steps must be taken to advertise the existence of the new SLAs and OLAs amongst the Service Desk and other support groups, with details of when they become operational. It may be helpful to extract key targets from these agreements into tables that can be on display in support areas, so that staff are always aware of the targets to which they are working. If support tools allow, these targets should be recorded within the tools, such as within a Service Catalogue or CMS, so that their content can be made widely available to all personnel. They should also be included as thresholds, and automatically alerted against when a target is threatened or actually breached. SLAs, OLAs and the targets they contain must also be publicized amongst the user community, so that users are aware of what they can expect from the services they use, and know at what point to start expressing dissatisfaction.

It is important that the Service Desk staff are committed to the SLM process, and become proactive ambassadors for SLAs, embracing the necessary service culture, as they are the first contact point for customers' incidents, complaints and queries. If the Service Desk staff are not fully aware of SLAs in place, and do not act on their contents, customers very quickly lose faith in SLAs.

4.2.9.1 Critical Success Factors

The main Critical Success Factors for the Service Catalogue Management process are:

- Manage the overall quality of IT services required
- Deliver the service as previously agreed at affordable costs
- Manage the interface with the business and users.

The risks associated with regard to the provision of an accurate Service Catalogue are:

- A lack of accurate input, involvement and commitment from the business and customers
- The tools and resources required to agree, document, monitor, report and review agreements and service levels
- The process becomes a bureaucratic, administrative process rather than an active and proactive process delivering measurable benefit to the business
- Access to and support of appropriate and up-to-date CMS and SKMS
- Bypassing the use of the SLM processes
- Business and customer measurements are too difficult to measure and improve, so are not recorded
- Inappropriate business and customer contacts and relationships are developed
- High customer expectations and low perception
- Poor and inappropriate communication is achieved with the business and customers.

4.3 CAPACITY MANAGEMENT

Capacity Management is a process that extends across the Service Lifecycle. A key success factor in managing capacity is ensuring it is considered during the design stage. It is for this reason that the Capacity Management process is included in this publication. Capacity Management is supported initially in Service Strategy where the decisions and analysis of business requirements and customer outcomes influence the development of patterns of business activity (PBA), levels of service (LOS) and service level packages (SLPs). This provides the predictive and ongoing capacity indicators needed to align capacity to demand.

4.3.1 Purpose/goal/objective

> 'The goal of the Capacity Management process is to ensure that cost-justifiable IT capacity in all areas of IT always exists and is matched to the current and future agreed needs of the business, in a timely manner'.

The purpose of Capacity Management is to provide a point of focus and management for all capacity- and performance-related issues, relating to both services and resources.

The objectives of Capacity Management are to:

- Produce and maintain an appropriate and up-to-date Capacity Plan, which reflects the current and future needs of the business
- Provide advice and guidance to all other areas of the business and IT on all capacity- and performance-related issues
- Ensure that service performance achievements meet or exceed all of their agreed performance targets, by managing the performance and capacity of both services and resources
- Assist with the diagnosis and resolution of performance- and capacity-related incidents and problems
- Assess the impact of all changes on the Capacity Plan, and the performance and capacity of all services and resources
- Ensure that proactive measures to improve the performance of services are implemented wherever it is cost-justifiable to do so.

4.3.2 Scope

The Capacity Management process should be the focal point for all IT performance and capacity issues. Technology management functions such as Network Support, Server Support or Operation Management may carry out the bulk of the day-to-day operational duties, but will provide performance information to the Capacity Management process. The process should encompass all areas of technology, both hardware and software, for all IT technology components and environments. Capacity Management should also consider space planning and environmental systems capacity as well as certain aspects of human resources, but only where a lack of human resources could result in a breach of SLA or OLA targets, a delay in the end-to-end performance or service response time, or an inability to meet future commitments and plans (e.g. overnight data backups not completed in time because no operators were present to load tapes).

In general, human resource management is a line management responsibility, though the staffing of a Service Desk should use identical Capacity Management techniques. The scheduling of human resources, staffing levels, skill levels and capability levels should therefore be included within the scope of Capacity Management. The driving force for Capacity Management should be the business requirements of the organization and to plan the resources needed to provide service levels in line with SLAs and OLAs. Capacity Management needs to understand the total IT and business environments, including:

- The current business operation and its requirements, through the patterns of business activity
- The future business plans and requirements via the Service Portfolio
- The service targets and the current IT service operation though SLAs and Standard Operating Procedures
- All areas of IT technology and its capacity and performance, including infrastructure, data, environment and applications.

Understanding all of this will enable Capacity Management to ensure that all the current and future capacity and performance aspects of services are provided cost-effectively.

Capacity Management is also about understanding the potential for the delivery of new services. New technology needs to be understood and, if appropriate, used to innovate and deliver the services required by the customer. Capacity Management needs to recognize that the rate of technological change will probably increase and that new technology should be harnessed to ensure that the IT services continue to satisfy changing business expectations. A direct link to the Service Strategy and Service Portfolio is needed to ensure that emerging technologies are considered in future service planning.

The Capacity Management process should include:

- Monitoring patterns of business activity and service-level plans through performance, utilization and throughput of IT services and the supporting infrastructure, environmental, data and applications components and the production of regular and ad hoc reports on service and component capacity and performance
- Undertaking tuning activities to make the most efficient use of existing IT resources
- Understanding the agreed current and future demands being made by the customer for IT resources, and producing forecasts for future requirements
- Influencing demand management, perhaps in conjunction with Financial Management
- Producing a Capacity Plan that enables the service provider to continue to provide services of the quality defined in SLAs and that covers a sufficient planning timeframe to meet future service levels required as defined in the Service Portfolio and SLRs
- Assistance with the identification and resolution of any incidents and problems associated with service or component performance
- The proactive improvement of service or component performance wherever it is cost-justifiable and meets the needs of the business.

Managing the capacity of large distributed IT infrastructures is a complex and demanding task, especially when the IT capacity and the financial investment required is ever-increasing. Therefore it makes even more sense to plan for growth. While the cost of the upgrade to an individual component in a distributed environment is usually less than the upgrade to a component in a mainframe environment, there are often many more components in the distributed environment that need to be upgraded. Also there could now be economies of scale, because the cost per individual component could be reduced when many components need to be purchased. Capacity Management should have input to the Service Portfolio and procurement process to ensure that the best deals with suppliers are negotiated.

Capacity Management provides the necessary information on current and planned resource utilization of individual components to enable organizations to decide, with confidence:

- Which components to upgrade (i.e. more memory, faster storage devices, faster processors, greater bandwidth)
- When to upgrade – ideally this is not too early, resulting in expensive over-capacity, nor too late, failing to take advantage of advances in new technology, resulting in bottle-necks, inconsistent performance and, ultimately, customer dissatisfaction and lost business opportunities
- How much the upgrade will cost – the forecasting and planning elements of Capacity Management feed into budgetary lifecycles, ensuring planned investment.

Many of the other processes are less effective if there is no input to them from the Capacity Management process. For example:

- Can the Change Management process properly assess the effect of any change on the available capacity?
- When a new service is implemented, can the SLM process be assured that the SLRs of the new service are achievable, and that the SLAs of existing services will not be affected?
- Can the Problem Management process properly diagnose the underlying cause of incidents caused by poor performance?

- Can the IT Service Continuity process accurately determine the capacity requirements of the key business processes?

Capacity Management is one of the forward-looking processes that, when properly carried out, can forecast business events and impacts often before they happen. Good Capacity Management ensures that there are no surprises with regard to service and component design and performance.

Capacity Management has a close, two-way relationship with the Service Strategy and planning processes within an organization. On a regular basis, the long-term strategy of an organization is encapsulated in an update of the business plans. The Service Strategy will reflect the business plans and strategy, which are developed from the organization's understanding of the external factors such as the competitive marketplace, economic outlook and legislation, and its internal capability in terms of manpower, delivery capability, etc. Often a shorter-term tactical plan, or business change plan is developed to implement the changes necessary in the short to medium term to progress the overall business plan and Service Strategy. Capacity Management needs to understand the short-, medium- and long-term plans of the business while providing information on the latest ideas, trends and technologies being developed by the suppliers of computing hardware and software.

The organization's business plans drive the specific IT Service Strategy, the contents of which Capacity Management needs to be familiar with, and to which Capacity Management needs to have had significant and ongoing input. The right level of capacity at the right time is critical. Service Strategy plans will be helpful to capacity planning by identifying the timing for acquiring and implementing new technologies, hardware and software.

4.3.3 Value to the business

Capacity Management is responsible for ensuring that IT resources are planned and scheduled to provide a consistent level of service that is matched to the current and future needs of the business, as agreed and documented within SLAs and OLAs. In conjunction with the business and their plans, Capacity Management provides a Capacity Plan that outlines the IT resources and funding needed to support the business plan, together with a cost justification of that expenditure.

4.3.4 Policies/principles/basic concepts

Capacity Management ensures that the capacity and performance of the IT services and systems matches the evolving agreed demands of the business in the most cost-effective and timely manner. Capacity Management is essentially a balancing act:

- **Balancing costs against resources needed:** the need to ensure that processing capacity that is purchased is cost-justifiable in terms of business need, and the need to make the most efficient use of those resources.
- **Balancing supply against demand:** the need to ensure that the available supply of IT processing power matches the demands made on it by the business, both now and in the future. It may also be necessary to manage or influence the demand for a particular resource.

Capacity Management processes and planning must be involved in all stages of the Service Lifecycle from Strategy and Design, through Transition and Operation to Improvement. From a strategic perspective, the Service Portfolio contains the IT resources and capabilities. The advent of Service Oriented Architecture, virtualization and the use of value networks in IT service provision are important factors in the management of capacity. The appropriate capacity and performance should be designed into services and components from the initial design stages. This will ensure not only that the performance of any new or changed service meets its expected targets, but also that all existing services continue to meet all of their targets. This is the basis of stable service provision.

The overall Capacity Management process is continually trying cost-effectively to match IT resources and capacity to the ever-changing needs and requirements of the business. This requires the tuning and optimization of the current resources and the effective estimation and planning of the future resources, as illustrated in Figure 4.8.

Capacity Management is an extremely technical, complex and demanding process, and in order to achieve results, it requires three supporting sub-processes.

One of the key activities of Capacity Management is to produce a plan that documents the current levels of resource utilization and service performance and, after consideration of the Service Strategy and plans, forecasts the future requirements for new IT resources to support the IT services that underpin the business activities. The plan should indicate clearly any assumptions made. It should also include any recommendations quantified in terms of resource required, cost, benefits, impact, etc.

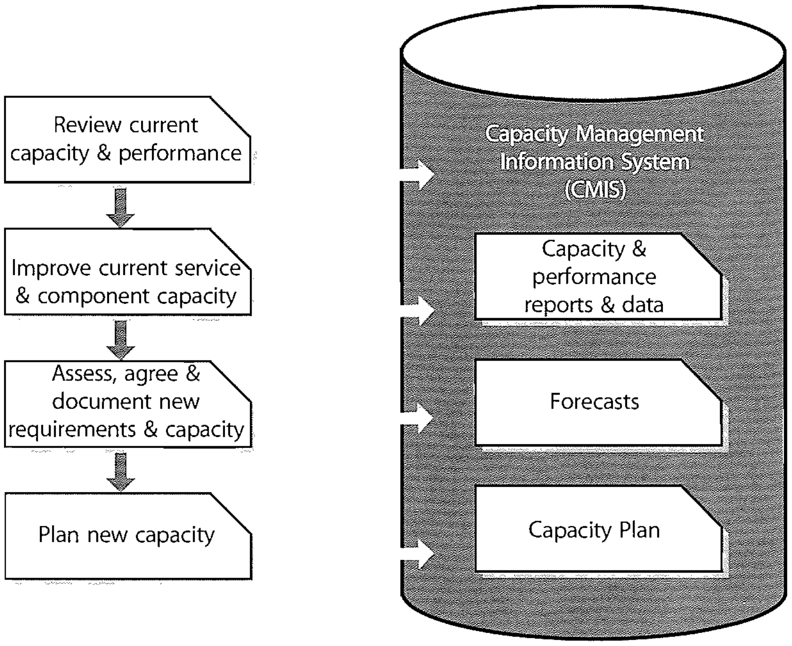

Review current capacity & performance
Improve current service & component capacity
Assess, agree & document new requirements & capacity
Plan new capacity
Capacity Management Information System (CMIS)
Capacity & performance reports & data
Forecasts
Capacity Plan

Figure 4.8 The Capacity Management process

The production and maintenance of a Capacity Plan should occur at pre-defined intervals. It is, essentially, an investment plan and should therefore be published annually, in line with the business or budget lifecycle, and completed before the start of negotiations on future budgets. A quarterly re-issue of the updated plan may be necessary to take into account changes in service plans, to report on the accuracy of forecasts and to make or refine recommendations. This takes extra effort but, if it is regularly updated, the Capacity Plan is more likely to be accurate and to reflect the changing business need.

The typical contents of a Capacity Plan are described in Appendix J.

4.3.4.1 Business Capacity Management

This sub-process translates business needs and plans into requirements for service and IT infrastructure, ensuring that the future business requirements for IT services are quantified, designed, planned and implemented in a timely fashion. This can be achieved by using the existing data on the current resource utilization by the various services and resources to trend, forecast, model or predict future requirements. These future requirements come from the Service Strategy and Service Portfolio detailing new processes and service requirements, changes, improvements, and also the growth in the existing services.

4.3.4.2 Service Capacity Management

The focus of this sub-process is the management, control and prediction of the end-to-end performance and capacity of the live, operational IT services usage and workloads. It ensures that the performance of all services, as detailed in service targets within SLAs and SLRs, is monitored and measured, and that the collected data is recorded, analysed and reported. Wherever necessary, proactive and reactive action should be instigated, to ensure that the performance of all services meets their agreed business targets. This is performed by staff with knowledge of all the areas of technology used in the delivery of end-to-end service, and often involves seeking advice from the specialists involved in Component Capacity Management. Wherever possible, automated thresholds should be used to manage all operational services, to ensure that situations where service targets are breached or threatened are rapidly identified and cost-effective actions to reduce or avoid their potential impact implemented.

4.3.4.3 Component Capacity Management

The focus in this sub-process is the management, control and prediction of the performance, utilization and capacity of individual IT technology components. It ensures that all components within the IT infrastructure that have finite resource are monitored and measured, and that the collected data is recorded, analysed and reported. Again, wherever possible, automated thresholds should be implemented to manage all components, to ensure that situations where service targets are breached or threatened by component usage or performance are rapidly identified, and cost-effective actions to reduce or avoid their potential impact are implemented.

There are many similar activities that are performed by each of the above sub-processes, but each sub-process has a very different focus. Business Capacity Management is focused on the current and future business requirements, while Service Capacity Management is focused on the delivery of the existing services that support the business, and Component Capacity Management is focused on the IT infrastructure that underpins service provision. The role that each of these sub-processes plays in the overall process and the use of management tools is illustrated in Figure 4.9.

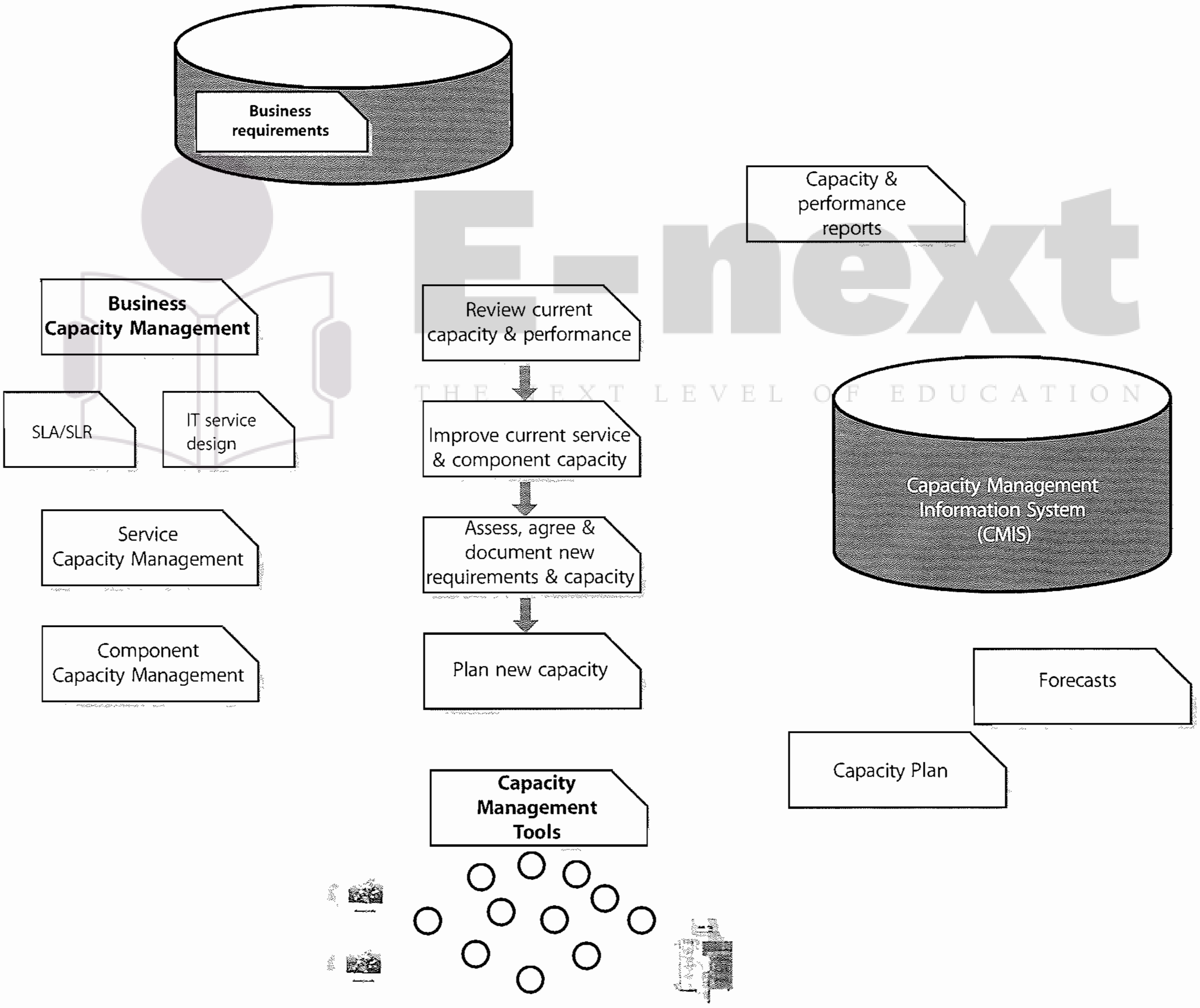

Business requirements
Capacity & performance reports
Business Capacity Management
Review current capacity & performance
SLA/SLR
IT service design
Improve current service & component capacity
Capacity Management Information System (CMIS)
Service Capacity Management
Assess, agree & document new requirements & capacity
Component Capacity Management
Plan new capacity
Forecasts
Capacity Plan
Capacity Management Tools

Figure 4.9 Capacity Management sub-processes

The tools used by Capacity Management need to conform to the organization's management architecture and integrate with other tools used for the management of IT systems and automating IT processes. The monitoring and control activities within Service Operation will provide a good basis for the tools to support and analyse information for Capacity Management.

4.3.5 Process activities, methods and techniques

Some activities in the Capacity Management process are reactive, while others are proactive. The proactive activities of Capacity Management should include:

- Pre-empting performance issues by taking the necessary actions before they occur
- Producing trends of the current component utilization and estimating the future requirements, using trends and thresholds for planning upgrades and enhancements
- Modelling and trending the predicted changes in IT services, and identifying the changes that need to be made to services and components of the IT infrastructure and applications to ensure that appropriate resource is available
- Ensuring that upgrades are budgeted, planned and implemented before SLAs and service targets are breached or performance issues occur
- Actively seeking to improve service performance wherever it is cost-justifiable
- Tuning and optimizing the performance of services and components.

The reactive activities of Capacity Management should include:

- Monitoring, measuring, reporting and reviewing the current performance of both services and components
- Responding to all capacity-related 'threshold' events and instigating corrective action
- Reacting to and assisting with specific performance issues. For example, the Service Desk may refer incidents of poor performance to Technology Management, which will employ Capacity Management techniques to resolve them.

Key message

The more successful the proactive and predictive activities of Capacity Management, the less need there will be for the reactive activities of Capacity Management.

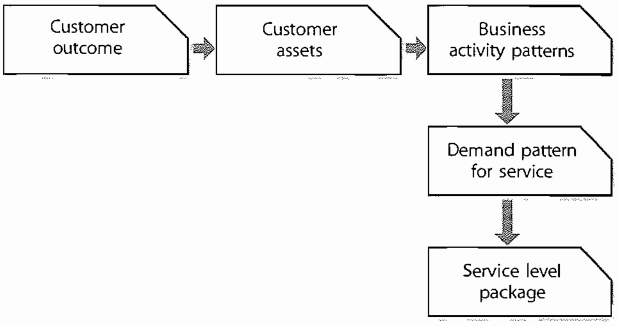


Figure 4.10 Capacity must support business requirements

4.3.5.1 Business Capacity Management

The main objective of the Business Capacity Management sub-process is to ensure that the future business requirements (customer outcomes) for IT services are considered and understood, and that sufficient IT capacity to support any new or changed services is planned and implemented within an appropriate timescale. Figure 4.10 illustrates that BCM is influenced by the business patterns of activity and how services are used.

The Capacity Management process must be responsive to changing requirements for capacity demand. New services or changed services will be required to underpin the changing business. Existing services will require modification to provide extra functionality. Old services will become obsolete, freeing up spare capacity. As a result, the ability to satisfy the customers' SLRs and SLAs will be affected. It is the responsibility of Capacity Management to predict the demand for capacity for such changes and manage the demand.

These new requirements may come to the attention of Capacity Management from many different sources and for many different reasons, but the principal sources of supply should be the Pattern of Business Activity from Demand Management and the Service Level Packages produced for the Service Portfolio. These indicate a window of future predictors for capacity. Such examples could be a recommendation to upgrade to take advantage of new technology, or the implementation of a tuning activity to resolve a performance problem. Figure 4.11 shows the cycle of demand management.

Capacity Management needs to be included in all strategic, planning and design activities, being involved as early as possible within each process, such as:

- Assisting and supporting the development of Service Strategy

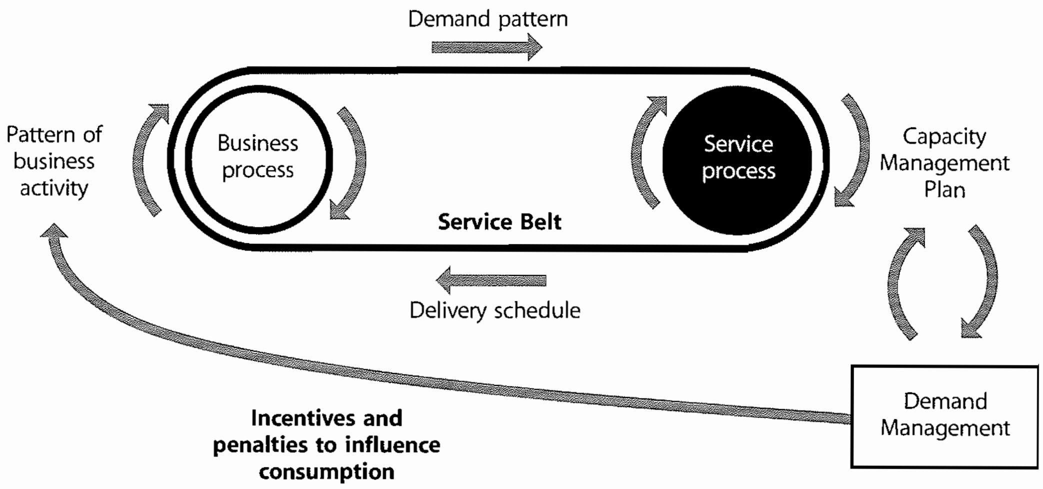


Figure 4.11 Capacity Management takes particular note of demand pattern

- Involvement in the review and improvement of IT strategies and policies
- Involvement in the review and improvement of technology architectures.

Key message

Capacity Management should not be a last-minute 'tick in the box' just prior to customer acceptance and operational acceptance.

If early involvement can be achieved from Capacity Management within these processes, then the planning and design of IT capacity can be closely aligned with business requirements and can ensure that service targets can be achieved and maintained.

Assist with agreeing Service Level Requirements

Capacity Management should assist SLM in understanding the customers' capacity and performance requirements, in terms of required service/system response times, expected throughput, patterns of usage and volume of users. Capacity Management should help in the negotiation process by providing possible solutions to a number of scenarios. For example, if the volume of users is less than 2,000, then response times can be guaranteed to be less than two seconds. If more than 2,000 users connect concurrently, then extra network bandwidth is needed to guarantee the required response time, or a slower response time will have to be accepted. Modelling, trending or application sizing techniques are often employed here to ensure that predictions accurately reflect the real situation.

Design, procure or amend service configuration

Capacity Management should be involved in the design of new or changing services and make recommendations for the procurement of hardware and software, where performance and/or capacity are factors. In some instances Capacity Management instigates the implementation of the new requirement through Change Management, where it is also involved as a member of the Change Advisory Board. In the interest of balancing cost and capacity, the Capacity Management process obtains the costs of alternative proposed solutions and recommends the most appropriate cost-effective solution.

Verify SLA

The SLA should include details of the anticipated service throughputs and the performance requirements. Capacity Management advises SLM on achievable targets that can be monitored and on which the Service Design has been based. Confidence that the Service Design will meet the SLRs and provide the ability for future growth can be gained by using modelling, trending or sizing techniques.

Support SLA negotiation

The results of the predictive techniques provide the verification of service performance capabilities. There may be a need for SLM to renegotiate SLAs based on these findings. Capacity Management provides support to SLM should renegotiations be necessary, by recommending potential solutions and associated cost information. Once assured that the requirements are achievable, it is the responsibility of SLM to agree the service levels and sign the SLA.

Control and implementation

All changes to service and resource capacity must follow all IT processes such as Change, Release, Configuration and Project Management to ensure that the right degree of control and coordination is in place on all changes and that any new or change components are recorded and tracked through their lifecycle.

4.3.5.2 Service Capacity Management

The main objective of the Service Capacity Management sub-process is to identify and understand the IT services, their use of resource, working patterns, peaks and troughs, and to ensure that the services meet their SLA targets, i.e. to ensure that the IT services perform as required. In this sub-process, the focus is on managing service performance, as determined by the targets contained in the agreed SLAs or SLRs.

The Service Capacity Management sub-process ensures that the services meet the agreed capacity service targets. The monitored service provides data that can identify trends from which normal service levels can be established. By regular monitoring and comparison with these levels, exception conditions can be defined, identified and reported on. Therefore Capacity Management informs SLM of any service breaches or near misses.

There will be occasions when incidents and problems are referred to Capacity Management from other processes, or it is identified that a service could fail to meet its SLA targets. On some of these occasions, the cause of the potential failure may not be resolved by Component Capacity Management. For example, when the failure is analysed it may be found that there is no lack of capacity, or no individual component is over-utilized. However, if the design or coding of an application is inefficient, then the service performance may need to be managed, as well as individual hardware or software resources. Service Capacity Management should also be monitoring service workloads and transactions to ensure that they remain within agreed limitations and thresholds.

The key to successful Service Capacity Management is to forecast issues, wherever possible, by monitoring changes in performance and monitoring the impact of changes. So this is another sub-process that has to be proactive and predictive, even pre-emptive, rather than reactive. However, there are times when it has to react to specific performance problems. From a knowledge and understanding of the performance requirements of each of the services being used, the effects of changes in the use of services can be estimated, and actions taken to ensure that the required service performance can be achieved.

4.3.5.3 Component Capacity Management

The main objective of Component Capacity Management (CCM) is to identify and understand the performance, capacity and utilization of each of the individual components within the technology used to support the IT services, including the infrastructure, environment, data and applications. This ensures the optimum use of the current hardware and software resources in order to achieve and maintain the agreed service levels. All hardware components and many software components in the IT infrastructure have a finite capacity that, when approached or exceeded, has the potential to cause performance problems.

This sub-process is concerned with components such as processors, memory, disks, network bandwidth, network connections etc. So information on resource utilization needs to be collected on a continuous basis. Monitors should be installed on the individual hardware and software components, and then configured to collect the necessary data, which is accumulated and stored over a period of time. This is an activity generally carried out through monitoring and control within Service Operation. A direct feedback to CCM should be applied within this sub-process.

As in Service Capacity Management, the key to successful CCM is to forecast issues, wherever possible, and it therefore has to be proactive and predictive as well. However, there are times when CCM has to react to specific problems that are caused by a lack of capacity, or the inefficient use of the component. From a knowledge and understanding of the use of resource by each of the services being run, the effects of changes in the use of services can be estimated and hardware or software upgrades can be budgeted and planned. Alternatively, services can be balanced across the existing resources to make most effective use of the current resources.

4.3.5.4 The underpinning activities of Capacity Management

The activities described in this section are necessary to support the sub-processes of Capacity Management, and these activities can be done both reactively or proactively, or even pre-emptively.

The major difference between the sub-processes is in the data that is being monitored and collected, and the perspective from which it is analysed. For example, the level of utilization of individual components in the infrastructure – such as processors, disks, and network links – is of interest in Component Capacity Management, while the transaction throughput rates and response times are of interest in Service Capacity Management. For Business Capacity Management, the transaction

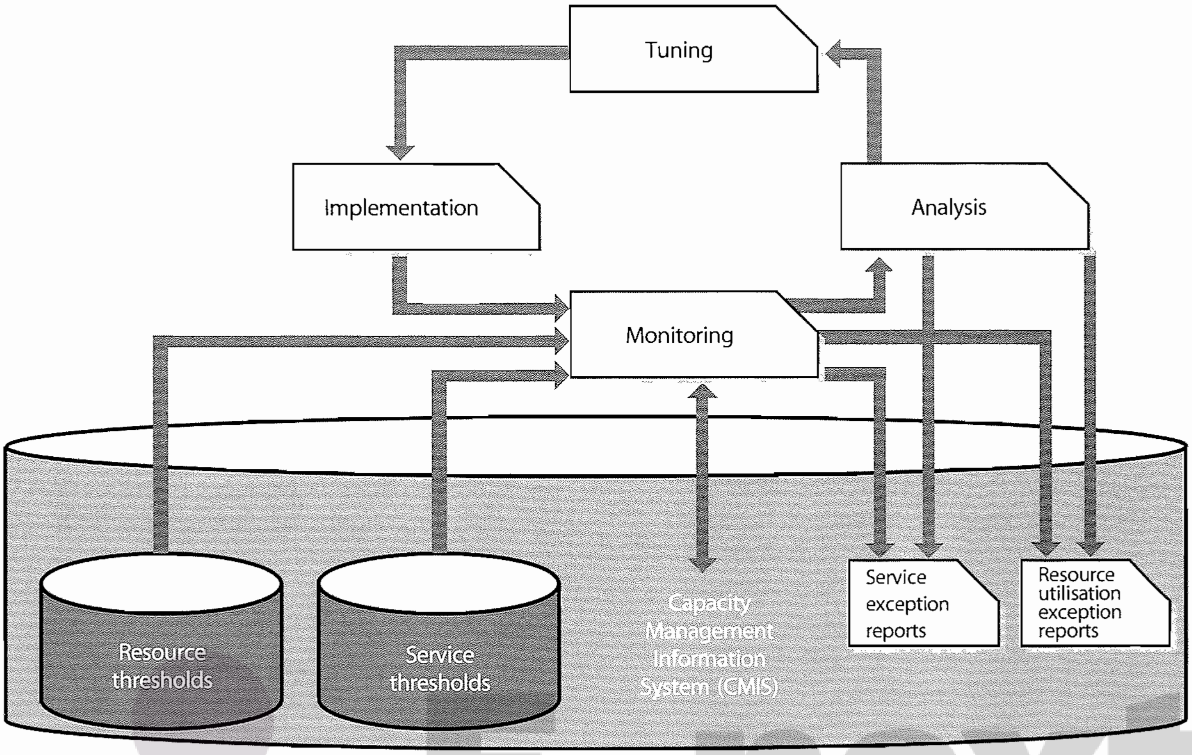

Tuning
Implementation
Analysis
Monitoring
Service exception reports
Resource utilisation exception reports
Resource thresholds
Service thresholds
Capacity Management Information System (CMIS)

Figure 4.12 Iterative ongoing activities of Capacity Management

throughput rates for the online service need to be translated into business volumes – for example, in terms of sales invoices raised or orders taken. The biggest challenge facing Capacity Management is to understand the relationship between the demands and requirements of the business and the business workload, and to be able to translate these in terms of the impact and effect of these on the service and resource workloads and utilizations, so that appropriate thresholds can be set at each level.

Tuning and optimization activities

A number of the activities need to be carried out iteratively and form a natural cycle, as illustrated in Figure 4.12.

These activities provide the basic historical information and triggers necessary for all of the other activities and processes within Capacity Management. Monitors should be established on all the components and for each of the services. The data should be analysed using, wherever possible, expert systems to compare usage levels against thresholds. The results of the analysis should be included in reports, and recommendations made as appropriate. Some form of control mechanism may then be put in place to act on the recommendations. This may take the form of balancing services, balancing workloads, changing concurrency levels and adding or removing resources. All of the information accumulated during these activities should be stored in the Capacity Management Information System (CMIS) and the cycle then begins again, monitoring any changes made to ensure they have had a beneficial effect and collecting more data for future actions.

Utilization monitoring

The monitors should be specific to particular operating systems, hardware configurations, applications, etc. It is important that the monitors can collect all the data required by the Capacity Management process, for a specific component or service. Typical monitored data includes:

- Processor utilization
- Memory utilization
- Per cent processor per transaction type
- IO rates (physical and buffer) and device utilization
- Queue lengths
- Disk utilization
- Transaction rates
- Response times

- Batch duration
- Database usage
- Index usage
- Hit rates
- Concurrent user numbers
- Network traffic rates.

In considering the data that needs to be included, a distinction needs to be drawn between the data collected to monitor capacity (e.g. throughput) and the data to monitor performance (e.g. response times). Data of both types is required by the Service and Component Capacity Management sub-processes. This monitoring and collection needs to incorporate all components in the service, thus monitoring the 'end-to-end' customer experience. The data should be gathered at total resource utilization level and at a more detailed profile for the load that each service places on each particular component. This needs to be carried out across the whole technology, host or server, the network, local server and client or workstation. Similarly the data needs to be collected for each service.

Part of the monitoring activity should be of thresholds and baselines or profiles of the normal operating levels. If these are exceeded, alarms should be raised and exception reports produced. These thresholds and baselines should have been determined from the analysis of previously recorded data, and can be set at both the component and service level. All thresholds should be set below the level at which the component or service is over-utilized, or below the targets in the SLAs. When the threshold is reached or threatened, there is still an opportunity to take corrective action before the SLA has been breached, or the resource has become over-utilised and there has been a period of poor performance. The monitoring and management of these events, thresholds and alarms is covered in detail in the Service Operation publication.

Often it is more difficult to get the data on the current business volumes as required by the Business Capacity Management sub-process. These statistics may need to be derived from the data available to the Service and Component Capacity Management sub-processes.

Response time monitoring

Many SLAs have user response times as one of the targets to be measured, but equally many organizations have great difficulty in supporting this requirement. User response times of IT and network services can be monitored and measured by the following:

- **Incorporating specific code within client and server applications software.** This can be used to provide complete 'end-to-end' service response times or intermediate timing points to break down the overall response into its constituent components. The figures obtained from these tools give the actual response times as perceived by the users of a service.
- **Using 'robotic scripted systems' with terminal emulation software.** These systems consist of client systems with terminal emulation software (e.g. browser or Telnet systems) and specialized scripted software for generating and measuring transactions and responses. These systems generally provide sample 'end-to-end' service response times and are useful for providing representative response times, particularly for multi-phase transactions or complex interactions. These only give sample response times, not the actual response times as perceived by the real users of the service.
- **Using distributed agent monitoring software.** Useful information on service response times can be obtained by distributing agent systems with monitoring software at different points of a network (e.g. within different countries on the internet). These systems can then be used to generate transactions from a number of locations and give periodic measurements of an internet site as perceived by international users of an internet website. However, again the times received are only indications of the response times and are not the real user response times.
- **Using specific passive monitoring systems.** Tracking a representative sample number of client systems. This method relies on the connection of specific network monitoring systems, often referred to as 'sniffers' being inserted at appropriate points within the network. These can then monitor, record and time all traffic passing a particular point within the network. Once recorded, this traffic can then be analysed to give detailed information on the service response times. Once again, however, these can only be used to give an approximation to the actual user response times, although these are often very close to the real-world situation, but this depends on the position of the monitor itself within the IT infrastructure.

In some cases, a combination of a number of systems may be used. The monitoring of response times is a complex process even if it is an in-house service running on a private network. If this is an external internet service, the process is much more complex because of the sheer number of different organizations and technologies involved.

Anecdote

A private company with a major website implemented a website monitoring service from an external supplier that would provide automatic alarms on the availability and responsiveness of their website. The availability and speed of the monitoring points were lower than those of the website being monitored. Therefore the figures produced by the service were of the availability and responsiveness of the monitoring service itself, rather than those of the monitored website.

Hints and tips

When implementing external monitoring services, ensure that the service levels and performance commitments of the monitoring service are in excess of those of the service(s) being monitored.

Analysis

The data collected from the monitoring should be analysed to identify trends from which the normal utilization and service levels, or baselines, can be established. By regular monitoring and comparison with this baseline, exception conditions in the utilization of individual components or service thresholds can be defined, and breaches or near misses in the SLAs can be reported and actioned. Also the data can be used to predict future resource usage, or to monitor actual business growth against predicted growth.

Analysis of the data may identify issues such as:

- 'Bottlenecks' or 'hot spots' within the infrastructure
- Inappropriate distribution of workload across available resources
- Inappropriate database indexing
- Inefficiencies in the application design
- Unexpected increase in workloads or transaction rates
- Inefficient scheduling or memory usage.

The use of each component and service needs to be considered over the short, medium and long term, and the minimum, maximum and average utilization for these periods recorded. Typically, the short-term pattern covers the utilization over a 24-hour period, while the medium term may cover a one- to four-week period, and the long term a year-long period. Over time, the trend in the use of the resource by the various IT services will become apparent. The usefulness of this information is further enhanced by recording any observed contributing factors to peaks or valleys in utilization – for example, if a change of business process or staffing coincides with any deviations from the normal utilization.

It is important to understand the utilization in each of these periods, so that changes in the use of any service can be related to predicted changes in the level of utilization of individual components. The ability to identify the specific hardware or software components on which a particular IT service depends is improved greatly by an accurate, up-to-date and comprehensive CMS.

When the utilization of a particular resource is considered, it is important to understand both the total level of utilization and the utilization by individual services of the resource.

Example

If a processor that is 75% loaded during the peak hour is being used by two different services, A and B, it is important to know how much of the total 75% is being used by each service. Assuming the system overhead on the processor is 5%, the remaining 70% load could be split evenly between the two services. If a change in either Service A or Service B is estimated to double its loading on the processor, then the processor would be overloaded.

However, if service A uses 60% and Service B uses 10% of the processor, then the processor would be overloaded if service A doubled its loading on the processor. But if service B doubled its loading on the processor, then the processor would not necessarily be overloaded.

Tuning

The analysis of the monitored data may identify areas of the configuration that could be tuned to better utilize the service, system and component resources or improve the performance of the particular service.

Tuning techniques that are of assistance include:

- Balancing workloads and traffic – transactions may arrive at the host or server at a particular gateway, depending on where the transaction was initiated; balancing the ratio of initiation points to gateways can provide tuning benefits
- Balancing disk traffic – storing data on disk efficiently and strategically, e.g. striping data across many spindles may reduce data contention
- Definition of an accepted locking strategy that specifies when locks are necessary and the appropriate level, e.g. database, page, file, record and row – delaying the lock until an update is necessary may provide benefits
- Efficient use of memory – may include looking to utilize more or less memory, depending on the circumstances.

Before implementing any of the recommendations arising from the tuning techniques, it may be appropriate to consider testing the validity of the recommendation. For example, 'Can Demand Management be used to avoid the need to carry out any tuning?' or 'Can the proposed change be modelled to show its effectiveness before it is implemented?'

Implementation

The objective of this activity is to introduce to the live operation services any changes that have been identified by the monitoring, analysis and tuning activities. The implementation of any changes arising from these activities must be undertaken through a strict, formal Change Management process. The impact of system tuning changes can have major implications on the customers of the service. The impact and risk associated with these types of changes are likely to be greater than that of other different type of changes.

It is important that further monitoring takes place, so that the effects of the change can be assessed. It may be necessary to make further changes or to regress some of the original changes.

Exploitation of new technology

This involves understanding new techniques and new technology and how they can be used to support the business and innovate improvements. It may be appropriate to introduce new technology to improve the provision and support of the IT services on which the organization is dependent. This information can be gathered by studying professional literature (magazine and press articles) and by attending:

- Promotional seminars by hardware and software suppliers
- User group meetings of suppliers of potential hardware and software
- User group meetings for other IT professionals involved in Capacity Management.

Each of these provides sources of information relating to potential techniques, technology, hardware and software, which might be advantageous for IT to implement to realize business benefits. However, at all times Capacity Management should recognize that the introduction and use of this new technology must be cost-justified and deliver real benefit to the business. It is not just the new technology itself that is important, but Capacity Management should also keep aware of the advantages to be obtained from the use of new technologies, using techniques such as 'grid computing', 'virtualization' and 'on-demand computing'.

Designing resilience

Capacity Management assists with the identification and improvement of the resilience within the IT infrastructure or any subset of it, wherever it is cost-justified. In conjunction with Availability Management, Capacity Management should use techniques such as Component Failure Impact Analysis (CFIA, as described in section 4.4 on Availability Management) to identify how susceptible the current configuration is to the failure or overload of individual components and make recommendations on any cost-effective solutions.

Capacity Management should be able to identify the impact on the available resources of particular failures, and the potential for running the most important services on the remaining resources. So the provision of spare capacity can act as resilience or fail-over in failure situations.

The requirements for resilience in the IT infrastructure should always be considered at the time of the service or system design. However, for many services, the resilience of the service is only considered after it is in live operational use. Incorporating resilience into Service Design is much more effective and efficient than trying to add it at a later date, once a service has become operational.

4.3.5.5 Threshold management and control

The technical limits and constraints on the individual services and components can be used by the monitoring activities to set the thresholds at which warnings and alarms are raised and exception reports are produced. However, care must be exercised when setting thresholds, because many thresholds are dependent on the work being run on the particular component.

The management and control of service and component thresholds is fundamental to the effective delivery of services to meet their agreed service levels. It ensures that all service and component thresholds are maintained at the appropriate levels and are continuously, automatically monitored, and alerts and warnings generated when breaches occur. Whenever monitored thresholds are breached or threatened, then alarms are raised and breaches, warnings and exception reports are produced. Analysis of the situation should then be completed and remedial action taken whenever justified, ensuring that the situation does not recur. The same data items can be used to identify when SLAs are breached or likely to be breached or when component performance degrades or is likely to be degraded. By setting thresholds below or above the actual targets, action can be taken and a breach of the SLA targets avoided. Threshold monitoring should not only alarm on exceeding a threshold, but should also monitor the rate of change and predict when the threshold will be reached. For example, a disk-space monitor should monitor the rate of growth and raise an alarm when the current rate will cause the disk to be full within the next N days. If a 1GB disk has reached 90% capacity, and is growing at 100KB per day, it will be 1,000 days before it is full. If it is growing at 10MB per day, it will only be 10 days before it is full. The monitoring and management of these events and alarms is covered in detail in the Service Operations publication.

There may be occasions when optimization of infrastructure components and resources is needed to maintain or improve performance or throughput. This can often be done through Workload Management, which is a generic term to cover such actions as:

- Rescheduling a particular service or workload to run at a different time of day or day of the week, etc. (usually away from peak times to off-peak windows) – which will often mean having to make adjustments to job-scheduling software
- Moving a service or workload from one location or set of CIs to another – often to balance utilization or traffic
- Technical 'virtualization': setting up and using virtualization techniques and systems to allow the movement of processing around the infrastructure to give better performance/resilience in a dynamic fashion
- Limiting or moving demand for components or resources through Demand Management techniques, in conjunction with Financial Management (see section 4.3.5.6).

It will only be possible to manage workloads effectively if a good understanding exists of which workloads will run at what time and how much resource utilization each workload places on the IT infrastructure. Diligent monitoring and analysis of workloads, together with a comprehensive CMIS, are therefore needed on an ongoing operational basis.

4.3.5.6 Demand Management

The prime objective of Demand Management is to influence user and customer demand for IT services and manage the impact on IT resources.

This activity can be carried out as a short-term requirement because there is insufficient current capacity to support the work being run, or, as a deliberate policy of IT management, to limit the required capacity in the long term.

Short-term Demand Management may occur when there has been a partial failure of a critical resource in the IT infrastructure. For example, if there has been a failure of a processor within a multi-processor server, it may not be possible to run the full range of services. However, a limited subset of the services could be run. Capacity Management should be aware of the business priority of each of the services, know the resource requirements of each service (in this case, the amount of processor power required to run the service) and then be able to identify which services can be run while there is a limited amount of processor power available.

Long-term Demand Management may be required when it is difficult to cost-justify an expensive upgrade. For example, many processors are heavily utilized for only a few hours each day, typically 10.00-12.00 and 14.00-16.00. Within these periods, the processor may be overloaded for only one or two hours. For the hours between 18.00-08.00, these processors are only very lightly loaded and the components are under-utilized. Is it possible to justify the cost of an upgrade to provide additional capacity for only a few hours in 24 hours? Or is it possible to influence the demand and spread the requirement for resource across 24 hours, thereby delaying or avoiding altogether the need for a costly upgrade?

Demand Management needs to understand which services are utilizing the resource and to what level, and the schedule of when they must be run. Then a decision can be made on whether it will be possible to influence the use of resource and, if so, which option is appropriate.

The influence on the services that are running could be exercised by:

- **Physical constraints:** for example, it may be possible to stop some services from being available at certain times, or to limit the number of customers who can use a particular service – for example, by limiting the number of concurrent users; the constraint could be implemented on a specific resource or component – for example, by limiting the number of physical connections to a network router or switch
- **Financial constraints:** if charging for IT services is in place, reduced rates could be offered for running work at times of the day when there is currently less demand for the resource. This is known as differential charging.

4.3.5.7 Modelling and trending

A prime objective of Capacity Management is to predict the behaviour of IT services under a given volume and variety of work. Modelling is an activity that can be used to beneficial effect in any of the sub-processes of Capacity Management.

The different types of modelling range from making estimates based on experience and current resource utilization information, to pilot studies, prototypes and full-scale benchmarks. The former is a cheap and reasonable approach for day-to-day small decisions, while the latter is expensive, but may be advisable when implementing a large new project or service. With all types of modelling, similar levels of accuracy can be obtained, but all are totally dependent on the skill of the person constructing the model and the information used to create it.

Baselining

The first stage in modelling is to create a baseline model that reflects accurately the performance that is being achieved. When this baseline model has been created, predictive modelling can be done, i.e. ask the 'What if?' questions that reflect failures, planned changes to the hardware and/or the volume/variety of workloads. If the baseline model is accurate, then the accuracy of the result of the potential failures and changes can be trusted.

Effective Capacity Management, together with modelling techniques, enables Capacity Management to answer the 'What if?' questions. What if the throughput of Service A doubles? What if Service B is moved from the current server onto a new server – what will be the effect on the response times of the two services?

Trend analysis

Trend analysis can be done on the resource utilization and service performance information that has been collected by the Capacity Management process. The data can be analysed in a spreadsheet, and the graphical and trending and forecasting facilities used to show the utilization of a particular resource over a previous period of time, and how it can be expected to change in the future.

Typically, trend analysis only provides estimates of future resource utilization information. Trend analysis is less effective in producing an accurate estimate of response times, in which case either analytical or simulation modelling should be used. Trend analysis is most effective when there is a linear relationship between a small number of variables, and less effective when there are non-linear relationships between variables or when there are many variables.

Analytical modelling

Analytical models are representations of the behaviour of computer systems using mathematical techniques, e.g. multi-class network queuing theory. Typically, a model is built using a software package on a PC, by specifying within the package the components and structure of the configuration that needs to be modelled, and the utilization of the components, e.g. processor, memory and disks, by the various workloads or applications. When the model is run, the queuing theory is used to calculate the response times in the computer system. If the response times predicted by the model are sufficiently close to the response times recorded in real life, the model can be regarded as an accurate representation of the computer system.

The technique of analytical modelling requires less time and effort than simulation modelling, but typically it gives less accurate results. Also, the model must be kept up-to-date. However, if the results are within 5% accuracy for utilization, and 15–20% for online application response times, the results are usually satisfactory.

Simulation modelling

Simulation involves the modelling of discrete events, e.g. transaction arrival rates, against a given hardware configuration. This type of modelling can be very accurate in sizing new applications or predicting the effects of changes on existing applications, but can also be very time-consuming and therefore costly.

When simulating transaction arrival rates, have a number of staff enter a series of transactions from prepared scripts, or use software to input the same scripted transactions with a random arrival rate. Either of these approaches takes time and effort to prepare and run. However, it can be cost-justified for organizations with very large services and systems where the major cost and the associated performance implications assume great importance.

4.3.5.8 Application sizing

Application sizing has a finite lifespan. It is initiated at the design stage for a new service, or when there is a major change to an existing service, and is completed when the application is accepted into the live operational environment. Sizing activities should include all areas of technology related to the applications, and not just the applications themselves. This should include the infrastructure, environment and data, and will often use modelling and trending techniques.

The primary objective of application sizing is to estimate the resource requirements to support a proposed change to an existing service or the implementation of a new service, to ensure that it meets its required service levels. To achieve this, application sizing has to be an integral part of the Service Lifecycle.

During the initial requirements and design, the required service levels must be specified in an SLR. This enables the Service Design and development to employ the pertinent technologies and products to achieve a design that meets the desired levels of service. It is much easier and less expensive to achieve the required service levels if Service Design considers the required service levels at the very beginning of the Service Lifecycle, rather than at some later stage.

Other considerations in application sizing are the resilience aspects that it may be necessary to build into the design of new services. Capacity Management is able to provide advice and guidance to the Availability Management process on the resources required to provide the required level of performance and resilience.

The sizing of the application should be refined as the design and development process progresses. The use of modelling can be used within the application sizing process.

The SLRs of the planned application developments should not be considered in isolation. The resources to be utilized by the application are likely to be shared with other services, and potential threats to existing SLA targets must be recognized and managed.

When purchasing software packages from external suppliers, it is just as important to understand the resource requirements needed to support the service. Often it can be difficult to obtain this information from the suppliers and it may vary, depending on throughput. Therefore, it is beneficial to identify similar customers of the product and to gain an understanding of the resource implications from them. It may be pertinent to benchmark, evaluate or trial the product prior to purchase.

Key message

Quality must be built in.

Some aspects of service quality can be improved after implementation (additional hardware can be added to improve performance, for example). Others – particularly aspects such as reliability and maintainability of applications software – rely on quality being 'built in', since to attempt to add it at a later stage is, in effect, redesign and redevelopment, normally at a much higher cost than the original development. Even in the hardware example quoted above, it is likely to cost more to add additional capacity after service implementation rather than as part of the original project.

4.3.6 Triggers, inputs, outputs and interfaces

There are many triggers that will initiate Capacity Management activities. These include:

- Service breaches, capacity or performance events and alerts, including threshold events
- Exception reports
- Periodic revision of current capacity and performance and the review of forecasts, reports and plans
- New and changed services requiring additional capacity
- Periodic trending and modelling
- Review and revision of business and IT plans and strategies
- Review and revision of designs and strategies
- Review and revision of SLAs, OLAs, contracts or any other agreements.

There are a number of sources of information that are relevant to the Capacity Management process. Some of these are as follows.

4.3.6.1 Inputs

- **Business information:** from the organization's business strategy, plans and financial plans, and information on their current and future requirements.
- **Service and IT information:** from Service Strategy, the IT strategy and plans and current budgets, covering all areas of technology and technology plans,

including the infrastructure, environment, data and applications, and the way in which they relate to business strategy and plans.

- **Component performance and capacity information:** of both existing and new technology, from manufacturers and suppliers.
- **Service performance issues:** the Incident and Problem Management processes, with incidents and problems relating to poor performance.
- **Service information:** from the SLM process, with details of the services from the Service Portfolio and the Service Catalogue and service level targets within SLAs and SLRs, and possibly from the monitoring of SLAs, service reviews and breaches of the SLAs.
- **Financial information:** from Financial Management, the cost of service provision, the cost of resources, components and upgrades, the resultant business benefit and the financial plans and budgets, together with the costs associated with service and component failure. Some of the costs of components and upgrades to components will be obtained from procurement, suppliers and manufacturers.
- **Change information:** from the Change Management process, with a Change Schedule and a need to assess all changes for their impact on the capacity of the technology.
- **Performance information:** from the Capacity Management Information System (CMIS) on the current performance of both all existing services and IT infrastructure components.
- **CMS:** containing information on the relationships between the business, the services, the supporting services and the technology.
- **Workload information:** from the IT Operations team, with schedules of all the work that needs to be run, and information on the dependencies between different services and information, and the interdependencies within a service.

4.3.6.2 Outputs

The outputs of Capacity Management are used within all other parts of the process, by many other processes and by other parts of the organization. Often this information is supplied as electronic reports or displays on shared areas, or as pages on intranet servers, to ensure the most up-to-date information is always used. The information provided is as follows:

- **The Capacity Management Information System (CMIS):** holds the information needed by all sub-processes within Capacity Management. For example, the data monitored and collected as part of Component and Service Capacity Management is used in Business Capacity Management to determine what infrastructure components or upgrades to components are needed, and when.
- **The Capacity Plan:** used by all areas of the business and IT management. and is acted on by the IT service provider and senior management of the organization to plan the capacity of the IT infrastructure. It also provides planning input to many other areas of IT and the business. It contains information on the current usage of service and components, and plans for the development of IT capacity to meet the needs in the growth of both existing service and any agreed new services. The Capacity Plan should be actively used as a basis for decision-making. Too often, Capacity Plans are created and never referred to or used.
- **Service performance information and reports:** used by many other processes. For example, the Capacity Management process assists Service Level Management with the reporting and reviewing of service performance and the development of new SLRs or changes to existing SLAs. It also assists the Financial Management process by identifying when money needs to be budgeted for IT infrastructure upgrades, or the purchase of new components.
- **Workload analysis and reports:** used by IT Operations to assess and implement changes in conjunction with Capacity Management to schedule or reschedule when services or workloads are run, to ensure that the most effective and efficient use is made of the available resources.
- **Ad hoc capacity and performance reports:** used by all areas of Capacity Management, IT and the business to analyse and resolve service and performance issues.
- **Forecasts and predictive reports:** used by all areas to analyse, predict and forecast particular business and IT scenarios and their potential solutions.
- **Thresholds, alerts and events.**

4.3.7 Key Performance Indicators

Some of the KPIs and metrics that can be used to judge the efficiency and effectiveness of the Capacity Management activities should include:

- Accurate business forecasts:
  - Production of workload forecasts on time
  - Percentage accuracy of forecasts of business trends
  - Timely incorporation of business plans into the Capacity Plan

  - Reduction in the number of variances from the business plans and Capacity Plans.
- Knowledge of current and future technologies:
  - Increased ability to monitor performance and throughput of all services and components
  - Timely justification and implementation of new technology in line with business requirements (time, cost and functionality)
  - Reduction in the use of old technology, causing breached SLAs due to problems with support or performance.
- Ability to demonstrate cost-effectiveness:
  - Reduction in last-minute buying to address urgent performance issues
  - Reduction in the over-capacity of IT
  - Accurate forecasts of planned expenditure
  - Reduction in the business disruption caused by a lack of adequate IT capacity
  - Relative reduction in the cost of production of the Capacity Plan.
- Ability to plan and implement the appropriate IT capacity to match business needs:
  - Percentage reduction in the number of incidents due to poor performance
  - Percentage reduction in lost business due to inadequate capacity
  - All new services implemented match Service Level Requirements (SLRs)
  - Increased percentage of recommendations made by Capacity Management are acted on
  - Reduction in the number of SLA breaches due to either poor service performance or poor component performance.

4.3.8 Information Management

The aim of the CMIS is to provide the relevant capacity and performance information to produce reports and support the Capacity Management process. These reports provide valuable information to many IT and Service Management processes. These reports should include the following.

Component-based reports

For each component there should be a team of technical staff responsible for its control and management. Reports must be produced to illustrate how components are performing and how much of their maximum capacity is being used.

Service-based reports

Reports and information must also be produced to illustrate how the service and its constituent components are performing with respect to their overall service targets and constraints. These reports will provide the basis of SLM and customer service reports.

Exception reports

Reports that show management and technical staff when the capacity and performance of a particular component or service becomes unacceptable are also a required from analysis of capacity data. Thresholds can be set for any component, service or measurement within the CMIS. An example threshold may be that processor percentage utilization for a particular server has breached 70% for three consecutive hours, or that the concurrent number of logged-in users exceeds the agreed limit.

In particular, exception reports are of interest to the SLM process in determining whether the targets in SLAs have been breached. Also the Incident and Problem Management processes may be able to use the exception reports in the resolution of incidents and problems.

Predictive and forecast reports

To ensure the IT service provider continues to provide the required service levels, the Capacity Management process must predict future workloads and growth. To do this, future component and service capacity and performance must be forecast. This can be done in a variety of ways, depending on the techniques and the technology used. Changes to workloads by the development and implementation of new functionality and services must be considered alongside growth in the current functionality and services driven by business growth. A simple example of a capacity forecast is a correlation between a business driver and a component utilization, e.g. processor utilization against the number of customer accounts. This data can be correlated to find the effect that an increase in the number of customer accounts will have on the processor utilization. If the forecasts on future capacity requirements identify a requirement for increased resource, this requirement needs to be input into the Capacity Plan and included within the IT budget cycle.

Often capacity reports are consolidated together and stored on an intranet site so that anyone can access and refer to them.

4.3.8.1 *Capacity Management Information System*

Often capacity data is stored in technology-specific tools and databases, and full value of the data, the information and its analysis is not obtained. The true value of the data can only be obtained when the data is combined into a single set of integrated, information repositories or set of databases.

The Capacity Management Information System (CMIS) is the cornerstone of a successful Capacity Management process. Information contained within the CMIS is stored and analysed by all the sub-processes of Capacity Management because it is a repository that holds a number of different types of data, including business, service, resource or utilization and financial data, from all areas of technology.

However, the CMIS is unlikely to be a single database, and probably exists in several physical locations. Data from all areas of technology, and all components that make up the IT services, can then be combined for analysis and provision of technical and management reporting. Only when all of the information is integrated can 'end-to-end' service reports be produced. The integrity and accuracy of the data within the CMIS needs to be carefully managed. If the CMIS is not part of an overall CMS or SKMS, then links between these systems need to be implemented to ensure consistency and accuracy of the information recorded within them.

The information in the CMIS is used to form the basis of performance and Capacity Management reports and views that are to be delivered to customers, IT management and technical personnel. Also, the data is utilized to generate future capacity forecasts and allow Capacity Management to plan for future capacity requirements. Often a web interface is provided to the CMIS to provide the different access and views required outside of the Capacity Management process itself.

The full range of data types stored within the CMIS is as follows.

Business data

It is essential to have quality information on the current and future needs of the business. The future business plans of the organization need to be considered and the effects on the IT services understood. The business data is used to forecast and validate how changes in business drivers affect the capacity and performance of the IT infrastructure. Business data should include business transactions or measurements such as the number of accounts, the number of invoices generated, the number of product lines.

Service data

To achieve a service-orientated approach to Capacity Management, service data should be stored within the CMIS. Typical service data are transaction response times, transaction rates, workload volumes, etc. In general, the SLAs and SLRs provide the service targets for which the Capacity Management process needs to record and monitor data. To ensure that the targets in the SLAs are achieved, SLM thresholds should be included, so that the monitoring activity can measure against these service thresholds and raise exception warnings and reports before service targets are breached.

Component utilization data

The CMIS also needs to record resource data consisting of utilization, threshold and limit information on all of the technological components supporting the services. Most of the IT components have limitations on the level to which they should be utilized. Beyond this level of utilization, the resource will be over-utilized and the performance of the services using the resource will be impaired. For example, the maximum recommended level of utilization on a processor could be 80%, or the utilization of a shared Ethernet LAN segment should not exceed 40%.

Also, components have various physical limitations beyond which greater connectivity or use is impossible. For example, the maximum number of connections through an application or a network gateway is 100, or a particular type of disk has a physical capacity of 15Gb. The CMIS should therefore contain, for each component and the maximum performance and capacity limits, current and past utilization rates and the associated component thresholds. Over time this can require vast amounts of data to be accumulated, so there need to be good techniques for analysing, aggregating and archiving this data.

Financial data

The Capacity Management process requires financial data. For evaluating alternative upgrade options, when proposing various scenarios in the Capacity Plan, the financial cost of the upgrades to the components of the IT infrastructure, together with information about the current IT hardware budget, must be known and included in the considerations. Most of this data may be available from the Financial Management for IT services process, but

Capacity Management needs to consider this information when managing the future business requirements.

4.3.9 Challenges, Critical Success Factors and risks

One of the major challenges facing Capacity Management is persuading the business to provide information on its strategic business plans, to enable the IT service provider organization to provide effective Business Continuity Management (BCM). This is particularly true in outsourced situations where there may be commercial or confidential reasons why this data cannot be shared. Even if the data on the strategic business plan is available, there may be issues with regard to the quality or accuracy of the data contained within the business plans with regard to BCM.

Another challenge is the combination of all of the CCM data into an integrated set of information that can be analysed in a consistent manner to provide details of the usage of all components of the services. This is particularly challenging when the information from the different technologies is provided by different tools in differing formats. Often the quality of component information on the performance of the technology is variable in both its quality and accuracy.

The amounts of information produced by BCM, and especially SCM and CCM, are huge and the analysis of this information is difficult to achieve. The people and the processes need to focus on the key resources and their usage, whilst not ignoring other areas. In order to do this, appropriate thresholds must be used, and reliance placed on the tools and technology to automatically manage the technology and provide warnings and alerts when things deviate significantly from the 'norm'.

The main CSFs for the Capacity Management process are:

- Accurate business forecasts
- Knowledge of current and future technologies
- Ability to demonstrate cost-effectiveness
- Ability to plan and implement the appropriate IT capacity to match business need.

Some of the major risks associated with Capacity Management include:

- A lack of commitment from the business to the Capacity Management process
- A lack of appropriate information from the business on future plans and strategies
- A lack of senior management commitment or a lack of resources and/or budget for the Capacity Management process
- SCM and CCM performed in isolation because BCM is difficult, or there is a lack of appropriate and accurate business information
- The processes become too bureaucratic or manually intensive
- The processes focus too much on the technology (CCM) and not enough on the services (SCM) and the business (BCM)
- The reports and information provided are too bulky or too technical and do not give the information required or appropriate to the customers and the business.

4.4 AVAILABILITY MANAGEMENT

4.4.1 Purpose/goal/objective

The goal of the Availability Management process is to ensure that the level of service availability delivered in all services is matched to or exceeds the current and future agreed needs of the business, in a cost-effective manner.

The purpose of Availability Management is to provide a point of focus and management for all availability-related issues, relating to both services and resources, ensuring that availability targets in all areas are measured and achieved.

The objectives of Availability Management are to:

- Produce and maintain an appropriate and up-to-date Availability Plan that reflects the current and future needs of the business
- Provide advice and guidance to all other areas of the business and IT on all availability-related issues
- Ensure that service availability achievements meet or exceed all their agreed targets, by managing services and resources-related availability performance
- Assist with the diagnosis and resolution of availability-related incidents and problems
- Assess the impact of all changes on the Availability Plan and the performance and capacity of all services and resources
- Ensure that proactive measures to improve the availability of services are implemented wherever it is cost-justifiable to do so.

Availability Management should ensure the agreed level of availability is provided. The measurement and monitoring of IT availability is a key activity to ensure availability levels are being met consistently. Availability Management should look to continually optimize and proactively improve the availability of the IT infrastructure, the services and the supporting organization, in order to provide cost-effective availability improvements that can deliver business and customer benefits.

4.4.2 Scope

The scope of the Availability Management process covers the design, implementation, measurement, management and improvement of IT service and component availability. Availability Management needs to understand the service and component availability requirements from the business perspective in terms of the:

- Current business processes, their operation and requirements
- Future business plans and requirements
- Service targets and the current IT service operation and delivery
- IT infrastructure, data, applications and environment and their performance
- Business impacts and priorities in relation to the services and their usage.

Understanding all of this will enable Availability Management to ensure that all the services and components are designed and delivered to meet their targets in terms of agreed business needs. The Availability Management process:

- Should be applied to all operational services and technology, particularly those covered by SLAs. It can also be applied to those IT services deemed to be business critical regardless of whether formal SLAs exist
- Should be applied to all new IT services and for existing services where Service Level Requirements (SLRs) or Service Level Agreements (SLAs) have been established
- Should be applied to all supporting services and the partners and suppliers (both internal and external) that form the IT support organization as a precursor to the creation of formal agreements
- Considers all aspects of the IT services and components and supporting organizations that may impact availability, including training, skills, process effectiveness, procedures and tools.

The Availability Management process does not include Business Continuity Management and the resumption of business processing after a major disaster. The support of BCM is included within IT Service Continuity Management (ITSCM). However, Availability Management does provide key inputs to ITSCM, and the two processes have a close relationship, particularly in the assessment and management of risks and in the implementation of risk reduction and resilience measures.

The Availability Management process should include:

- Monitoring of all aspects of availability, reliability and maintainability of IT services and the supporting components, with appropriate events, alarms and escalation, with automated scripts for recovery
- Maintenance of a set of methods, techniques and calculations for all availability measurements, metrics and reporting
- Assistance with risk assessment and management activities
- Collection of measurements, analysis and production of regular and ad hoc reports on service and component availability
- Understanding the agreed current and future demands of the business for IT services and their availability
- Influencing the design of services and components to align with business needs
- Producing an Availability Plan that enables the service provider to continue to provide and improve services in line with availability targets defined in Service Level Agreements (SLAs), and to plan and forecast future availability levels required, as defined in Service Level Requirements (SLRs)
- Maintaining a schedule of tests for all resilient and fail-over components and mechanisms
- Assistance with the identification and resolution of any incidents and problems associated with service or component unavailability
- Proactive improvement of service or component availability wherever it is cost-justifiable and meets the needs of the business.

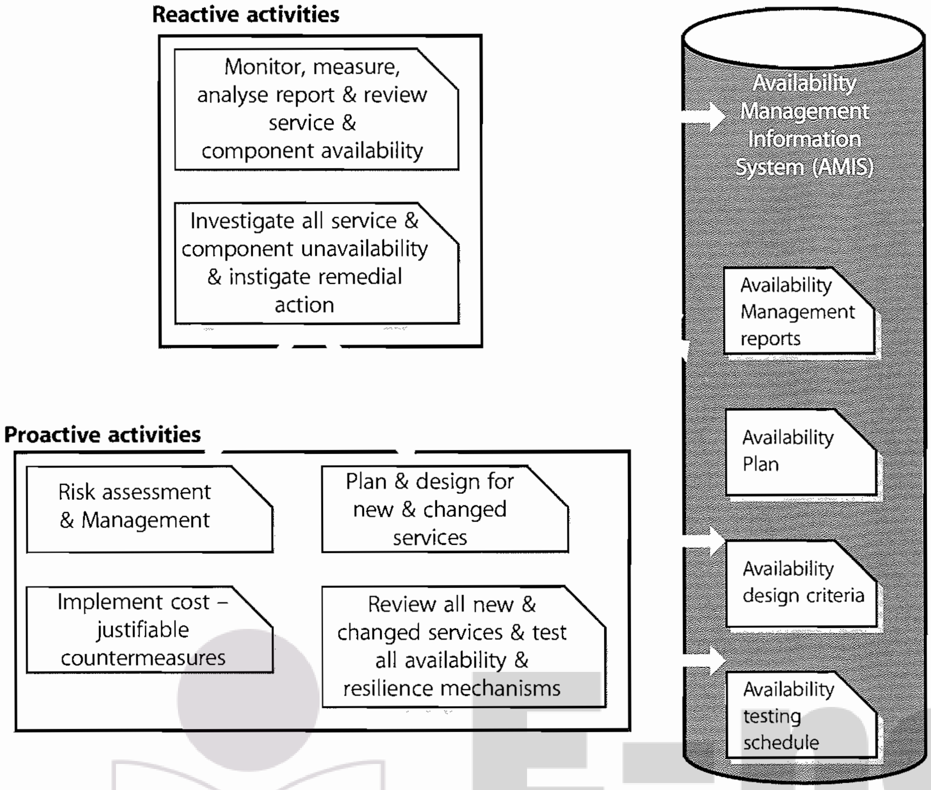

Reactive activities
Monitor, measure, analyse report & review service & component availability
Investigate all service & component unavailability & instigate remedial action
Proactive activities
Risk assessment & Management
Plan & design for new & changed services
Implement cost – justifiable countermeasures
Review all new & changed services & test all availability & resilience mechanisms
Availability Management Information System (AMIS)
Availability Management reports
Availability Plan
Availability design criteria
Availability testing schedule

Figure 4.13 The Availability Management process

4.4.3 Value to the business

The Availability Management process ensures that the availability of systems and services matches the evolving agreed needs of the business. The role of IT within the business is now pivotal. The availability and reliability of IT services can directly influence customer satisfaction and the reputation of the business. This is why Availability Management is essential in ensuring IT delivers the right levels of service availability required by the business to satisfy its business objectives and deliver the quality of service demanded by its customers. In today's competitive marketplace, customer satisfaction with service(s) provided is paramount. Customer loyalty can no longer be relied on, and dissatisfaction with the availability and reliability of IT service can be a key factor in customers taking their business to a competitor.

The Availability Management process and planning, just like Capacity Management, must be involved in all stages of the Service Lifecycle, from Strategy and Design, through Transition and Operation to Improvement. The appropriate availability and resilience should be designed into services and components from the initial design stages. This will ensure not only that the availability of any new or changed service meets its expected targets, but also that all existing services and components continue to meet all of their targets. This is the basis of stable service provision.

4.4.4 Policies/principles/basic concepts

The Availability Management process is continually trying to ensure that all operational services meet their agreed availability targets, and that new or changed services are designed appropriately to meet their intended targets, without compromising the performance of existing services. In order to achieve this, Availability Management should perform the reactive and proactive activities illustrated in Figure 4.13.

The reactive activities of Availability Management consist of monitoring, measuring, analysing, reporting and reviewing all aspects of component and service availability. This is to ensure that all agreed service targets are measured and achieved. Wherever deviations or breaches are detected, these are investigated and remedial action instigated. Most of these activities are conducted within the Operations stage of the lifecycle and are linked into the monitoring and control activities, Event and Incident Management processes. (See the Service Operation publication.)

The proactive activities consist of producing recommendations, plans and documents on design guidelines and criteria for new and changed services, and the continual improvement of service and reduction of risk in existing services wherever it can be cost-justified. These are key aspects to be considered within Service Design activities.

An effective Availability Management process, consisting of both the reactive and proactive activities, can 'make a big difference' and will be recognized as such by the business, if the deployment of Availability Management within an IT organization has a strong emphasis on the needs of the business and customers. To reinforce this emphasis, there are several guiding principles that should underpin the Availability Management process and its focus:

- Service availability is at the core of customer satisfaction and business success: there is a direct correlation in most organizations between the service availability and customer and user satisfaction, where poor service performance is defined as being unavailable.
- Recognizing that when services fail, it is still possible to achieve business, customer and user satisfaction and recognition: the way a service provider reacts in a failure situation has a major influence on customer and user perception and expectation.
- Improving availability can only begin after understanding how the IT services support the operation of the business.
- Service availability is only as good as the weakest link on the chain: it can be greatly increased by the elimination of Single Points of Failure (SPoFs) or an unreliable or weak component.
- Availability is not just a reactive process. The more proactive the process, the better service availability will be. Availability should not purely react to service and component failure. The more events and failures are predicted, pre-empted and prevented, the higher the level of service availability.
- It is cheaper to design the right level of service availability into a service from the start rather than try and 'bolt it on' subsequently. Adding resilience into a service or component is invariably more expensive than designing it in from the start. Also, once a service gets a bad name for unreliability, it becomes very difficult to change the image. Resilience is also a key consideration of ITSCM, and this should be considered at the same time.

The scope of Availability Management covers the design, implementation, measurement and management of IT service and infrastructure availability. This is reflected in the process description shown in Figure 4.13 and described in the following paragraphs.

The Availability Management process has two key elements:

- Reactive activities: the reactive aspect of Availability Management involves the monitoring, measuring, analysis and management of all events, incidents and problems involving unavailability. These activities are principally involved within operational roles.
- Proactive activities: the proactive activities of Availability Management involve the proactive planning, design and improvement of availability. These activities are principally involved within design and planning roles.

Availability Management is completed at two inter-connected levels:

- **Service availability:** involves all aspects of service availability and unavailability and the impact of component availability, or the potential impact of component unavailability on service availability
- **Component availability:** involves all aspects of component availability and unavailability.

Availability Management relies on the monitoring, measurement, analysis and reporting of the following aspects:

Availability: the ability of a service, component or CI to perform its agreed function when required. It is often measured and reported as a percentage:

$$\textbf{Availability (\%)} = \frac{\textbf{(Agreed Service Time (AST) – downtime)}}{\textbf{Agreed Service Time (AST)}} \textbf{ X 100 \%}$$

Note: Downtime should only be included in the above calculation when it occurs within the Agreed Service Time (AST). However, total downtime should also be recorded and reported.

Reliability: a measure of how long a service, component or CI can perform its agreed function without interruption. The reliability of the service can be improved by increasing the reliability of individual components or by increasing the resilience of the service to individual component failure (i.e. increasing the component redundancy, e.g. by using load-balancing techniques). It is often measured and reported as Mean Time Between Service Incidents (MTBSI) or Mean Time Between Failures (MTBF):

$$\textbf{Reliability (MTBSI in hours)} = \frac{\textbf{Available time in hours}}{\textbf{Number of breaks}}$$

$$\textbf{Reliability (MTBF in hours)} = \frac{\textbf{Available time in hours} - \textbf{Total downtime in hours}}{\textbf{Number of breaks}}$$

Maintainability: a measure of how quickly and effectively a service, component or CI can be restored to normal working after a failure. It is measured and reported as Mean Time to Restore Service (MTRS) and should be calculated using the following formula:

$$\textbf{Maintainability (MTRS in hours)} = \frac{\textbf{Total downtime in hours}}{\textbf{Number of service breaks}}$$

MTRS should be used to avoid the ambiguity of the more common industry term Mean Time To Repair (MTTR), which in some definitions includes only repair time, but in others includes recovery time. The downtime in MTRS covers all the contributory factors that make the service, component or CI unavailable:

- Time to record
- Time to respond
- Time to resolve
- Time to physically repair or replace
- Time to recover.

Example: A situation where a 24 x 7 service has been running for a period of 5,020 hours with only two breaks, one of six hours and one of 14 hours, would give the following figures:

Availability = (5,020-(6+14)) / 5,020 x 100 = 99.60%

Reliability (MTBSI) = 5,020 / 2 = 2,510 hours

Reliability (MTBF) = 5,020-(6+14) / 2 = 2,500 hours

Maintainability (MTRS) = (6+14) / 2 = 10 hours

Serviceability: the ability of a third-party supplier to meet the terms of their contract. Often this contract will include agreed levels of availability, reliability and/or maintainability for a supporting service or component.

These aspects and their inter-relationships are illustrated in Figure 4.14.

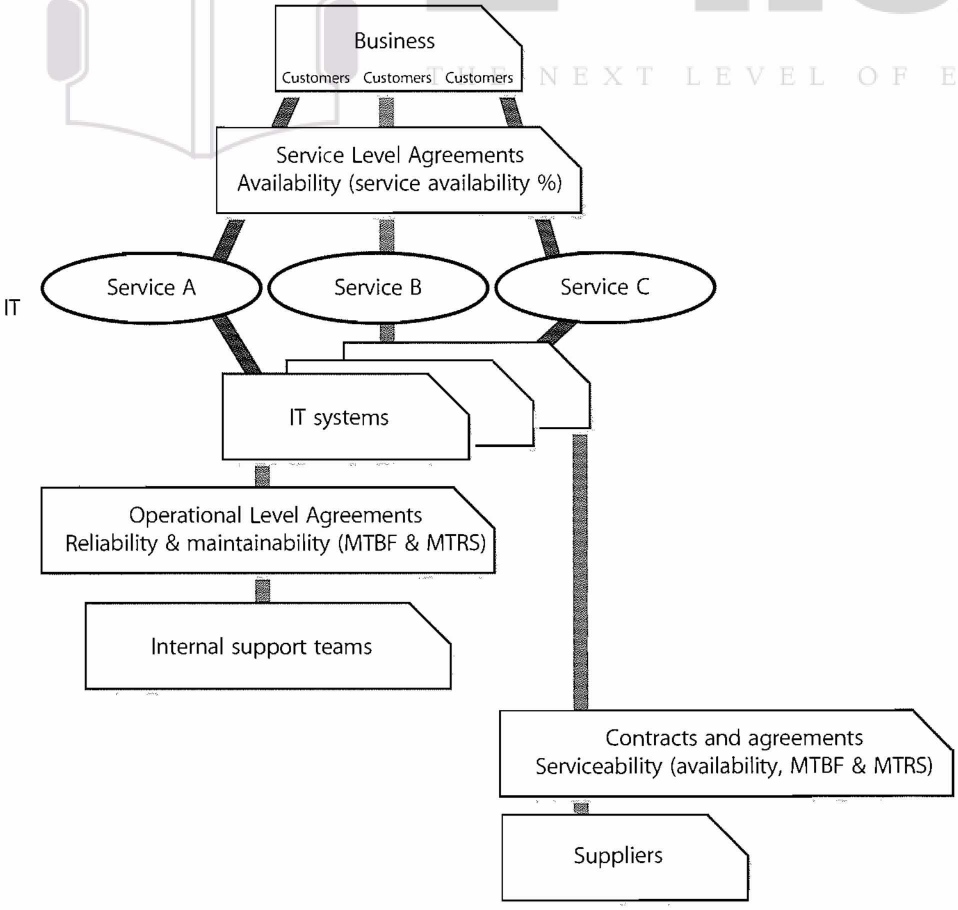


Figure 4.14 Availability terms and measurements

Although the principal service target contained within SLAs for the customers and business is availability, as illustrated in Figure 4.14, some customers also require reliability and maintainability targets to be included as well. Where these are included they should relate to service reliability and maintainability targets, whereas the reliability and maintainability targets contained in OLAs and contracts relate to component and supporting service targets and can often include availability targets relating to the relevant components or supporting services.

The term Vital Business Function (VBF) is used to reflect the business critical elements of the business process supported by an IT service. An IT service may support a number of business functions that are less critical. For example, an automated teller machine (ATM) or cash dispenser service VBF would be the dispensing of cash. However, the ability to obtain a statement from an ATM may not be considered as vital. This distinction is important and should influence availability design and associated costs. The more vital the business function generally, the greater the level of resilience and availability that needs to be incorporated into the design required in the supporting IT services. For all services, whether VBFs or not, the availability requirements should be determined by the business and not by IT. The initial availability targets are often set at too high a level, and this leads to either over-priced services or an iterative discussion between the service provider and the business to agree an appropriate compromise between the service availability and the cost of the service and its supporting technology.

Certain VBFs may need special designs, which are now being used as a matter of course within Service Design plans, incorporating:

- **High availability:** a characteristic of the IT service that minimizes or masks the effects of IT component failure to the users of a service.
- **Fault tolerance:** the ability of an IT service, component or CI to continue to operate correctly after failure of a component part.
- **Continuous operation:** an approach or design to eliminate planned downtime of an IT service. Note that individual components or CIs may be down even though the IT service remains available.
- **Continuous availability:** an approach or design to achieve 100% availability. A continuously Available IT service has no planned or unplanned downtime.

> **Industry view**
>
> Many suppliers commit to high availability or continuous availability solutions only if stringent environmental standards and resilient processes are used. They often agree to such contracts only after a site survey has been completed and additional, sometimes costly, improvements have been made.

Availability Management commences as soon as the availability requirements for an IT service are clear enough to be articulated. It is an ongoing process, finishing only when the IT service is decommissioned or retired. The key activities of the Availability Management process are:

- Determining the availability requirements from the business for a new or enhanced IT service and formulating the availability and recovery design criteria for the supporting IT components
- Determining the VBFs, in conjunction with the business and ITSCM
- Determining the impact arising from IT service and component failure in conjunction with ITSCM and, where appropriate, reviewing the availability design criteria to provide additional resilience to prevent or minimize impact to the business
- Defining the targets for availability, reliability and maintainability for the IT infrastructure components that underpin the IT service to enable these to be documented and agreed within SLAs, OLAs and contracts
- Establishing measures and reporting of availability, reliability and maintainability that reflect the business, user and IT support organization perspectives
- Monitoring and trend analysis of the availability, reliability and maintainability of IT components
- Reviewing IT service and component availability and identifying unacceptable levels
- Investigating the underlying reasons for unacceptable availability
- Producing and maintaining an Availability Plan that prioritizes and plans IT availability improvements.

4.4.5 Process activities, methods and techniques

The Availability Management process depends heavily on the measurement of service and component achievements with regard to availability.

Key messages

'If you don't measure it, you can't manage it'
'If you don't measure it, you can't improve it'
'If you don't measure it, you probably don't care'
'If you can't influence or control it, then don't measure it'

'What to measure and how to report it' inevitably depends on which activity is being supported, who the recipients are and how the information is to be utilized. It is important to recognize the differing perspectives of availability to ensure measurement and reporting satisfies these varied needs:

- The business perspective considers IT service availability in terms of its contribution or impact on the VBFs that drive the business operation.
- The user perspective considers IT service availability as a combination of three factors, namely the frequency, the duration and the scope of impact, i.e. all users, some users, all business functions or certain business functions – the user also considers IT service availability in terms of response times. For many performance-centric applications, poor response times are considered equal in impact to failures of technology.
- The IT service provider perspective considers IT service and component availability with regard to availability, reliability and maintainability.

In order to satisfy the differing perspectives of availability, Availability Management needs to consider the spectrum of measures needed to report the 'same' level of availability in different ways. Measurements need to be meaningful and add value if availability measurement and reporting are ultimately to deliver benefit to the IT and business organizations. This is influenced strongly by the combination of 'what you measure' and 'how you report it'.

4.4.5.1 The reactive activities of Availability Management

Monitor, measure, analyse and report service and component availability

A key output from the Availability Management process is the measurement and reporting of IT availability. Availability measures should be incorporated into SLAs, OLAs and any underpinning contracts. These should be reviewed regularly at Service Level review meetings. Measurement and reporting provide the basis for:

- Monitoring the actual availability delivered versus agreed targets
- Establishing measures of availability and agreeing availability targets with the business
- Identifying unacceptable levels of availability that impact the business and users
- Reviewing availability with the IT support organization
- Continual improvement activities to optimize availability.

The IT service provider organizations have, for many years, measured and reported on their perspective of availability. Traditionally these measures have concentrated on component availability and have been somewhat divorced from the business and user views. Typically these traditional measures are based on a combination of an availability percentage (%), time lost and the frequency of failure. Some examples of these traditional measures are as follows:

- **Per cent available** – the truly 'traditional' measure that represents availability as a percentage and, as such, much more useful as a component availability measure than a service availability measure. It is typically used to track and report achievement against a service level target. It tends to emphasize the 'big number' such that if the service level target was 98.5% and the achievement was 98.3%, then it does not seem that bad. This can encourage a complacent behaviour within the IT support organization.
- **Per cent unavailable** – the inverse of the above. This representation, however, has the benefit of focusing on non-availability. Based on the above example, if the target for non-availability is 1.5% and the achievement was 1.7%, then this is a much larger relative difference. This method of reporting is more likely to create awareness of the shortfall in delivering the level of availability required.
- **Duration** – achieved by converting the percentage unavailable into hours and minutes. This provides a more 'human' measure that people can relate to. If the weekly downtime target is two hours, but one week the actual downtime was four hours; this would represent a trend leading to an additional four days of non-availability to the business over a full year. This type of measure and reporting is more likely to encourage focus on service improvement.
- **Frequency of failure** – used to record the number of interruptions to the IT service. It helps provide a good indication of reliability from a user perspective. It is best used in combination with 'duration' to take a balanced view of the level of service interruptions and the duration of time lost to the business.
- **Impact of failure** – this is the true measure of service unavailability. It depends on mature incident recording

where the inability of users to perform their business tasks is the most important piece of information captured. All other measures suffer from a potential to mask the real effects of service failure and are often converted to a financial impact.

The business may have, for many years, accepted that the IT availability that they experience is represented in terms of component availability rather than overall service or business availability. However, this is no longer being viewed as acceptable and the business is keen to better represent availability in measure(s) that demonstrate the positive and negative consequences of IT availability on their business and users.

Key messages

The most important availability measurements are those that reflect and measure availability from the business and user perspective.

Availability Management needs to consider availability from both a business/IT service provider perspective and from an IT component perspective. These are entirely different aspects, and while the underlying concept is similar, the measurement, focus and impact are entirely different.

The sole purpose of producing these availability measurements and reports, including those from the business perspective, is to improve the quality and availability of IT service provided to the business and users. All measures, reports and activities should reflect this purpose.

Availability, when measured and reported to reflect the experience of the user, provides a more representative view on overall IT service quality. The user view of availability is influenced by three factors:

- Frequency of downtime
- Duration of downtime
- Scope of impact.

Measurements and reporting of user availability should therefore embrace these factors. The methodology employed to reflect user availability could consider two approaches:

- **Impact by user minutes lost:** this is to base calculations on the duration of downtime multiplied by the number of users impacted. This can be the basis to report availability as lost user productivity, or to calculate the availability percentage from a user perspective, and can also include the costs of recovery for lost productivity (e.g. increased overtime payments).
- **Impact by business transaction:** this is to base calculations on the number of business transactions that could not be processed during the period of downtime. This provides a better indication of business impact reflecting differing transaction processing profiles across the time of day, week etc. In many instances it may be the case that the user impact correlates to a VBF, e.g. if the user takes customer purchase orders and a VBF is customer sales. This single measure is the basis to reflect impact to the business operation and user.

The method employed should be influenced by the nature of the business operation. A business operation supporting data entry activity is well suited to reporting that reflects user productivity loss. Business operations that are more customer-facing, e.g. ATM services, benefit from reporting transaction impact. It should also be noted that not all business impact is user-related. With increasing automation and electronic processing, the ability to process automated transactions or meet market cut-off times can also have a large financial impact that may be greater than the ability of users to work.

The IT support organization needs to have a keen awareness of the user experience of availability. However, the real benefits come from aggregating the user view into the overall business view. A guiding principle of the Availability Management process is that **'Improving availability can only begin when the way technology supports the business is understood'.** Therefore Availability Management isn't just about understanding the availability of each IT component, but is all about understanding the impact of component failure on service and user availability. From the business perspective, an IT service can only be considered available when the business is able to perform all vital business functions required to drive the business operation. For the IT service to be available, it therefore relies on all components on which the service depends being available, i.e. systems, key components, network, data and applications.

The traditional IT approach would be to measure individually the availability of each of these components. However, the true measure of availability has to be based on the positive and negative impacts on the VBFs on which the business operation is dependent. This approach ensures that SLAs and IT availability reporting are based on measures that are understood by both the business and IT. By measuring the VBFs that rely on IT services, measurement and reporting becomes business-driven, with the impact of failure reflecting the consequences to the business. It is also important that the availability of the services is defined and agreed with the business and reflected within SLAs. This definition of availability should include:

- What is the minimum available level of functionality of the service?
- At what level of service response is the service considered unavailable?
- Where will this level of functionality and response be measured?
- What are the relative weightings for partial service unavailability?
- If one location or office is impacted, is the whole service considered unavailable, or is this considered to be 'partial unavailability'? This needs to be agreed with the customers.

Reporting and analysis tools are required for the manipulation of data stored in the various databases utilized by Availability Management. These tools can either be platform- or PC-based and are often a combination of the two. This will be influenced by the database repository technologies selected and the complexity of data processing and reporting required. Availability Management, once implemented and deployed, will be required to produce regular reports on an agreed basis, e.g. monthly availability reports, Availability Plan, Service Failure Analysis (SFA) status reports, etc. The activities involved within these reporting activities can require much manual effort and the only solution is to automate as much of the report generation activity as possible. For reporting purposes, organizational reporting standards should be used wherever possible. If these don't exist, IT standards should be developed so that IT reports can be developed using standard tools and techniques. This means that the integration and consolidation of reports will subsequently be much easier to achieve.

Unavailability analysis

All events and incidents causing unavailability of services and components should be investigated, with remedial actions being implemented within either the Availability Plan or the overall SIP. Trends should be produced from this analysis to direct and focus activities such as Service Failure Analysis (SFA) to those areas causing the most impact or disruption to the business and the users.

The overall costs of an IT service are influenced by the levels of availability required and the investments required in technology and services provided by the IT support organization to meet this requirement. Availability certainly does not come for free. However, it is important to reflect that the unavailability of IT also has a cost, therefore unavailability isn't free either. For highly critical business processes and VBFs, it is necessary to consider not only the cost of providing the service, but also the costs that are incurred from failure. The optimum balance to strike is the cost of the availability solution weighed against the costs of unavailability.

Before any SLR is accepted, and ultimately the SLR or SLA is negotiated and agreed between the business and the IT organization, it is essential that the availability requirements of the business are analysed to assess if/how the IT service can deliver the required levels of availability. This applies not only to new IT services that are being introduced, but also to any requested changes to the availability requirements of existing IT services.

The cost of an IT failure could simply be expressed as the number of business or IT transactions impacted, either as an actual figure (derived from instrumentation) or based on an estimation. When measured against the VBFs that support the business operation, this can provide an obvious indication of the consequence of failure. The advantage of this approach is the relative ease of obtaining the impact data and the lack of any complex calculations. It also becomes a 'value' that is understood by both the business and IT organization. This can be the stimulus for identifying improvement opportunities and can become a key metric in monitoring the availability of IT services.

The major disadvantage of this approach is that it offers no obvious monetary value that would be needed to justify any significant financial investment decisions for improving availability. Where significant financial investment decisions are required, it is better to express the cost of failure arising from service, system, application or function loss to the business as a monetary 'value'.

The monetary value can be calculated as a combination of the tangible costs associated with failure, but can also include a number of intangible costs. The monetary value should also reflect the cost impact to the whole organization, i.e. the business and IT organization.

Tangible costs can include:

- Lost user productivity
- Lost IT staff productivity
- Lost revenue
- Overtime payments
- Wasted goods and material
- Imposed fines or penalty payments.

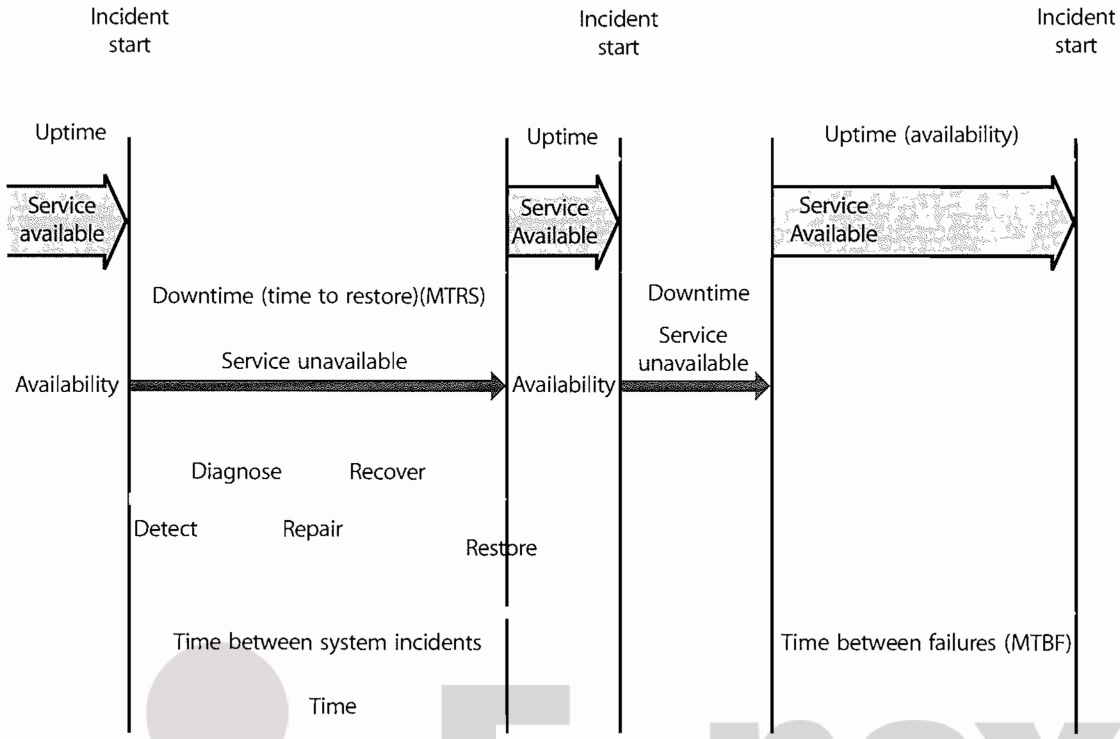


Figure 4.15 The expanded incident lifecycle

These costs are often well understood by the finance area of the business and IT organization, and in relative terms are easier to obtain and aggregate than the intangible costs associated with an IT failure.

Intangible costs can include:

- Loss of customers
- Loss of customer goodwill (customer dissatisfaction)
- Loss of business opportunity (to sell, gain new customers or revenue, etc.)
- Damage to business reputation
- Loss of confidence in IT service provider
- Damage to staff morale.

It is important not simply to dismiss the intangible costs (and the potential consequences) on the grounds that they may be difficult to measure. The overall unavailability of service, the total tangible cost and the total intangible costs arising from service unavailability are all key metrics in the measurement of the effectiveness of the Availability Management process.

The expanded incident lifecycle

A guiding principle of Availability Management is to recognize that it is still possible to gain customer satisfaction even when things go wrong. One approach to help achieve this requires Availability Management to ensure that the duration of any incident is minimized to enable normal business operations to resume as quickly as possible. An aim of Availability Management is to ensure the duration and impact from incidents impacting IT services are minimized, to enable business operations to resume as quickly as is possible. The analysis of the 'expanded incident lifecycle' enables the total IT service downtime for any given incident to be broken down and mapped against the major stages through which all incidents progress (the lifecycle). Availability Management should work closely with Incident Management and Problem Management in the analysis of all incidents causing unavailability.

A good technique to help with the technical analysis of incidents affecting the availability of components and IT services is to take an incident 'lifecycle' view. Every incident passes through several major stages. The time elapsed in these stages may vary considerably. For Availability Management purposes, the standard incident 'lifecycle', as described within Incident Management, has been expanded to provide additional help and guidance particularly in the area of 'designing for recovery'. Figure 4.15 illustrates the expanded incident lifecycle.

From the above it can be seen that an incident can be broken down into individual stages within a lifecycle that can be timed and measured. This lifecycle view provides an important framework in determining, amongst others, systems management requirements for event and incident detection, diagnostic data capture requirements and tools for diagnosis, recovery plans to aid speedy recovery and how to verify that IT service has been restored. The individual stages of the lifecycle are considered in more detail as follows.

- **Incident detection:** the time at which the IT service provider organization is made aware of an incident. Systems management tools positively influence the ability to detect events and incidents and therefore to improve levels of availability that can be delivered. Implementation and exploitation should have a strong focus on achieving high availability and enhanced recovery objectives. In the context of recovery, such tools should be exploited to provide automated failure detection, assist failure diagnosis and support automated error recovery, with scripted responses. Tools are very important in reducing all stages of the incident lifecycle, but principally the detection of events and incidents. Ideally the event is automatically detected and resolved, before the users have noticed it or have been impacted in any way.
- **Incident diagnosis:** the time at which diagnosis to determine the underlying cause has been completed. When IT components fail, it is important that the required level of diagnostics is captured, to enable problem determination to identify the root cause and resolve the issue. The use and capability of diagnostic tools and skills is critical to the speedy resolution of service issues. For certain failures, the capture of diagnostics may extend service downtime. However, the non-capture of the appropriate diagnostics creates and exposes the service to repeat failures. Where the time required to take diagnostics is considered excessive, or varies from the target, a review should be instigated to identify if techniques and/or procedures can be streamlined to reduce the time required. Equally the scope of the diagnostic data available for capture can be assessed to ensure only the diagnostic data considered essential is taken. The additional downtime required to capture diagnostics should be included in the recovery metrics documented for each IT component.
- **Incident repair:** the time at which the failure has been repaired/fixed. Repair times for incidents should be continuously monitored and compared against the targets agreed within OLAs, underpinning contracts and other agreements. This is particularly important with respect to externally provided services and supplier performance. Wherever breaches are observed, techniques should be used to reduce or remove the breaches from similar incidents in the future.
- **Incident recovery:** the time at which component recovery has been completed. The backup and recovery requirements for the components underpinning a new IT service should be identified as early as possible within the design cycle. These requirements should cover hardware, software and data and recovery targets. The outcome from this activity should be a documented set of recovery requirements that enables the development of appropriate recovery plans. To anticipate and prepare for performing recovery such that reinstatement of service is effective and efficient requires the development and testing of appropriate recovery plans based on the documented recovery requirements. Wherever possible, the operational activities within the recovery plan should be automated. The testing of the recovery plans also delivers approximate timings for recovery. These recovery metrics can be used to support the communication of estimated recovery of service and validate or enhance the Component Failure Impact Analysis documentation. Availability Management must continuously seek and promote faster methods of recovery for all potential Incidents. This can be achieved via a range of methods, including automated failure detection, automated recovery, more stringent escalation procedures, exploitation of new and faster recovery tools and techniques. Availability requirements should also contribute to determining what spare parts are kept within the Definitive Spares to facilitate quick and effective repairs, as described within the Service Transition publication.
- **Incident restoration:** the time at which normal business service is resumed. An incident can only be considered 'closed' once service has been restored and normal business operation has resumed. It is important that the restored IT service is verified as working correctly as soon as service restoration is completed and before any technical staff involved in the incident are stood down. In the majority of cases, this is simply a case of getting confirmation from the affected users. However, the users for some services may be customers of the business, i.e. ATM services, internet-based services. For these types of services, it is recommended that IT service verification procedures are developed to enable the IT service provider organization to verify that a restored IT service is now working as expected. These could simply be visual checks of transaction throughput or user simulation scripts that validate the end-to-end service.

Each stage, and the associated time taken, influences the total downtime perceived by the user. By taking this approach it is possible to see where time is being 'lost' for the duration of an incident. For example, the service was unavailable to the business for 60 minutes, yet it only took five minutes to apply a fix – where did the other 55 minutes go?

Using this approach identifies possible areas of inefficiency that combine to make the loss of service experienced by the business greater than it need be. These could cover areas such as poor automation (alerts, automated recovery etc.), poor diagnostic tools and scripts, unclear escalation procedures (which delay the escalation to the appropriate technical support group or supplier), or lack of comprehensive operational documentation. Availability Management needs to work in close association with Incident and Problem Management to ensure repeat occurrences are eliminated. It is recommended that these measures are established and captured for all availability incidents. This provides Availability Management with metrics for both specific incidents and trending information. This information can be used as input to SFA assignments, SIP activities and regular Availability Management reporting and to provide an impetus for continual improvement activity to pursue cost-effective improvements. It can also enable targets to be set for specific stages of the incident lifecycle. While accepting that each incident may have a wide range of technical complexity, the targets can be used to reflect the consistency of how the IT service provider organization responds to incidents.

An output from the Availability Management process is the real-time monitoring requirements for IT services and components. To achieve the levels of availability required and/or ensure the rapid restoration of service following an IT failure requires investment and exploitation of a systems management toolset. Systems management tools are an essential building block for IT services that require a high level of availability and can provide an invaluable role in reducing the amount of downtime incurred. Availability Management requirements cover the detection and alerting of IT service and component exceptions, automated escalation and notification of IT failures and the automated recovery and restoration of components from known IT failure situations. This makes it possible to identify where 'time is being lost' and provides the basis for the identification of factors that can improve recovery and restoration times. These activities are performed on a regular basis within Service Operation.

Service Failure Analysis

Service Failure Analysis (SFA) is a technique designed to provide a structured approach to identifying the underlying causes of service interruptions to the user. SFA utilizes a range of data sources to assess where and why shortfalls in availability are occurring. SFA enables a holistic view to be taken to drive not just technology improvements, but also improvements to the IT support organization, processes, procedures and tools. SFA is run as an assignment or project, and may utilize other Availability Management methods and techniques to formulate the recommendations for improvement. The detailed analysis of service interruptions can identify opportunities to enhance levels of availability. SFA is a structured technique to identify improvement opportunities in end-to-end service availability that can deliver benefits to the user. Many of the activities involved in SFA are closely aligned with those of Problem Management, and in a number of organizations these activities are performed jointly by Problem and Availability Management.

The high-level objectives of SFA are:

- To improve the overall availability of IT services by producing a set of improvements for implementation or input to the Availability Plan
- To identify the underlying causes of service interruption to users
- To assess the effectiveness of the IT support organization and key processes
- To produce reports detailing the major findings and recommendations
- That availability improvements derived from SFA-driven activities are measured.

SFA initiatives should use input from all areas and all processes including, most importantly, the business and users. Each SFA assignment should have a recognized sponsor(s) (ideally, joint sponsorship from the IT and business) and involve resources from many technical and process areas. The use of the SFA approach:

- Provides the ability to deliver enhanced levels of availability without major cost
- Provides the business with visible commitment from the IT support organization
- Develops in-house skills and competencies to avoid expensive consultancy assignments related to availability improvement
- Encourages cross-functional team working and breaks barriers between teams, and is an enabler to lateral

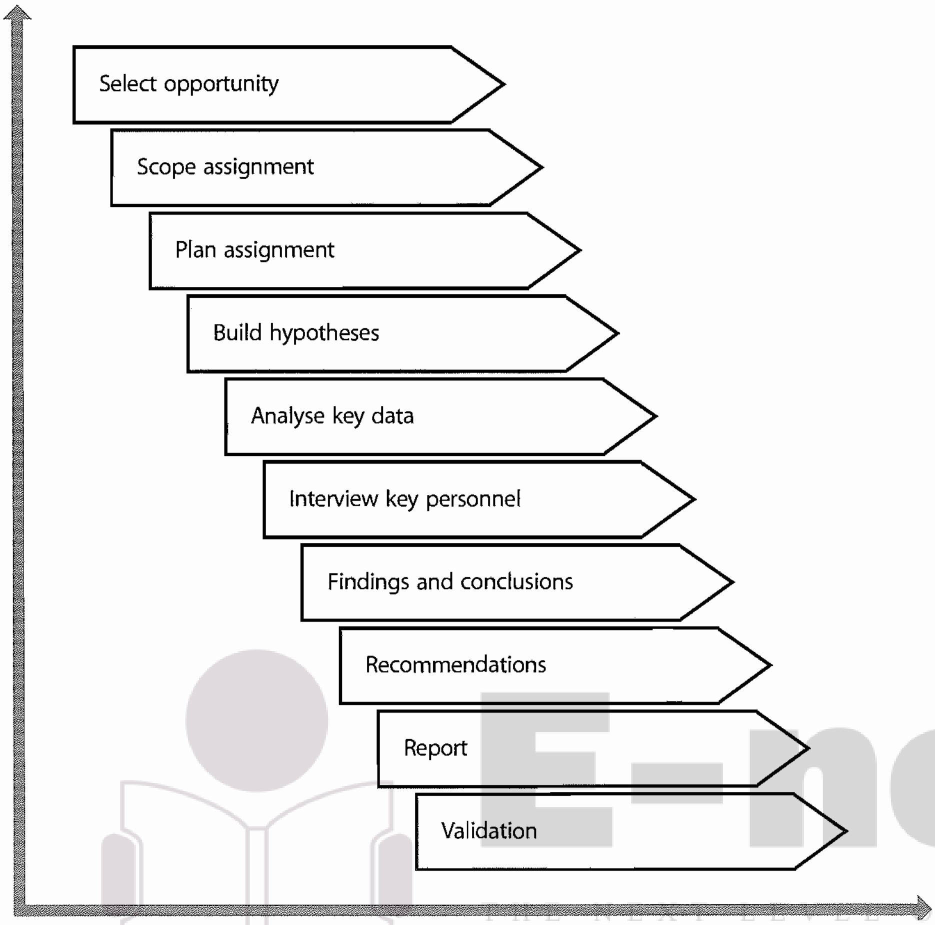


Figure 4.16 The structured approach to Service Failure Analysis (SFA)

thinking, challenging traditional thoughts and providing innovative, and often inexpensive, solutions

- Provides a programme of improvement opportunities that can make a real difference to service quality and user perception
- Provides opportunities that are focused on delivering benefit to the user
- Provides an independent 'health check' of IT Service Management processes and is the stimulus for process improvements.

To maximize both the time of individuals allocated to the SFA assignment and the quality of the delivered report, a structured approach is required. This structure is illustrated in Figure 4.16. This approach is similar to many consultancy models utilized within the industry, and in many ways Availability Management can be considered as providing via SFA a form of internal consultancy.

The above high-level structure is described briefly as follows.

- **Select opportunity:** prior to scheduling an SFA assignment, there needs to be agreement as to which IT service or technology is to be selected. It is recommended that an agreed number of assignments are scheduled per year within the Availability Plan and, if possible, the IT services are selected in advance as part of the proactive approach to Availability Management. Before commencing with the SFA, it is important that the assignment has a recognized sponsor from within the IT organization and/or the business and that they are involved and regularly updated with progress of the SFA activity. This ensures organizational visibility to the SFA and ensures recommendations are endorsed at a senior level within the organization.
- **Scope assignment:** this is to state explicitly what areas are and are not covered within the assignment. This is normally documented in Terms of Reference issued prior to the assignment.
- **Plan assignment:** the SFA assignment needs to be planned a number of weeks in advance of the assignment commencing, with an agreed project plan and a committed set of resources. The project should look at identifying improvement opportunities that benefit the user. It is therefore important that an end-to-end view of the data and Management Information

System (MIS) requirements is taken. The data and documentation should be collected from all areas and analysed from the user and business perspective. A 'virtual' SFA team should be formed from all relevant areas to ensure that all aspects and perspectives are considered. The size of the team should reflect the scope and complexity of the SFA assignment.

- **Build hypothesis:** this is a useful method of building likely scenarios, which can help the study team draw early conclusions within the analysis period. These scenarios can be built from discussing the forthcoming assignment with key roles, e.g. senior management and users, or by using the planning session to brainstorm the list from the assembled team. The completed hypotheses list should be documented and input to the analysis period to provide some early focus on the data and Management Information System (MIS) that match the individual scenarios. It should be noted that this approach also eliminates perceived issues, i.e. no data or MIS substantiates what is perceived to be a service issue.
- **Analyse data:** the number of individuals that form the SFA team dictates how to allocate specific analysis responsibilities. During this analysis period the hypotheses list should be used to help draw some early conclusions.
- **Interview key personnel:** it is essential that key business representatives and users are interviewed to ensure the business and user perspectives are captured. It is surprising how this dialogue can identify quick win opportunities, as often what the business views as a big issue can be addressed by a simple IT solution. Therefore these interviews should be initiated as soon as possible within the SFA assignment. The study team should also seek input from key individuals within the IT service provider organization to identify additional problem areas and possible solutions that can be fed back to the study team. The dialogue also helps capture those issues that are not easily visible from the assembled data and MIS reports.
- **Findings and conclusions:** after analysis of the data and MIS provided, interviews and continual revision of the hypothesis list, the study team should be in a position to start documenting initial findings and conclusions. It is recommended that the team meet immediately after the analysis period to share their individual findings and then take an aggregate view to form the draft findings and conclusions. It is important that all findings can be evidenced by facts gathered during the analysis. During this phase of the assignment, it may be necessary to validate finding(s) by additional analysis to ensure the SFA team can back up all findings with clear documented evidence.
- **Recommendations:** after all findings and conclusions have been validated, the SFA team should be in a position to formulate recommendations. In many cases, the recommendations to support a particular finding are straightforward and obvious. However, the benefit of bringing a cross-functional team together for the SFA assignment is to create an environment for innovative lateral-thinking approaches. The SFA assignment leader should facilitate this session with the aim of identifying recommendations that are practical and sustainable once implemented.
- **Report:** the final report should be issued to the sponsor with a management summary. Reporting styles are normally determined by the individual organizations. It is important that the report clearly shows where loss of availability is being incurred and how the recommendations address this. If the report contains many recommendations, an attempt should be made to quantify the availability benefit of each recommendation, together with the estimated effort to implement. This enables informed choices to be made on how to take the recommendations forward and how these should be prioritized and resourced.
- **Validation:** it is recommended that for each SFA, key measures that reflect the business and user perspectives prior to the assignment are captured and recorded as the 'before' view. As SFA recommendations are progressed, the positive impacts on availability should be captured to provide the 'after' view for comparative purposes. Where anticipated benefits have not been delivered, this should be investigated and remedial action taken. Having invested time and effort in completing the SFA assignment, it is important that the recommendations, once agreed by the sponsor, are then taken forward for implementation. The best mechanism for achieving this is by incorporating the recommendations as activities to be completed within the Availability Plan or the overall SIP.

The success of the SFA assignment as a whole should be monitored and measured to ensure its continued effectiveness.

Hints and tips

Consider categorizing the recommendations under the following headings:
DETECTION: Recommendations that, if implemented, will provide enhanced reporting of key indicators to ensure underlying IT service issues are detected early to enable a proactive response.
REDUCTION: Recommendations that, if implemented, will reduce or minimize the user impact from IT service interruption, e.g. recovery and/or restoration can be enhanced to reduce impact duration.
AVOIDANCE: Recommendations that, if implemented, will eliminate this particular cause of IT service interruption.

4.4.5.2 The proactive activities of Availability Management

The capability of the Availability Management process is positively influenced by the range and quality of proactive methods and techniques utilized by the process. The following activities are the proactive techniques and activities of the Availability Management process.

Identifying Vital Business Functions (VBFs)

The term Vital Business Function (VBF) is used to reflect the business critical elements of the business process supported by an IT service. The service may also support less critical business functions and processes, and it is important that the VBFs are recognized and documented to provide the appropriate business alignment and focus.

Designing for availability

The level of availability required by the business influences the overall cost of the IT service provided. In general, the higher the level of availability required by the business, the higher the cost. These costs are not just the procurement of the base IT technology and services required to underpin the IT infrastructure. Additional costs are incurred in providing the appropriate Service Management processes, systems management tools and high-availability solutions required to meet the more stringent availability requirements. The greatest level of availability should be included in the design of those services supporting the most critical of the VBFs.

When considering how the availability requirements of the business are to be met, it is important to ensure that the level of availability to be provided for an IT service is at the level actually required, and is affordable and cost-justifiable to the business. Figure 4.17 indicates the products and processes required to provide varying levels of availability and the cost implications.

Base product and components

The procurement or development of the base products, technology and components should be based on their capability to meet stringent availability and reliability requirements. These should be considered as the cornerstone of the availability design. The additional investment required to achieve even higher levels of availability will be wasted and availability levels not met if these base products and components are unreliable and prone to failure.

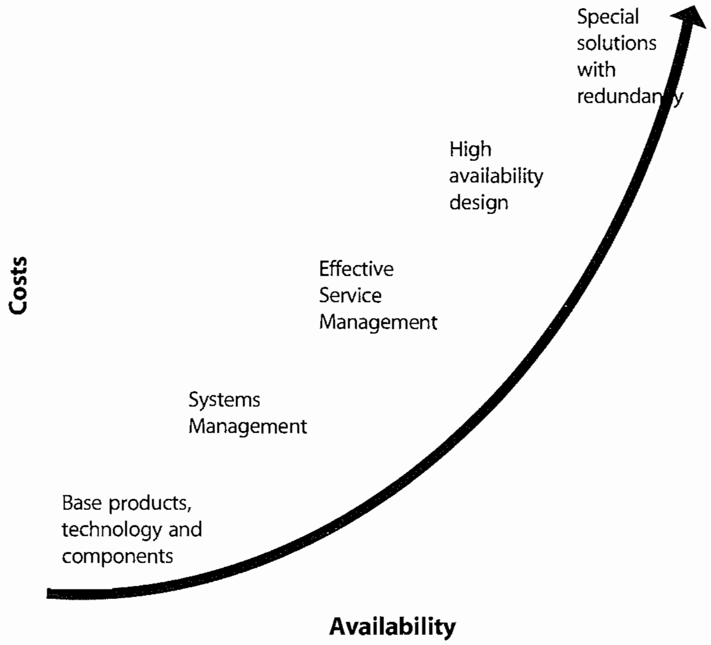


Figure 4.17 Relationship between levels of availability and overall costs

Systems management

Systems management should provide the monitoring, diagnostic and automated error recovery to enable fast detection and speedy resolution of potential and actual IT failure.

Service Management processes

Effective Service Management processes contribute to higher levels of availability. Processes such as Availability Management, Incident Management, Problem Management, Change Management, Configuration Management, etc. play a crucial role in the overall management of the IT service.

High-availability design

The design for high availability needs to consider the elimination of SPoFs and/or the provision of alternative components to provide minimal disruption to the business operation should an IT component failure occur. The design also needs to eliminate or minimize the effects of planned downtime to the business operation normally required to accommodate maintenance activity, the implementation of changes to the IT infrastructure or business application. Recovery criteria should define rapid recovery and IT service reinstatement as a key objective within the designing for recovery phase of design.

Special solutions with full redundancy

To approach continuous availability in the range of 100% requires expensive solutions that incorporate full mirroring or redundancy. Redundancy is the technique of improving availability by using duplicate components. For stringent availability requirements to be met, these need to be working autonomously in parallel. These solutions are not just restricted to the IT components, but also to the IT environments, i.e. data centres, power supplies, air conditioning and telecommunications.

Where new IT services are being developed, it is essential that Availability Management takes an early and participating design role in determining the availability requirements. This enables Availability Management to influence positively the IT infrastructure design to ensure that it can deliver the level of availability required. The importance of this participation early in the design of the IT infrastructure cannot be underestimated. There needs to be a dialogue between IT and the business to determine the balance between the business perception of the cost of unavailability and the exponential cost of delivering higher levels of availability.

As illustrated in Figure 4.17, there is a significant increase in costs when the business requirement is higher than the optimum level of availability that the IT infrastructure can deliver. These increased costs are driven by major redesign of the technology and the changing of requirements for the IT support organization.

It is important that the level of availability designed into the service is appropriate to the business needs, the criticality of the business processes being supported and the available budget. The business should be consulted early in the Service Design lifecycle so that the business availability needs of a new or enhanced IT service can be costed and agreed. This is particularly important where stringent availability requirements may require additional investment in Service Management processes, IT service and System Management tools, high-availability design and special solutions with full redundancy.

It is likely that the business need for IT availability cannot be expressed in technical terms. Availability Management therefore provides an important role in being able to translate the business and user requirements into quantifiable availability targets and conditions. This is an important input into the IT Service Design and provides the basis for assessing the capability of the IT design and IT support organization in meeting the availability requirements of the business.

The business requirements for IT availability should contain at least:

- A definition of the VBFs supported by the IT service
- A definition of IT service downtime, i.e. the conditions under which the business considers the IT service to be unavailable
- The business impact caused by loss of service, together with the associated risk
- Quantitative availability requirements, i.e. the extent to which the business tolerates IT service downtime or degraded service
- The required service hours, i.e. when the service is to be provided
- An assessment of the relative importance of different working periods
- Specific security requirements
- The service backup and recovery capability.

Once the IT technology design and IT support organization are determined, the service provider organization is then in a position to confirm if the availability requirements can be met. Where shortfalls are identified, dialogue with the business is required to present the cost options that exist to enhance the proposed design to meet the availability requirements. This enables the business to reassess if lower or higher levels of availability are required, and to understand the appropriate impact and costs associated with their decision.

Determining the availability requirements is likely to be an iterative process, particularly where there is a need to balance the business availability requirement against the associated costs. The necessary steps are:

- Determine the business impact caused by loss of service
- From the business requirements, specify the availability, reliability and maintainability requirements for the IT service and components supported by the IT support organization
- For IT services and components provided externally, identify the serviceability requirements
- Estimate the costs involved in meeting the availability, reliability, maintainability and serviceability requirements
- Determine, with the business, if the costs identified in meeting the availability requirements are justified
- Determine, from the business, the costs likely to be incurred from loss or degradation of service
- Where these are seen as cost-justified, define the availability, reliability, maintainability and serviceability requirements in agreements and negotiate into contracts.

Hints and tips

If costs are seen as prohibitive, either:

- Reassess the IT infrastructure design and provide options for reducing costs and assess the consequences on availability; or
- Reassess the business use and reliance on the IT service and renegotiate the availability targets within the SLA.

The SLM process is normally responsible for communicating with the business on how its availability requirements for IT services are to be met and negotiating the SLR/SLA for the IT Service Design process. Availability Management therefore provides important support and input to the both SLM and design processes during this period. While higher levels of availability can often be provided by investment in tools and technology, there is no justification for providing a higher level of availability than that needed and afforded by the business. The reality is that satisfying availability requirements is always a balance between cost and quality. This is where Availability Management can play a key role in optimizing availability of the IT Service Design to meet increasing availability demands while deferring an increase in costs.

Designing service for availability is a key activity driven by Availability Management. This ensures that the required level of availability for an IT service can be met. Availability Management needs to ensure that the design activity for availability looks at the task from two related, but distinct, perspectives:

- **Designing for availability:** this activity relates to the technical design of the IT service and the alignment of the internal and external suppliers required to meet the availability requirements of the business. It needs to cover all aspects of technology, including infrastructure, environment, data and applications.
- **Designing for recovery:** this activity relates to the design points required to ensure that in the event of an IT service failure, the service and its supporting components can be reinstated to enable normal business operations to resume as quickly as is possible. This again needs to cover all aspects of technology.

Additionally, the ability to recover quickly may be a crucial factor. In simple terms, it may not be possible or cost-justified to build a design that is highly resilient to failure(s). The ability to meet the availability requirements within the cost parameters may rely on the ability consistently to recover in a timely and effective manner. All aspects of availability should be considered in the Service Design process and should consider all stages within the Service Lifecycle.

The contribution of Availability Management within the design activities is to provide:

- The specification of the availability requirements for all components of the service
- The requirements for availability measurement points (instrumentation)
- The requirements for new/enhanced systems and Service Management
- Assistance with the IT infrastructure design
- The specification of the reliability, maintainability and serviceability requirements for components supplied by internal and external suppliers
- Validation of the final design to meet the minimum levels of availability required by the business for the IT service.

If the availability requirements cannot be met, the next task is to re-evaluate the Service Design and identify cost-justified design changes. Improvements in design to meet the availability requirements can be achieved by reviewing the capability of the technology to be deployed in the proposed IT design. For example:

- The exploitation of fault-tolerant technology to mask the impact of planned or unplanned component downtime
- Duplexing, or the provision of alternative IT infrastructure components to allow one component to take over the work of another component
- Improving component reliability by enhancing testing regimes
- Improved software design and development
- Improved processes and procedures
- Systems management enhancements/exploitation
- Improved externally supplied services, contracts or agreements
- Developing the capability of the people with more training.

Hints and tips

Consider documenting the availability design requirements and considerations for new IT services and making them available to the design and implementation functions. Longer term seek to mandate these requirements and integrate within the appropriate governance mechanisms that cover the introduction of new IT services.

Part of the activity of designing for availability must ensure that all business, data and information security requirements are incorporated within the Service Design. The overall aim of IT security is 'balanced security in depth', with justifiable controls implemented to ensure that the Information Security Policy is enforced and that continued IT services within secure parameters (i.e. confidentiality, integrity and availability) continue to operate. During the gathering of availability requirements for new IT services, it is important that requirements that cover IT security are defined. These requirements need to be applied within the design phase for the supporting technology. For many organizations, the approach taken to IT security is covered by an Information Security Policy owned and maintained by Information Security Management. In the execution of the security policy, Availability Management plays an important role in its operation for new IT services.

Where the business operation has a high dependency on IT service availability, and the cost of failure or loss of business reputation is considered not acceptable, the business may define stringent availability requirements. These factors may be sufficient for the business to justify the additional costs required to meet these more demanding levels of availability. Achieving agreed levels of availability begins with the design, procurement and/or development of good-quality products and components. However, these in isolation are unlikely to deliver the sustained levels of availability required. To achieve a consistent and sustained level of availability requires investment in and deployment of effective Service Management processes, systems management tools, high-availability design and ultimately special solutions with full mirroring or redundancy.

Designing for availability is a key activity, driven by Availability Management, which ensures that the stated availability requirements for an IT service can be met. However, Availability Management should also ensure that within this design activity there is focus on the design elements required to ensure that when IT services fail, the service can be reinstated to enable normal business operations to resume as quickly as is possible. 'Designing for recovery' may at first sound negative. Clearly good availability design is about avoiding failures and delivering, where possible, a fault-tolerant IT infrastructure. However, with this focus is too much reliance placed on technology, and has as much emphasis been placed on the fault tolerance aspects of the IT infrastructure? The reality is that failures will occur. The way the IT organization manages failure situations can have a positive effect on the perception of the business, customers and users of the IT services.

Key message

Every failure is an important 'moment of truth' – an opportunity to make or break your reputation with the business.

By providing focus on the 'designing for recovery' aspects of the overall availability, design can ensure that every failure is an opportunity to maintain and even enhance business and user satisfaction. To provide an effective 'design for recovery', it is important to recognize that both the business and the IT organization have needs that must be satisfied to enable an effective recovery from IT failure.

These are informational needs that the business requires to help them manage the impact of failure on their business and set expectation within the business, user community and their business customers. These are the skills, knowledge, processes, procedures and tools required to enable the technical recovery to be completed in an optimal time.

Hints and tips

Consider documenting the recovery design requirements and considerations for new IT services and make them available to the areas responsible for design and implementation. In the longer term, seek to mandate these requirements and integrate them within the appropriate governance mechanisms that cover the introduction of new IT services.

A key aim is to prevent minor incidents from becoming major incidents by ensuring the right people are involved early enough to avoid mistakes being made and to ensure the appropriate business and technical recovery procedures are invoked at the earliest opportunity. The instigation of these activities is the responsibility of the Incident Management process and a role of the Service Desk. To ensure business needs are met during major IT service failures, and to ensure the most optimal recovery, the Incident Management process and Service Desk need to have defined and to execute effective procedures for assessing and managing all incidents.

Key message

The above are not the responsibilities of Availability Management. However, the effectiveness of the Incident Management process and Service Desk can strongly influence the overall recovery period. The use of Availability Management methods and techniques to further optimize IT recovery may be the stimulus for subsequent continual improvement activities to the Incident Management process and the Service Desk.

In order to remain effective, the maintainability of IT services and components should be monitored, and their impact on the 'expanded incident lifecycle' understood, managed and improved.

Component Failure Impact Analysis

Component Failure Impact Analysis (CFIA) can be used to predict and evaluate the impact on IT service arising from component failures within the technology. The output from a CFIA can be used to identify where additional resilience should be considered to prevent or minimize the impact of component failure to the business operation and users. This is particularly important during the Service Design stage, where it is necessary to predict and evaluate the impact on IT service availability arising from component failures within the proposed IT Service Design. However, the technique can also be applied to existing services and infrastructure.

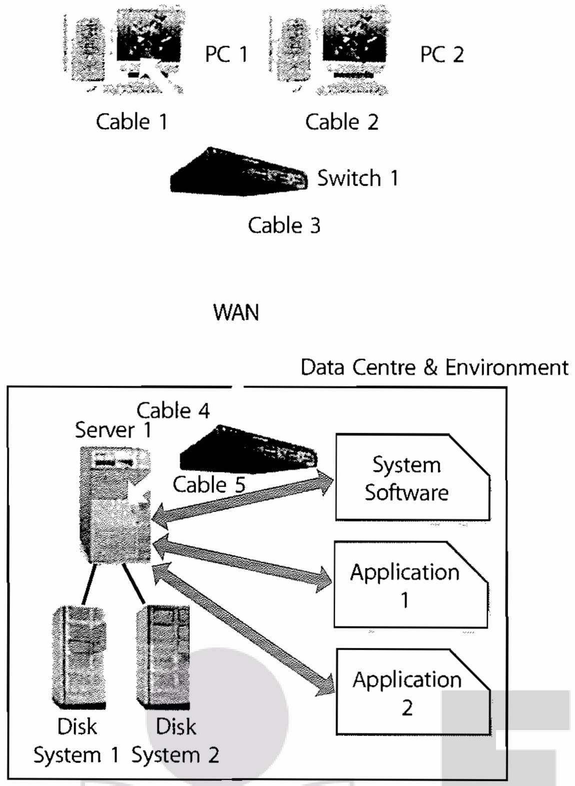
CFIA is a relatively simple technique that can be used to provide this information. IBM devised CFIA in the early 1970s, with its origins based on hardware design and configuration. However, it is recommended that CFIA be used in a much wider context to reflect the full scope of the IT infrastructure, i.e. hardware, network, software, applications, data centres and support staff. Additionally the technique can also be applied to identify impact and dependencies on IT support organization skills and competencies amongst staff supporting the new IT service. This activity is often completed in conjunction with ITSCM and possibly Capacity Management.

The output from a CFIA provides vital information to ensure that the availability and recovery design criteria for the new IT service is influenced to prevent or minimize the impact of failure to the business operation and users. CFIA achieves this by providing and indicating:

- SPoFs that can impact availability
- The impact of component failure on the business operation and users
- Component and people dependencies
- Component recovery timings
- The need to identify and document recovery options
- The need to identify and implement risk reduction measures.

The above can also provide the stimulus for input to ITSCM to consider the balance between recovery options and risk reduction measures, i.e. where the potential business impact is high there is a need to concentrate on high-availability risk reduction measures, i.e. increased resilience or standby systems.

Having determined the IT infrastructure configuration to be assessed, the first step is to create a grid with CIs on one axis and the IT services that have a dependency on the CI on the other, as illustrated in Figure 4.18. This information should be available from the CMS, or alternatively it can be built using documented configuration charts and SLAs.



CI	Service 1	Service 2
PC1	M	M
PC2	M	M
Cable 1	M	M
Cable 2	M	M
Switch 1	X	X
Cable 3	X	X
WAN	X	X
Cable 4	X	X
Switch 2	X	X
Cable 5	X	X
Data Centre	X	X
Server 1	X	X
Disk 1	A	A
Disk 2	A	A
System S/W	X	X
Application 1	X	
Application 2		X

Figure 4.18 Component Failure Impact Analysis

The next step is to perform the CFIA and populate the grid as follows:

- Leave a blank when a failure of the CI does not impact the service in any way
- Insert an 'X' when the failure of the CI causes the IT service to be inoperative
- Insert an 'A' when there is an alternative CI to provide the service
- Insert an 'M' when there is an alternative CI, but the service requires manual intervention to be recovered.

Having built the grid, CIs that have a large number of Xs are critical to many services and can result in high impact should the CI fail. Equally, IT services having high counts of Xs are complex and are vulnerable to failure. This basic approach to CFIA can provide valuable information in quickly identifying SPoFs, IT services at risk from CI failure and what alternatives are available should CIs fail. It should also be used to assess the existence and validity of recovery procedures for the selected CIs. The above example assumes common infrastructure supporting multiple IT services. The same approach can be used for a single IT service by mapping the component CIs against the VBFs and users supported by each component, thus understanding the impact of a component failure on the business and user. The approach can also be further refined and developed to include and develop 'component availability weighting' factors that can be used to assess and calculate the overall effect of the component failure on the total service availability.

To undertake an advanced CFIA requires the CFIA matrix to be expanded to provide additional fields required for the more detailed analysis. This could include fields such as:

- **Component availability weighting:** a weighting factor appropriate to the impact of failure of the component on the total service availability. For example, if the failure of a switch can cause 2,000 users to lose service out of a total service user base of 10,000, then the weighting factor should be 0.2, or 20%
- **Probability of failure:** this can be based on the reliability of the component as measured by the Mean Time Between Failures (MTBF) information if available or on the current trends. This can be expressed as a low/medium/high indicator or as a numeric representation
- **Recovery time:** this is the estimated recovery time to recover the CI. This can be based on recent recovery timings, recovery information from disaster recovery testing or a scheduled test recovery

- **Recovery procedures:** this is to verify that up-to-date recovery procedures are available for the CI
- **Device independence:** where software CIs have duplex files to provide resilience, this is to ensure that file placements have been verified as being on separate hardware disk configurations. This also applies to power supplies – it should be verified that alternate power supplies are connected correctly
- **Dependency:** this is to show any dependencies between CIs. If one CI failed, there could be an impact on other CIs – for example, if the security CI failed, the operating system might prevent tape processing.

Single Point of Failure analysis

A Single Point of Failure (SPoF) is any component within the IT infrastructure that has no backup or fail-over capability, and has the potential to cause disruption to the business, customers or users when it fails. It is important that no unrecognized SPoFs exist within the IT infrastructure design or the actual technology, and that they are avoided wherever possible.

The use of SPoF analysis or CFIA as techniques to identify SPoFs is recommended. SPoF and CFIA analysis exercises should be conducted on a regular basis, and wherever SPoFs are identified, CFIA can be used to identify the potential business, customer or user impact and help determine what alternatives can or should be considered to cater for this weakness in the design or the actual infrastructure. Countermeasures should then be implemented wherever they are cost-justifiable. The impact and disruption caused by the potential failure of the SPoF should be used to cost-justify its implementation.

Fault Tree Analysis

Fault Tree Analysis (FTA) is a technique that can be used to determine the chain of events that causes a disruption to IT services. FTA, in conjunction with calculation methods, can offer detailed models of availability. This can be used to assess the availability improvement that can be achieved by individual technology component design options. Using FTA:

- Information can be provided that can be used for availability calculations
- Operations can be performed on the resulting fault tree; these operations correspond with design options
- The desired level of detail in the analysis can be chosen.

FTA makes a representation of a chain of events using Boolean notation. Figure 4.19 gives an example of a fault tree.

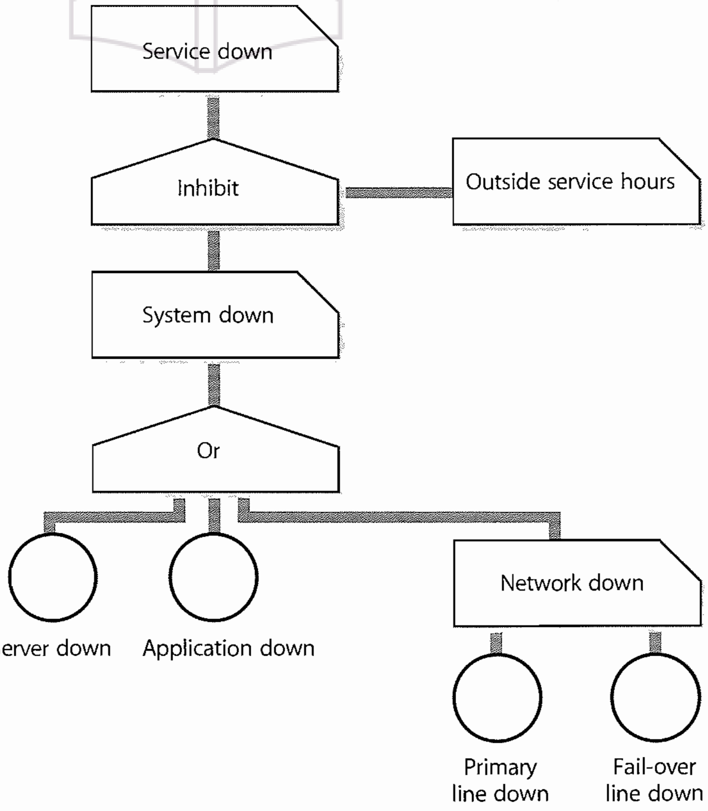


Figure 4.19 Example Fault Tree Analysis

Essentially FTA distinguishes the following events:

- **Basic events** – terminal points for the fault tree, e.g. power failure, operator error. Basic events are not investigated in great depth. If basic events are investigated in further depth, they automatically become resulting events.
- **Resulting events** – intermediate nodes in the fault tree, resulting from a combination of events. The highest point in the fault tree is usually a failure of the IT service.
- **Conditional events** – events that only occur under certain conditions, e.g. failure of the air-conditioning equipment only affects the IT service if equipment temperature exceeds the serviceable values.
- **Trigger events** – events that trigger other events, e.g. power failure detection equipment can trigger automatic shutdown of IT services.

These events can be combined using logic operators, i.e.:

- **AND-gate** – the resulting event only occurs when all input events occur simultaneously
- **OR-gate** – the resulting event occurs when one or more of the input events occurs
- **Exclusive OR-gate** – the resulting event occurs when one and only one of the input events occurs
- **Inhibit gate** – the resulting event only occurs when the input condition is not met.

This is the basic FTA technique. This technique can also be refined, but complex FTA and the mathematical evaluation of fault trees are beyond the scope of this publication.

Modelling

To assess if new components within a design can match the stated requirements, it is important that the testing regime instigated ensures that the availability expected can be delivered. Simulation, modelling or load testing tools to generate the expected user demand for the new IT service should be seriously considered to ensure components continue to operate under anticipated volume and stress conditions.

Modelling tools are also required to forecast availability and to assess the impact of changes to the IT infrastructure. Inputs to the modelling process include descriptive data of the component reliability, maintainability and serviceability. A spreadsheet package to perform calculations is usually sufficient. If more detailed and accurate data is required, a more complex modelling tool may need to be developed or acquired. The lack of readily available availability modelling tools in the marketplace may require such a tool to be developed and maintained 'in-house', but this is a very expensive and time-consuming activity that should only be considered where the investment can be justified. Unless there is a clearly perceived benefit from such a development and the ongoing maintenance costs, the use of existing tools and spreadsheets should be sufficient. However, some System Management tools do provide modelling capability and can provide useful information on trending and forecasting availability needs.

Risk Analysis and Management

To assess the vulnerability of failure within the configuration and capability of the IT service and support organization it is recommended that existing or proposed IT infrastructure, service configurations, Service Design and supporting organization (internal and external suppliers) are subject to formal Risk Analysis and Management exercises. Risk Analysis and Management is a technique that can be used to identify and quantify risks and justifiable countermeasures that can be implemented to protect the availability of IT systems. The identification of risks and the provision of justified countermeasures to reduce or eliminate the threats posed by such risks can play an important role in achieving the required levels of availability for a new or enhanced IT service. Risk Analysis should be undertaken during the design phase for the IT technology and service to identify:

- Risks that may incur unavailability for IT components within the technology and Service Design
- Risks that may incur confidentiality and/or integrity exposures within the IT technology and Service Design.

Most risk assessment and management methodologies involve the use of a formal approach to the assessment of risk and the subsequent mitigation of risk with the implementation of subsequent cost-justifiable counter-measures, as illustrated in Figure 4.20.

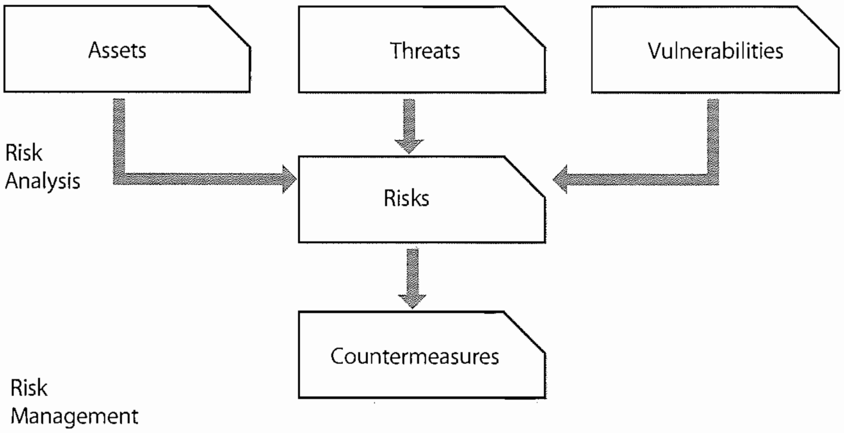


Figure 4.20 Risk Analysis and Management

Risk Analysis involves the identification and assessment of the level (measure) of the risks calculated from the assessed values of assets and the assessed levels of threats to, and vulnerabilities of, those assets. Risk is also determined to a certain extent by its acceptance. Some organizations and businesses may be more willing to accept risk whereas others cannot.

Risk management involves the identification, selection and adoption of countermeasures justified by the identified risks to assets in terms of their potential impact on services if failure occurs, and the reduction of those risks to an acceptable level. Risk management is an activity that is associated with many other activities, especially ITSCM, Security Management and Service Transition. All of these risk assessment exercises should be coordinated rather than being separate activities.

This approach, when applied via a formal method, ensures coverage is complete, together with sufficient confidence that:

- All possible risks and countermeasures have been identified
- All vulnerabilities have been identified and their levels accurately assessed
- All threats have been identified and their levels accurately assessed
- All results are consistent across the broad spectrum of the technology reviewed
- All expenditure on selected countermeasures can be justified.

Formal Risk Analysis and Management methods are now an important element in the overall design and provision of IT services. The assessment of risk is often based on the probability and potential impact of an event occurring. Counter-measures are implemented wherever they are cost-justifiable, to reduce the impact of an event, or the probability of an event occurring, or both.

Management of Risk (M_o_R) provides an alternative generic framework for the management of risk across all parts of an organization – strategic, programme, project and operational. It incorporates all the activities required to identify and control the exposure to any type of risk, positive or negative, that may have an impact on the achievement of your organization's business objectives.

M_o_R provides a framework that is tried, tested and effective to help you eliminate – or manage – the risks involved in reaching your goals. M_o_R adopts a systematic application of principles, approach and processes to the task of identifying, assessing and then planning and implementing risk responses. Guidance stresses a collaborative approach and focuses on the following key elements:

- Developing a framework that is transparent, repeatable and adaptable
- Clearly communicating the policy and its benefits to all staff
- Nominating key individuals in senior management to 'own' risk management initiatives and ensure they move forwards
- Ensuring the culture engages with and supports properly considered risk, including innovation
- Embedding risk management systems in management and applying them consistently
- Ensuring that risk management supports objectives – rather than vice versa
- Explicitly assessing the risks involved in working with other organizations
- Adopting a no-blame approach to monitoring and reviewing risk assessment activity.

Availability testing schedule

A key deliverable from the Availability Management process is the 'availability testing schedule'. This is a schedule for the regular testing of all availability mechanisms. Some availability mechanisms, such as 'load balancing', 'mirroring' and 'grid computing', are used in the provision of normal service on a day-by-day basis; others are used on a fail-over or manual reconfiguration basis. It is essential, therefore, that all availability mechanisms are tested in a regular and scheduled manner to ensure that when they are actually needed for real they work. This schedule needs to be maintained and widely circulated so that all areas are aware of its content and so that all other proposed activities can be synchronized with its content, such as:

- The change schedule
- Release plans and the release schedule
- All transition plans, projects and programmes
- Planned and preventative maintenance schedules
- The schedule for testing IT service continuity and recovery plans
- Business plans and schedules.

Planned and preventative maintenance

All IT components should be subject to a planned maintenance strategy. The frequency and levels of maintenance required varies from component to component, taking into account the technologies involved, criticality and the potential business benefits that may be introduced. Planned maintenance activities enable the IT support organization to provide:

- Preventative maintenance to avoid failures
- Planned software or hardware upgrades to provide new functionality or additional capacity
- Business requested changes to the business applications
- Implementation of new technology and functionality for exploitation by the business.

The requirement for planned downtime clearly influences the level of availability that can be delivered for an IT service, particularly those that have stringent availability requirements. In determining the availability requirements for a new or enhanced IT service, the amount of downtime and the resultant loss of income required for planned maintenance may not be acceptable to the business. This is becoming a growing issue in the area of 24 x 7 service operation. In these instances, it is essential that continuous operation is a core design feature to enable maintenance activity to be performed without impacting the availability of IT services.

Where the required service hours for IT services are less than 24 hours per day and/or seven days per week, it is likely that the majority of planned maintenance can be accommodated without impacting IT service availability. However, where the business needs IT services available on a 24-hour and seven-day basis, Availability Management needs to determine the most effective approach in balancing the requirements for planned maintenance against the loss of service to the business. Unless mechanisms exist to allow continuous operation, scheduled downtime for planned maintenance is essential if high levels of availability are to be achieved and sustained. For all IT services, there should logically be a 'low-impact' period for the implementation of maintenance. Once the requirements for managing scheduled maintenance have been defined and agreed, these should be documented as a minimum in:

- SLAs
- OLAs
- Underpinning contracts
- Change Management schedules
- Release and Deployment Management schedules.

Hints and tips

Availability Management should ensure that building in preventative maintenance is one of the prime design considerations for a '24 x 7' IT service.

The most appropriate time to schedule planned downtime is clearly when the impact on the business and its customers is least. This information should be provided initially by the business when determining the availability requirements. For an existing IT service, or once the new service has been established, monitoring of business and customer transactions helps establish the hours when IT service usage is at its lowest. This should determine the most appropriate time for the component(s) to be removed for planned maintenance activity.

To accommodate the individual component requirements for planned downtime while balancing the IT service availability requirements of the business provides an opportunity to consider scheduling planned maintenance to multiple components concurrently. The benefit of this approach is that the number of service disruptions required to meet the maintenance requirements is reduced. While this approach has benefits, there are potential risks that need to be assessed. For example:

- The capability of the IT support organization to coordinate the concurrent implementation of a high number of changes

- The ability to perform effective problem determination where the IT service is impacted after the completion of multiple changes
- The impact of change dependency across multiple components where back-out of a failed change requires multiple changes to be removed.

The effective management of planned downtime is an important contribution in meeting the required levels of availability for an IT service. Where planned downtime is required on a cyclic basis to an IT component(s), the time that the component is unavailable to enable the planned maintenance activity to be undertaken should be defined and agreed with the internal or external supplier. This becomes a stated objective that can be formalized, measured and reported. All planned maintenance should be scheduled, managed and controlled to ensure that the individual objectives and time slots are not exceeded and to ensure that activities are coordinated with all other schedules of activity to minimize clashes and conflict (e.g. change and release schedules, testing schedules.) In addition they provide an early warning during the maintenance activity of the time allocated to the planned outage duration being breached. This can enable an early decision to be made on whether the activity is allowed to complete with the potential to further impact service or to abort the activity and instigate the back-out plan. Planned downtime and performance against the stated objectives for each component should be recorded and used in service reporting.

Production of the Projected Service Outage (PSO) document

Availability Management should produce and maintain the PSO document. This document consists of any variations from the service availability agreed within SLAs. This should be produced based on input from:

- The change schedule
- The release schedules
- Planned and preventative maintenance schedules
- Availability testing schedules
- ITSCM and Business Continuity Management testing schedules.

The PSO contains details of all scheduled and planned service downtime within the agreed service hours for all services. These documents should be agreed with all the appropriate areas and representatives of both the business and IT. Once the PSO has been agreed, the Service Desk should ensure that it is communicated to all relevant parties so that everyone is made aware of any additional, planned service downtime.

Continual review and improvement

Changing business needs and customer demand may require the levels of availability provided for an IT service to be reviewed. Such reviews should form part of the regular service reviews with the business undertaken by SLM. Other input should also be considered on a regular basis from ITSCM, particularly from the updated Business Impact Analysis and Risk Analysis exercises. The criticality of services will often change and it is important that the design and the technology supporting such services is regularly reviewed and improved by Availability Management to ensure that the change of importance in the service is reflected within a revised design and supporting technology. Where the required levels of availability are already being delivered, it may take considerable effort and incur significant cost to achieve a small incremental improvement within the level of availability.

A key activity for Availability Management is continually to look at opportunities to optimize the availability of the IT infrastructure in conjunction with Continual Service Improvement activities. The benefits of this regular review approach are that, sometimes, enhanced levels of availability may be achievable, but with much lower costs. The optimization approach is a sensible first step to delivering better value for money. A number of Availability Management techniques can be applied to identify optimization opportunities. It is recommended that the scope should not be restricted to the technology, but also include a review of both the business process and other end-to-end business-owned responsibilities. To help achieve these aims, Availability Management needs to be recognized as a leading influence over the IT service provider organization to ensure continued focus on availability and stability of the technology.

Availability Management can provide the IT support organization with a real business and user perspective on how deficiencies within the technology and the underpinning process and procedure impact on the business operation and ultimately their customers. The use of business-driven metrics can demonstrate this impact in real terms and, importantly, also help quantify the benefits of improvement opportunities. Availability Management can play an important role in helping the IT service provider organization recognize where it can add value by exploiting its technical skills and competencies in an availability context. The continual improvement technique can be used by Availability Management to harness this technical capability. This can be used with either small groups of technical staff or a wider group within a workshop or SFA environment.

The impetus to improve availability comes from one or more of the following:

- The inability for existing or new IT services to meet SLA targets on a consistent basis
- Period(s) of IT service instability resulting in unacceptable levels of availability
- Availability measurement trends indicating a gradual deterioration in availability
- Unacceptable IT service recovery and restoration times
- Requests from the business to increase the level of availability provided
- Increasing impact on the business and its customers of IT service failures as a result of growth and/or increased business priorities or functionality
- A request from SLM to improve availability as part of an overall SIP
- Availability Management monitoring and trend analysis.

Availability Management should take a proactive role in identifying and progressing cost-justified availability improvement opportunities within the Availability Plan. The ability to do this places reliance on having appropriate and meaningful availability measurement and reporting. To ensure availability improvements deliver benefits to the business and users, it is important that measurement and reporting reflects not just IT component availability but also availability from a business operation and user perspective.

Where the business has a requirement to improve availability, the process and techniques to reassess the technology and IT service provider organization capability to meet these enhanced requirements should be followed. An output of this activity is enhanced availability and recovery design criteria. To satisfy the business requirement for increased levels of availability may require additional financial investment to enhance the underpinning technology and/or extend the services provided by the IT service provider organization. It is important that any additional investment to improve the levels of availability delivered can be cost-justified. Determining the cost of unavailability as a result of IT failure(s) can help support any financial investment decision in improving availability.

4.4.6 Triggers, inputs, outputs and interfaces

Many events may trigger Availability Management activity. These include:

- New or changed business needs or new or changed services
- New or changed targets within agreements, such as SLRs, SLAs, OLAs or contracts
- Service or component breaches, availability events and alerts, including threshold events, exception reports
- Periodic activities such as reviewing, revising or reporting
- Review of Availability Management forecasts, reports and plans
- Review and revision of business and IT plans and strategies
- Review and revision of designs and strategies
- Recognition or notification of a change of risk or impact of a business process or VBF, an IT service or component
- Request from SLM for assistance with availability targets and explanation of achievements.

The key interfaces that Availability Management has with other processes are:

- **Incident and Problem Management:** in providing assistance with the resolution and subsequent justification and correction of availability incidents and problems
- **Capacity Management:** with the provision of resilience and spare capacity
- **IT Service Continuity Management:** with the assessment of business impact and risk and the provision of resilience, fail-over and recovery mechanisms
- **Service Level Management:** assistance with the determining of availability targets and the investigation and resolution of service and component breaches.

4.4.6.1 Inputs

A number of sources of information are relevant to the Availability Management process. Some of these are as follows:

- **Business information:** from the organization's business strategy, plans and financial plans, and information on their current and future requirements, including the availability requirements for new or enhanced IT services
- **Business impact information:** from BIAs and assessment of VBFs underpinned by IT services
- **Previous Risk Analysis** and Assessment reports and a risk register
- **Service information:** from the Service Portfolio and the Service Catalogue,
- **Service information:** from the SLM process, with

details of the services from the Service Portfolio and the Service Catalogue, service level targets within SLAs and SLRs, and possibly from the monitoring of SLAs, service reviews and breaches of the SLAs

- **Financial information:** from Financial Management, the cost of service provision, the cost of resources and components
- **Change and release information:** from the Change Management process with a Change Schedule, the Release Schedule from Release Management and a need to assess all changes for their impact on service availability
- **Configuration Management:** containing information on the relationships between the business, the services, the supporting services and the technology
- **Service targets:** from SLAs, SLRs, OLAs and contracts
- **Component information:** on the availability, reliability and maintainability requirements for the technology components that underpin IT service(s)
- **Technology information:** from the CMS on the topology and the relationships between the components and the assessment of the capabilities of new technology
- **Past performance:** from previous measurements, achievements and reports and the Availability Management Information System (AMIS)
- **Unavailability and failure information:** from incidents and problems.

4.4.6.2 Outputs

The outputs produced by Availability Management should include:

- The Availability Management Information System (AMIS)
- The Availability Plan for the proactive improvement of IT services and technology
- Availability and recovery design criteria and proposed service targets for new or changed services
- Service availability, reliability and maintainability reports of achievements against targets, including input for all service reports
- Component availability, reliability and maintainability reports of achievements against targets
- Revised risk analysis reviews and reports and an updated risk register
- Monitoring, management and reporting requirements for IT services and components to ensure that deviations in availability, reliability and maintainability are detected, actioned, recorded and reported
- An Availability Management test schedule for testing all availability, resilience and recovery mechanisms
- The planned and preventative maintenance schedules
- The Projected Service Outage (PSO) in conjunction with Change and Release Management
- Details of the proactive availability techniques and measures that will be deployed to provide additional resilience to prevent or minimize the impact of component failures on the IT service availability
- Improvement actions for inclusion within the SIP.

4.4.7 Key Performance Indicators

Many KPIs can be used to measure the effectiveness and efficiency of Availability Management, including the following examples:

Manage availability and reliability of IT service:

- Percentage reduction in the unavailability of services and components
- Percentage increase in the reliability of services and components
- Effective review and follow-up of all SLA, OLA and underpinning contract breaches
- Percentage improvement in overall end-to-end availability of service
- Percentage reduction in the number and impact of service breaks
- Improvement in the MTBF (Mean Time Between Failures)
- Improvement in the MTBSI (Mean Time Between Systems Incidents)
- Reduction in the MTRS (Mean Time to Restore Service).

Satisfy business needs for access to IT services:

- Percentage reduction in the unavailability of services
- Percentage reduction of the cost of business overtime due to unavailable IT
- Percentage reduction in critical time failures, e.g. specific business peak and priority availability needs are planned for
- Percentage improvement in business and users satisfied with service (by CSS results).

Availability of IT infrastructure achieved at optimum costs:

- Percentage reduction in the cost of unavailability
- Percentage improvement in the Service Delivery costs
- Timely completion of regular Risk Analysis and system review
- Timely completion of regular cost-benefit analysis

established for infrastructure Component Failure Impact Analysis (CFIA)

- Percentage reduction in failures of third-party performance on MTRS/MTBF against contract targets
- Reduced time taken to complete (or update) a Risk Analysis
- Reduced time taken to review system resilience
- Reduced time taken to complete an Availability Plan
- Timely production of management reports
- Percentage reduction in the incidence of operational reviews uncovering security and reliability exposures in application designs.

4.4.8 Information Management

The Availability Management process should maintain an AMIS that contains all of the measurements and information required to complete the Availability Management process and provide the appropriate information to the business on the level of IT service provided. This information, covering services, components and supporting services, provides the basis for regular, ad hoc and exception availability reporting and the identification of trends within the data for the instigation of improvement activities. These activities and the information contained within the AMIS provide the basis for developing the content of the Availability Plan.

In order to provide structure and focus to a wide range of initiatives that may need to be undertaken to improve availability, an Availability Plan should be formulated and maintained. The Availability Plan should have aims, objectives and deliverables and should consider the wider issues of people, processes, tools and techniques as well as having a technology focus. In the initial stages it may be aligned with an implementation plan for Availability Management, but the two are different and should not be confused. As the Availability Management process matures, the plan should evolve to cover the following:

- Actual levels of availability versus agreed levels of availability for key IT services. Availability measurements should always be business- and customer-focused and report availability as experienced by the business and users.
- Activities being progressed to address shortfalls in availability for existing IT services. Where investment decisions are required, options with associated costs and benefits should be included.
- Details of changing availability requirements for existing IT services. The plan should document the options available to meet these changed requirements. Where investment decisions are required, the associated costs of each option should be included.
- Details of the availability requirements for forthcoming new IT services. The plan should document the options available to meet these new requirements. Where investment decisions are required, the associated costs of each option should be included.
- A forward-looking schedule for the planned SFA assignments.
- Regular reviews of SFA assignments should be completed to ensure that the availability of technology is being proactively improved in conjunction with the SIP.
- A technology futures section to provide an indication of the potential benefits and exploitation opportunities that exist for planned technology upgrades. Anticipated availability benefits should be detailed, where possible based on business-focused measures, in conjunction with Capacity Management. The effort required to realize these benefits where possible should also be quantified.

During the production of the Availability Plan, it is recommended that liaison with all functional, technical and process areas is undertaken. The Availability Plan should cover a period of one to two years, with a more detailed view and information for the first six months. The plan should be reviewed regularly, with minor revisions every quarter and major revisions every half year. Where the technology is only subject to a low level of change, this may be extended as appropriate.

It is recommended that the Availability Plan is considered complementary to the Capacity Plan and Financial Plan, and that publication is aligned with the capacity and business budgeting cycle. If a demand is foreseen for high levels of availability that cannot be met due to the constraints of the existing IT infrastructure or budget, then exception reports may be required for the attention of both senior IT and business management.

In order to facilitate the production of the Availability Plan, Availability Management may wish to consider having its own database repository. The AMIS can be utilized to record and store selected data and information required to support key activities such as report generation, statistical analysis and availability forecasting and planning. The AMIS should be the main repository for the recording of IT availability metrics, measurements, targets and documents, including the Availability Plan, availability measurements, achievement reports, SFA assignment reports, design criteria, action plans and testing schedules.

Hints and tips

Be pragmatic, define the initial tool requirements and identify what is already deployed that can be used and shared to get started as quickly as possible. Where basic tools are not already available, work with the other IT service and systems management processes to identify common requirements with the aim of selecting shared tools and minimizing costs. The AMIS should address the specific reporting needs of Availability Management not currently provided by existing repositories and integrate with them and their contents.

4.4.9 Challenges, Critical Success Factors and risks

Availability Management faces many challenges, but probably the main challenge is to actually meet the expectations of the customers, the business and senior management. These expectations are that services will always be available not just during their agreed service hours, but that all services will be available on a 24-hour, 365-day basis. When they aren't, it is assumed that they will be recovered within minutes. This is only the case when the appropriate level of investment and design has been applied to the service, and this should only be made where the business impact justifies that level of investment. However, the message needs to be publicized to all customers and areas of the business, so that when services do fail they have the right level of expectation on their recovery. It also means that Availability Management must have access to the right level of quality information on the current business need for IT services and its plans for the future. This is another challenge faced by many Availability Management processes.

Another challenge facing Availability Management is the integration of all of the availability data into an integrated set of information (AMIS) that can be analysed in a consistent manner to provide details on the availability of all services and components. This is particularly challenging when the information from the different technologies is often provided by different tools in differing formats.

Yet another challenge facing Availability Management is convincing the business and senior management of the investment needed in proactive availability measures. Investment is always recognized once failures have occurred, but by then it is really too late. Persuading businesses and customers to invest in resilience to avoid the possibility of failures that may happen is a difficult challenge. Availability Management should work closely with Service Continuity Management, Security Management and Capacity Management in producing the justifications necessary to secure the appropriate investment.

The main CSFs for the Availability Management process are:

- Manage availability and reliability of IT service
- Satisfy business needs for access to IT services
- Availability of IT infrastructure, as documented in SLAs, provided at optimum costs.

Some of the major risks associated with Availability Management include:

- A lack of commitment from the business to the Availability Management process
- A lack of commitment from the business and a lack of appropriate information on future plans and strategies
- A lack of senior management commitment or a lack of resources and/or budget to the Availability Management process
- The reporting processes become very labour-intensive
- The processes focus too much on the technology and not enough on the services and the needs of the business
- The Availability Management information (AMIS) is maintained in isolation and is not shared or consistent with other process areas, especially ITSCM, Security Management and Capacity Management. This investment is particularly important when considering the necessary service and component backup and recovery tools, technology and processes to meet the agreed needs.

4.5 IT SERVICE CONTINUITY MANAGEMENT

4.5.1 Purpose/goal/objective

> 'The goal of ITSCM is to support the overall Business Continuity Management process by ensuring that the required IT technical and service facilities (including computer systems, networks, applications, data repositories, telecommunications, environment, technical support and Service Desk) can be resumed within required, and agreed, business timescales.'

As technology is a core component of most business processes, continued or high availability of IT is critical to the survival of the business as a whole. This is achieved by introducing risk reduction measures and recovery options. Like all elements of ITSM, successful implementation of ITSCM can only be achieved with senior management commitment and the support of all members of the organization. Ongoing maintenance of the recovery capability is essential if it is to remain effective. The purpose of ITSCM is to maintain the necessary ongoing recovery capability within the IT services and their supporting components.

The objectives of ITSCM are to:

- Maintain a set of IT Service Continuity Plans and IT recovery plans that support the overall Business Continuity Plans (BCPs) of the organization
- Complete regular Business Impact Analysis (BIA) exercises to ensure that all continuity plans are maintained in line with changing business impacts and requirements
- Conduct regular Risk Analysis and Management exercises, particularly in conjunction with the business and the Availability Management and Security Management processes, that manage IT services within an agreed level of business risk
- Provide advice and guidance to all other areas of the business and IT on all continuity- and recovery-related issues
- Ensure that appropriate continuity and recovery mechanisms are put in place to meet or exceed the agreed business continuity targets
- Assess the impact of all changes on the IT Service Continuity Plans and IT recovery plans
- Ensure that proactive measures to improve the availability of services are implemented wherever it is cost-justifiable to do so
- Negotiate and agree the necessary contracts with suppliers for the provision of the necessary recovery capability to support all continuity plans in conjunction with the Supplier Management process.

4.5.2 Scope

ITSCM focuses on those events that the business considers significant enough to be considered a disaster. Less significant events will be dealt with as part of the Incident Management process. What constitutes a disaster will vary from organization to organization. The impact of a loss of a business process, such as financial loss, damage to reputation or regulatory breach, is measured through a BIA exercise, which determines the minimum critical requirements. The specific IT technical and service requirements are supported by ITSCM. The scope of ITSCM within an organization is determined by the organizational structure, culture and strategic direction (both business and technology) in terms of the services provided and how these develop and change over time.

ITSCM primarily considers the IT assets and configurations that support the business processes. If (following a disaster) it is necessary to relocate to an alternative working location, provision will also be required for items such as office and personnel accommodation, copies of critical paper records, courier services and telephone facilities to communicate with customers and third parties.

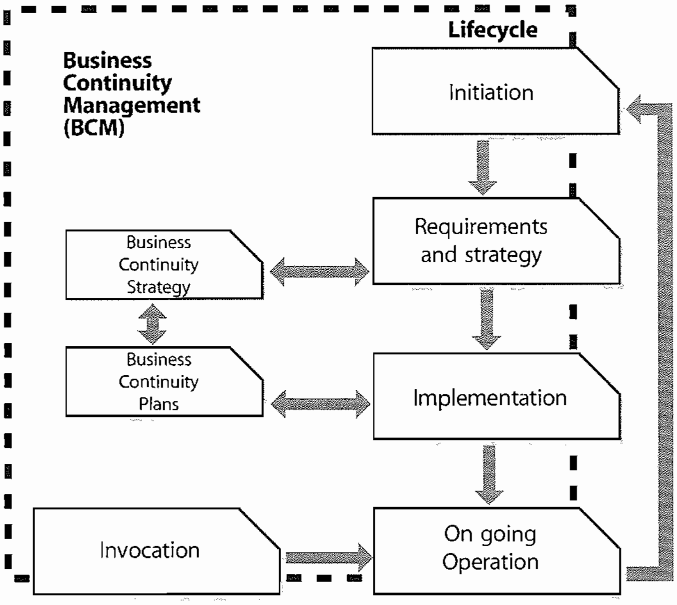
The scope will need to take into account the number and location of the organization's offices and the services performed in each.

ITSCM does not usually directly cover longer-term risks such as those from changes in business direction, diversification, restructuring, major competitor failure, and so on. While these risks can have a significant impact on IT service elements and their continuity mechanisms, there is usually time to identify and evaluate the risk and include risk mitigation through changes or shifts in business and IT strategies, thereby becoming part of the overall business and IT Change Management programme.

Similarly, ITSCM does not usually cover minor technical faults (for example, non critical disk failure), unless there is a possibility that the impact could have a major impact on the business. These risks would be expected to be covered mainly through the Service Desk and the Incident Management process, or resolved through the planning associated with the processes of Availability Management, Problem Management, Change Management, Configuration Management and 'business as usual' operational management.

The ITSCM process includes:

- The agreement of the scope of the ITSCM process and the policies adopted.
- Business Impact Analysis (BIA) to quantify the impact loss of IT service would have on the business.
- Risk Analysis (RA) – the risk identification and risk assessment to identify potential threats to continuity and the likelihood of the threats becoming reality. This also includes taking measures to manage the identified threats where this can be cost-justified.
- Production of an overall ITSCM strategy that must be integrated into the BCM strategy. This can be produced following the two steps identified above, and is likely to include elements of risk reduction as well as selection of appropriate and comprehensive recovery options.



Key activities

- Policy setting
- Scope
- Initiate a project

- Business Impact Analysis
- Risk Assessment
- IT Service Continuity Strategy

- Develop IT Service Continuity Plans
- Develop IT plans, recovery plans and procedures
- Organisation Planning
- Testing strategy

- Education, awareness and training
- Review and audit
- Testing
- Change Management

Figure 4.21 Lifecycle of Service Continuity Management

- Production of an ITSCM plan, which again must be integrated with the overall BCM plans.
- Testing of the plans.
- Ongoing operation and maintenance of the plans.

4.5.3 Value to the business

ITSCM provides an invaluable role in supporting the Business Continuity Planning process. In many organizations, ITSCM is used to raise awareness of continuity and recovery requirements and is often used to justify and implement a Business Continuity Planning process and Business Continuity Plans. The ITSCM should be driven by business risk as identified by Business Continuity Planning, and ensures that the recovery arrangements for IT services are aligned to identified business impacts, risks and needs.

4.5.4 Policies/principles/basic concepts

A lifecycle approach should be adopted to the setting up and operation of an ITSCM process. Figure 4.21 shows the lifecycle of ITSCM, from initiation through to continual assurance that the protection provided by the plan is current and reflects all changes to services and service levels. ITSCM is a cyclic process through the lifecycle to ensure that once service continuity and recovery plans have been developed they are kept aligned with Business Continuity Plans (BCPs) and business priorities. Figure 4.21 also shows the role played within the ITSCM process of BCM.

Initiation and requirements stages are principally BCM activities. ITSCM should only be involved in these stages to support the BCM activities and to understand the relationship between the business processes and the impacts caused on them by loss of IT service. As a result of these initial BIA and Risk Analysis activities, BCM should produce a Business Continuity Strategy, and the first real ITSCM task is to produce an ITSCM strategy that underpins the BCM strategy and its needs.

The Business Continuity Strategy should principally focus on business processes and associated issues (e.g. business process continuity, staff continuity, buildings continuity). Once the Business Continuity Strategy has been produced, and the role that IT services has to provide within the strategy has been determined, an ITSCM strategy can be produced that supports and enables the Business Continuity Strategy. This ensures that cost-effective decisions can be made, considering all the 'resources' to deliver a business process. Failure to do this tends to encourage ITSCM options that are faster, more elaborate and expensive than are actually needed.

The activities to be considered during initiation depend on the extent to which continuity facilities have been applied within the organization. Some parts of the business may have established individual Business Continuity Plans based around manual work-arounds, and IT may have developed continuity plans for systems perceived to be critical. This is good input to the process. However, effective ITSCM depends on supporting critical business functions. The only way of implementing effective ITSCM is through the identification of critical business processes and the analysis and coordination of the required technology and supporting IT services.

This situation may be even more complicated in outsourcing situations where an ITSCM process within an external service provider or outsourcer organization has to meet the needs not only of the customer BCM process and strategy, but also of the outsourcer's own BCM process and strategy. These needs may be in conflict with one another, or may conflict with the BCM needs of one of the other outsourcing organization's customers.

However, in many organizations BCM is absent or has very little focus, and often ITSCM is required to fulfil many of the requirements and activities of BCM. The rest of this section has assumed that ITSCM has had to perform many of the activities required by BCM. Where a BCM process is established with Business Continuity Strategies and Plans in place, these documents should provide the focus and drive for establishing ITSCM.

4.5.5 Process activities, methods and techniques

The following sections contain details of each of the stages within the ITSCM lifecycle.

4.5.5.1 Stage 1 – Initiation

The initiation process covers the whole of the organization and consists of the following activities:

- **Policy setting** – this should be established and communicated as soon as possible so that all members of the organization involved in, or affected by, Business Continuity issues are aware of their responsibilities to comply with and support ITSCM. As a minimum, the policy should set out management intention and objectives.
- **Specify terms of reference and scope** – this includes defining the scope and responsibilities of all staff in the organization. It covers such tasks as undertaking a Risk Analysis and Business Impact Analysis and determination of the command and control structure required to support a business interruption. There is also a need to take into account such issues as outstanding audit points, regulatory or client requirements and insurance organization stipulations, and compliance with standards such as ISO 27001, the Standard on Information Security Management, which also addresses Service Continuity requirements.
- **Allocate resources** – the establishment of an effective Business Continuity environment requires considerable resource in terms of both money and manpower. Depending on the maturity of the organization, with respect to ITSCM, there may be a requirement to familiarize and/or train staff to accomplish the Stage 2 tasks. Alternatively, the use of experienced external consultants may assist in completing the analysis more quickly. However, it is important that the organization can then maintain the process going forward without the need to rely totally on external support.
- **Define the project organization and control structure** – ITSCM and BCM projects are potentially complex and need to be well organized and controlled. It is strongly advisable to use a recognized standard project planning methodology such as Projects IN a Controlled Environment (PRINCE2®) or Project Management Body Of Knowledge (PMBOK®).
- **Agree project and quality plans** – plans enable the project to be controlled and variances addressed. Quality plans ensure that the deliverables are achieved and to an acceptable level of quality. They also provide a mechanism for communicating project resource requirements and deliverables, thereby obtaining 'buy-in' from all necessary parties.

4.5.5.2 Stage 2 – Requirements and strategy

Ascertaining the business requirements for IT service continuity is a critical component in order to determine how well an organization will survive a business interruption or disaster and the costs that will be incurred. If the requirements analysis is incorrect, or key information has been missed, this could have serious consequences on the effectiveness of ITSCM mechanisms.

This stage can effectively be split into two sections:

- **Requirements** – perform Business Impact Analysis and risk assessment
- **Strategy** – following the requirements analysis, the strategy should document the required risk reduction measures and recovery options to support the business.

Requirements – Business Impact Analysis

The purpose of a Business Impact Analysis (BIA) is to quantify the impact to the business that loss of service would have. This impact could be a 'hard' impact that can be precisely identified – such as financial loss – or 'soft' impact – such as public relations, morale, health and safety or loss of competitive advantage. The BIA will identify the most important services to the organization and will therefore be a key input to the strategy.

The BIA identifies:

- The form that the damage or loss may take – for example:
  - Lost income
  - Additional costs
  - Damaged reputation
  - Loss of goodwill
  - Loss of competitive advantage
  - Breach of law, health and safety

  - Risk to personal safety
  - Immediate and long-term loss of market share
  - Political, corporate or personal embarrassment
  - Loss of operational capability – for example, in a command and control environment
- How the degree of damage or loss is likely to escalate after a service disruption, and the times of the day, week, month or year when disruption will be most severe
- The staffing, skills, facilities and services (including the IT services) necessary to enable critical and essential business processes to continue operating at a minimum acceptable level
- The time within which minimum levels of staffing, facilities and services should be recovered
- The time within which all required business processes and supporting staff, facilities and services should be fully recovered
- The relative business recovery priority for each of the IT services.

One of the key outputs from a BIA exercise is a graph of the anticipated business impact caused by the loss of a business process or the loss of an IT service over time, as illustrated in Figure 4.22.

This graph can then be used to drive the business and IT continuity strategies and plans. More preventative measures need to be adopted with regard to those processes and services with earlier and higher impacts, whereas greater emphasis should be placed on continuity and recovery measures for those where the impact is lower and takes longer to develop. A balanced approach of both measures should be adopted to those in between.

These items provide the drivers for the level of ITSCM mechanisms that need to be considered or deployed. Once presented with these options, the business may decide that lower levels of service or increased delays are more acceptable, based on a cost-benefit analysis, or it maybe that comprehensive disaster prevention measures will need to be implemented.

These assessments enable the mapping of critical service, application and technology components to critical business processes, thus helping to identify the ITSCM elements that need to be provided. The business requirements are ranked and the associated ITSCM elements confirmed and prioritized in terms of risk reduction and recovery planning. The results of the BIA, discussed earlier, are invaluable input to several areas of process design including Service Level Management to understand the required service levels.

Impacts should be measured against particular scenarios for each business process, such as an inability to settle trades in a money market dealing process, or an inability to invoice for a period of days. An example is a money market dealing environment where loss of market data information could mean that the organization starts to lose money immediately as trading cannot continue. In

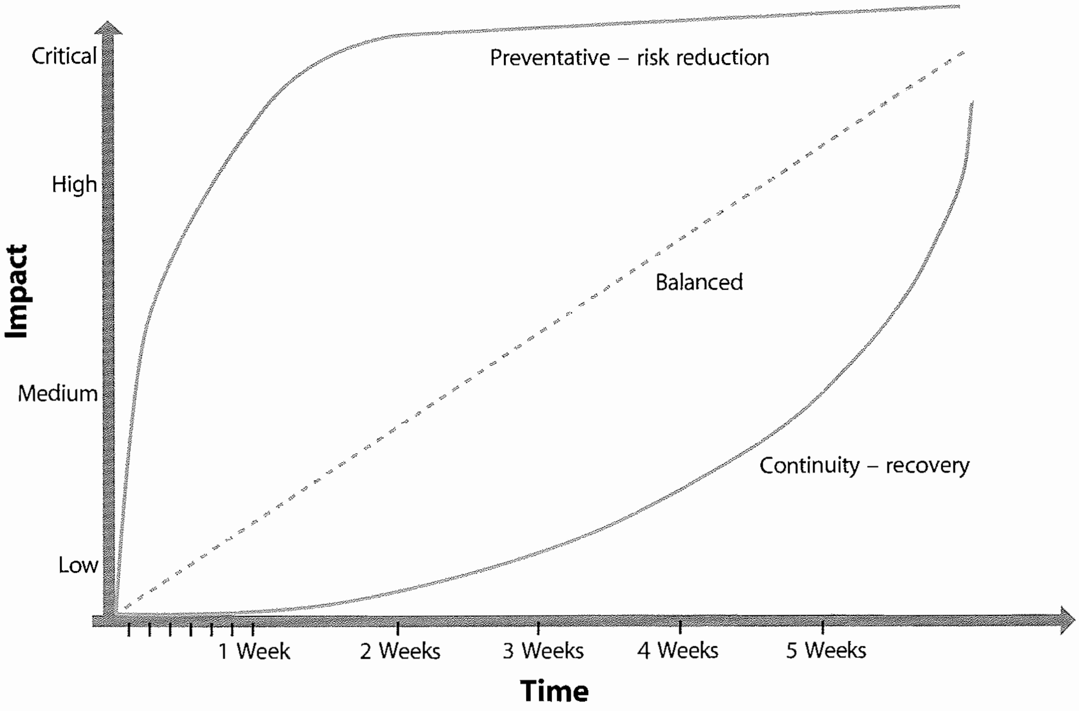


Figure 4.22 Graphical representation of business impacts

addition, customers may go to another organization, which would mean potential loss of core business. Loss of the settlement system does not prevent trading from taking place, but if trades already conducted cannot be settled within a specified period of time, the organization may be in breach of regulatory rules or settlement periods and suffer fines and damaged reputation. This may actually be a more significant impact than the inability to trade because of an inability to satisfy customer expectations.

It is also important to understand how impacts may change over time. For instance, it may be possible for a business to function without a particular process for a short period of time. In a balanced scenario, impacts to the business will occur and become greater over time. However, not all organizations are affected in this way. In some organizations, impacts are not apparent immediately. At some point, however, for any organization, the impacts will accrue to such a level that the business can no longer operate. ITSCM ensures that contingency options are identified so that the appropriate measure can be applied at the appropriate time to keep business impacts from service disruption to a minimum level.

When conducting a BIA, it is important that senior business area representatives' views are sought on the impact following loss of service. It is also equally important that the views of supervisory staff and more junior staff are sought to ensure all aspects of the impact following loss of service are ascertained. Often different levels of staff will have different views on the impact, and all will have to be taken into account when producing the overall strategy.

In many organizations it will be impossible, or it will not be cost-justifiable, to recover the total service in a very short timescale. In many cases, business processes can be re-established without a full complement of staff, systems and other facilities, and still maintain an acceptable level of service to clients and customers. The business recovery objectives should therefore be stated in terms of:

- The time within which a pre-defined team of core staff and stated minimum facilities must be recovered
- The timetable for recovery of remaining staff and facilities.

It may not always be possible to provide the recovery requirements to a detailed level. There is a need to balance the potential impact against the cost of recovery to ensure that the costs are acceptable. The recovery objectives do, however, provide a starting point from which different business recovery and ITSCM options can be evaluated.

Requirements – Risk Analysis

The second driver in determining ITSCM requirements is the likelihood that a disaster or other serious service disruption will actually occur. This is an assessment of the level of threat and the extent to which an organization is vulnerable to that threat. Risk Analysis can also be used in assessing and reducing the chance of normal operational incidents and is a technique used by Availability Management to ensure the required availability and reliability levels can be maintained. Risk Analysis is also a key aspect of Information Security Management. A diagram on Risk Analysis and Management (Figure 4.20) is contained within the Availability Management process in section 4.4.

A number of Risk Analysis and Management methods are available for both the commercial and government sectors. Risk Analysis is the assessment of the risks that may give rise to service disruption or security violation. Risk management is concerned with identifying appropriate risk responses or cost-justifiable counter-measures to combat those risks.

A standard methodology, such as the Management of Risk (M_o_R), should be used to assess and manage risks within an organization. The M_o_R framework is illustrated in Figure 4.23.

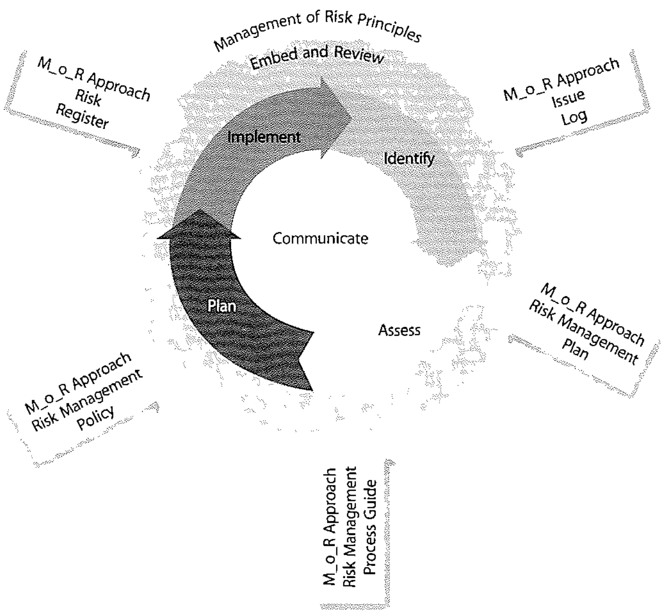

Management of Risk Principles
Embed and Review
M_o_R Approach Risk Register
M_o_R Approach Issue Log
Implement
Identify
Communicate
Plan
Assess
M_o_R Approach Risk Management Plan
M_o_R Approach Risk Management Policy
M_o_R Approach Risk Management Process Guide

Figure 4.23 Management of Risk

The M_o_R approach is based around the above framework, which consists of the following:

- **M_o_R principles:** these principles are essential for the development of good risk management practice and are derived from corporate governance principles.
- **M_o_R approach:** an organization's approach to these principles needs to be agreed and defined within the following living documents:
  - Risk Management Policy
  - Process Guide
  - Plans
  - risk registers
  - Issue Logs.
- **M_o_R Processes:** the following four main steps describe the inputs, outputs and activities that ensure that risks are controlled:
  - **Identify:** the threats and opportunities within an activity that could impact the ability to reach its objective
  - **Assess:** the understanding of the net effect of the identified threats and opportunities associated with an activity when aggregated together
  - **Plan:** to prepare a specific management response that will reduce the threats and maximize the opportunities
  - **Implement:** the planned risk management actions, monitor their effectiveness and take corrective action where responses do not match expectations.
- **Embedding and reviewing M_o_R:** having put the principles, approach and processes in place, they need to be continually reviewed and improved to ensure they remain effective.
- **Communication:** having the appropriate communication activities in place to ensure that everyone is kept up-to-date with changes in threats, opportunities and any other aspects of risk management.

This M_o_R method requires the evaluation of risks and the development of a risk profile, such as the example in Figure 4.24.

Figure 4.24 shows an example risk profile, containing many risks that are outside the defined level of 'acceptable risk'. Following the Risk Analysis it is possible to determine appropriate risk responses or risk reduction measures (ITSCM mechanisms) to manage the risks, i.e. reduce the risk to an acceptable level or mitigate the risk. Wherever possible, appropriate risk responses should be implemented to reduce either the impact or the likelihood, or both, of these risks from manifesting themselves.

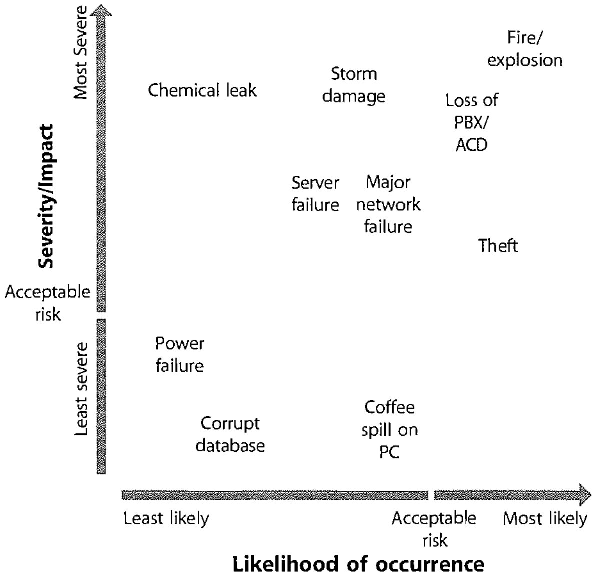


Figure 4.24 Example summary risk profile

In the context of ITSCM, there are a number of risks that need to be taken into consideration. The following is not a comprehensive list but does give some examples of risks and threats that need to be addressed by the ITSCM process.

IT Service Continuity Strategy

The results of the Business Impact Analysis and the Risk Analysis will enable appropriate Business and IT Service Continuity strategies to be produced in line with the business needs. The strategy will be an optimum balance of risk reduction and recovery or continuity options. This includes consideration of the relative service recovery priorities and the changes in relative service priority for the time of day, day of the week, and monthly and annual variations. Those services that have been identified as high impacts in the short term within the BIA will want to concentrate efforts on preventative risk reduction methods – for example, through full resilience and fault tolerance – while an organization that has low short-term impacts would be better suited to comprehensive recovery options, as described in the following sections. Similar advice and guidance can be found in the Business Continuity Institute's BCI Good Practice Guidelines.

Risk response measures

Most organizations will have to adopt a balanced approach where risk reduction and recovery are complementary and both are required. This entails reducing, as far as possible, the risks to the continued

Table 4.1 Examples of risks and threats

Risk	Threat
Loss of internal IT systems/networks, PABXs, ACDs, etc.	Fire Power failure Arson and vandalism Flood Aircraft impact Weather damage, e.g. hurricane Environmental disaster Terrorist attack Sabotage Catastrophic failure Electrical damage, e.g. lightning Accidental damage Poor-quality software
Loss of external IT systems/networks, e.g. e-commerce servers, cryptographic systems	All of the above Excessive demand for services Denial of service attack, e.g. against an internet firewall Technology failure, e.g. cryptographic system
Loss of data	Technology failure Human error Viruses, malicious software, e.g. attack applets
Loss of network services	Damage or denial of access to network service provider's premises Loss of service provider's IT systems/networks Loss of service provider's data Failure of the service provider
Unavailability of key technical and support staff	Industrial action Denial of access to premises Resignation Sickness/injury Transport difficulties
Failure of service providers, e.g. outsourced IT	Commercial failure, e.g. insolvency Denial of access to premises Unavailability of service provider's staff Failure to meet contractual service levels

provision of the IT service and is usually achieved through Availability Management. However well planned, it is impossible to completely eliminate all risks – for example, a fire in a nearby building will probably result in damage, or at least denial of access, as a result of the implementation of a cordon. As a general rule, the invocation of a recovery capability should only be taken as a last resort. Ideally, an organization should assess all of the risks to reduce the potential requirement to recover the business, which is likely to include the IT services.

The risk reduction measures need to be implemented and should be instigated in conjunction with Availability Management, as many of these reduce the probability of failure affecting the availability of service. Typical risk reduction measures include:

- Installation of UPS and backup power to the computer
- Fault-tolerant systems for critical applications where even minimal downtime is unacceptable – for example, a banking system
- RAID arrays and disk mirroring for LAN servers to prevent against data loss and to ensure continued availability of data
- Spare equipment/components to be used in the event of equipment or component failure – for example, a spare LAN server already configured with the standard configuration and available to replace a faulty server with minimum build and configuration time

- The elimination of SpoFs, such as single access network points or single power supply into a building
- Resilient IT systems and networks
- Outsourcing services to more than one provider
- Greater physical and IT-based security controls
- Better controls to detect service disruptions, such as fire detection systems, coupled with suppression systems
- A comprehensive backup and recovery strategy, including off-site storage.

The above measures will not necessarily solve an ITSCM issue and remove the risk totally, but all or a combination of them may significantly reduce the risks associated with the way in which services are provided to the business.

Off-site storage

One risk response method is to ensure all vital data is backed up and stored off-site. Once the recovery strategy has been defined, an appropriate backup strategy should be adopted and implemented to support it. The backup strategy must include regular (probably daily) removal of data (including the CMS to ease recovery) from the main data centres to a suitable off-site storage location. This will ensure retrieval of data following relatively minor operational failure as well as total and complete disasters. As well as the electronic data, all other important information and documents should be stored off-site, with the main example being the ITSCM plans.

ITSCM recovery options

An organization's ITSCM strategy is a balance between the cost of risk reduction measures and recovery options to support the recovery of critical business processes within agreed timescales. The following is a list of the potential IT recovery options that need to be considered when developing the strategy.

Manual work-arounds

For certain types of services, manual work-arounds can be an effective interim measure for a limited timeframe until the IT service is resumed. For instance, a Service Desk call-logging service could survive for a limited time using paper forms linked to a laptop computer with a spreadsheet.

Reciprocal arrangements

In the past, reciprocal arrangements were typical contingency measures where agreements were put in place with another organization using similar technology. This is no longer effective or possible for most types of IT systems, but can still be used in specific cases – for example, setting up an agreement to share high-speed printing facilities. Reciprocal arrangements can also be used for the off-site storage of backups and other critical information.

Gradual recovery

This option (sometimes referred to as 'cold standby') includes the provision of empty accommodation, fully equipped with power, environmental controls and local network cabling infrastructure, telecommunications connections, and available in a disaster situation for an organization to install its own computer equipment. It does not include the actual computing equipment, so is not applicable for services requiring speedy recovery, as set-up time is required before recovery of services can begin. This recovery option is only recommended for services that can bear a delay of recovery time in days or weeks, not hours. Any non-critical service that can bear this type of delay should take into account the cost of this option versus the benefit to the business before determining if a gradual recovery option should be included in the ITSCM options for the organization.

The accommodation may be provided commercially by a third party, for a fee, or may be private, (established by the organization itself) and provided as either a fixed or portable service.

A portable facility is typically a prefabricated building provided by a third party and located, when needed, at a predetermined site agreed with the organization. This may be in another location some distance from the home site, perhaps another owned building. The replacement computer equipment will need to be planned, but suppliers of computing equipment do not always guarantee replacement equipment within a fixed deadline, though they would normally do so under their best efforts.

Intermediate recovery

This option (sometimes referred to as 'warm standby') is selected by organizations that need to recover IT facilities within a predetermined time to prevent impacts to the business process. The predetermined time will have been agreed with the business during the BIA.

Most common is the use of commercial facilities, which are offered by third-party recovery organizations to a number of subscribers, spreading the cost across those subscribers. Commercial facilities often include operation, system management and technical support. The cost varies depending on the facilities requested, such as processors, peripherals, communications, and how quickly the services must be restored.

The advantage of this service is that the customer can have virtually instantaneous access to a site, housed in a secure building, in the event of a disaster. It must be understood, however, that the restoration of services at the site may take some time, as delays may be encountered while the site is re-configured for the organization that invokes the service, and the organization's applications and data will need to be restored from backups.

One potentially major disadvantage is the security implications of running IT services at a third party's data centre. This must be taken into account when planning to use this type of facility. For some organizations, the external intermediate recovery option may not be appropriate for this reason.

If the site is invoked, there is often a daily fee for use of the service in an emergency, although this may be offset against additional cost of working insurance.

Commercial recovery services can be provided in self-contained, portable or mobile form where an agreed system is delivered to a customer's site, within an agreed time.

Fast recovery

This option (sometimes referred to as 'hot standby') provides for fast recovery and restoration of services and is sometimes provided as an extension to the intermediate recovery provided by a third-party recovery provider. Some organizations will provide their own facilities within the organization, but not on an alternative site to the one used for the normal operations. Others implement their own internal second locations on an alternative site to provide more resilient recovery.

Where there is a need for a fast restoration of a service, it is possible to 'rent' floor space at the recovery site and install servers or systems with application systems and communications already available, and data mirrored from the operational servers. In the event of a system failure, the customers can then recover and switch over to the backup facility with little loss of service. This typically involves the re-establishment of the critical systems and services within a 24-hour period.

Immediate recovery

This option (also often referred to as 'hot standby', 'mirroring', 'load balancing' or 'split site') provides for immediate restoration of services, with no loss of service. For business critical services, organizations requiring continuous operation will provide their own facilities within the organization, but not on the same site as the normal operations. Sufficient IT equipment will be 'dual located 'in either an owned or hosted location to run the compete service from either location in the event of loss of one facility, with no loss of service to the customer. The second site can then be recovered whilst the service is provided from the single operable location. This is an expensive option, but may be justified for critical business processes or VBFs where non-availability for a short period could result in a significant impact, or where it would not be appropriate to be running IT services on a third party's premises for security or other reasons. The facility needs to be located separately and far enough away from the home site that it will not be affected by a disaster affecting that location. However, these mirrored servers and sites options should be implemented in close liaison with Availability Management as they support services with high levels of availability.

The strategy is likely to include a combination of risk response measures and a combination of the above recovery options, as illustrated in Figure 4.25.

	Manual	Immediate	Fast	Intermediate	Gradual
Service Desk	Yes		Yes	Yes	Yes
Mainframe payroll	Yes			Yes	Yes
Financial system			Yes		Yes
Dealer system		Yes		Yes	Yes

Figure 4.25 Example set of recovery options

Figure 4.25 shows that a number of options may be used to provide continuity of service. An example from Figure 4.25 shows that, initially, continuity of the Service Desk is provided using manual processes such as a set of forms, and maybe a spreadsheet operating from a laptop computer, whilst recovery plans for the service are completed on an alternative 'fast recovery' site. Once the alternative site has become operational, the Service Desk can switch back to using the IT service. However, use of the external 'fast recovery' alternative site is probably limited in duration, so while running temporarily from this site, the 'intermediate site' can be made operational and long-term operations can be transferred there.

Different services within an organization require different in-built resilience and different recovery options. Whatever option is chosen, the solution will need to be cost-justified. As a general rule, the longer the business can survive without a service, the cheaper the solution will be. For example, a critical healthcare system that requires continuous operation will be very costly, as potential loss of service will need to be eliminated by the use of immediate recovery, whereas a service the absence of which does not severely affect the business for a week or so could be supported by a much cheaper solution, such as intermediate recovery.

As well as the recovery of the computing equipment, planning needs to include the recovery of accommodation and infrastructure for both IT and user staff. Other areas to be taken into account include critical services such as power, telecommunications, water, couriers, post, paper records and reference material.

It is important to remember that the recovery is based around a series of stand-by arrangements including accommodation, procedures and people, as well as systems and telecommunications. Certain actions are necessary to implement the stand-by arrangements. For example:

- Negotiating for third-party recovery facilities and entering into a contractual arrangement
- Preparing and equipping the stand-by accommodation
- Purchasing and installing stand-by computer systems.

4.5.5.3 Stage 3 – Implementation

Once the strategy has been approved, the IT Service Continuity Plans need to be produced in line with the Business Continuity Plans.

ITSCM plans need to be developed to enable the necessary information for critical systems, services and facilities to either continue to be provided or to be reinstated within an acceptable period to the business. An example ITSCM recovery plan is contained in Appendix K. Generally the Business Continuity Plans rely on the availability of IT services, facilities and resources. As a consequence of this, ITSCM plans need to address all activities to ensure that the required services, facilities and resources are delivered in an acceptable operational state and are 'fit for purpose' when accepted by the business. This entails not only the restoration of services and facilities, but also the understanding of dependencies between them, the testing required prior to delivery (performance, functional, operational and acceptance testing) and the validation of data integrity and consistency.

It should be noted that the continuity plans are more than just recovery plans, and should include documentation of the resilience measures and the measures that have been put into place to enable recovery, together with explanations of why a particular approach has been taken (this facilitates decisions should invocation determine that the particular situation requires a modification to the plan). However, the format of the plan should enable rapid access to the recovery information itself, perhaps as an appendix that can be accessed directly. All key staff should have access to copies of all the necessary recovery documentation.

Management of the distribution of the plans is important to ensure that copies are available to key staff at all times. The plans should be controlled documents (with formalized documents maintained under Change Management and Configuration Management control) to ensure that only the latest versions are in circulation and each recipient should ensure that a personal copy is maintained off-site.

The plan should ensure that all details regarding recovery of the IT services following a disaster are fully documented. It should have sufficient details to enable a technical person unfamiliar with the systems to be able to follow the procedures. The recovery plans include key details such as the data recovery point, a list of dependent systems, the nature of the dependency and their data recovery points, system hardware and software requirements, configuration details and references to other relevant or essential information about the service and systems.

It is a good idea to include a checklist that covers specific actions required during all stages of recovery for the service and system. For example, after the system has been restored to an operational state, connectivity checks, functionality checks or data consistency and integrity checks should be carried out prior to handing the service over to the business.

There are a number of technical plans that may already exist within an organization, documenting recovery procedures from a normal operational failure. The development and maintenance of these plans will be the responsibility of the specialist teams, but will be coordinated by the Business Continuity Management team. These will be useful additions or appendices to the main plan. Additionally, plans that will need to be integrated with the main BCP are:

- **Emergency Response Plan:** to interface to all emergency services and activities
- **Damage Assessment Plan:** containing details of damage assessment contacts, processes and plans
- **Salvage Plan:** containing information on salvage contacts, activities and processes
- **Vital Records Plan:** details of all vital records and information, together with their location, that are critical to the continued operation of the business
- **Crisis Management and Public Relations Plan:** the plans on the command and control of different crisis situations and management of the media and public relations
- **Accommodation and Services Plan:** detailing the management of accommodation, facilities and the services necessary for their continued operation
- **Security Plan:** showing how all aspects of security will be managed on all home sites and recovery sites
- **Personnel Plan:** containing details of how all personnel issues will be managed during a major incident
- **Communication Plan:** showing how all aspects of communication will be handled and managed with all relevant areas and parties involved during a major incident
- **Finance and Administration Plan:** containing details of alternative methods and processes for obtaining possible emergency authorization and access to essential funds during a major incident.

Finally, each critical business area is responsible for the development of a plan detailing the individuals who will be in the recovery teams and the tasks to be undertaken on invocation of recovery arrangements.

The ITSCM Plan must contain all the information needed to recover the IT systems, networks and telecommunications in a disaster situation once a decision to invoke has been made, and then to manage the business return to normal operation once the service disruption has been resolved. One of the most important inputs into the plan development is the results of the Business Impact Analysis. Additionally other areas will need to be analysed, such as Service Level Agreements (SLA), security requirements, operating instructions and procedures and external contracts. It is likely that a separate SLA with alternative targets will have been agreed if running at a recovery site following a disaster.

Other areas that will need to be implemented following the approval of the strategy are:

Organization planning

During the disaster recovery process, the organizational structure will inevitably be different from normal operation and is based around:

- Executive – including senior management/executive board, with overall authority and control within the organization and responsible for crisis management and liaison with other departments, divisions, organizations, the media, regulators, emergency services etc.
- Coordination – typically one level below the executive group and responsible for coordinating the overall recovery effort within the organization
- Recovery – a series of business and service recovery teams, representing the critical business functions and the services that need to be established to support these functions. Each team is responsible for executing the plans within their own areas and for liaison with staff, customers and third parties. Within IT the recovery teams should be grouped by IT service and application. For example, the infrastructure team may have one or more people responsible for recovering external connections, voice services, local area networks, etc. and the support teams may be split by platform, operating system or application. In addition, the recovery priorities for the service, application or its components identified during the Business Impact Analysis should be documented within the recovery plans and applied during their execution.

Testing

Experience has shown that recovery plans that have not been fully tested do not work as intended, if at all. Testing is therefore a critical part of the overall ITSCM process and the only way of ensuring that the selected strategy, standby arrangements, logistics, business recovery plans and procedures will actually work in practice.

The IT service provider is responsible for ensuring that the IT services can be recovered in the required timescales with the required functionality and the required performance following a disaster.

There are four basic types of tests that can be undertaken:

- **Walk-through tests** can be conducted when the plan has been produced simply by getting the relevant people together to see if the plan(s) at least work in a simulated way.
- **Full tests** should be conducted as soon as possible after the plan production and at regular intervals of at least annually thereafter. They should involve the business units to assist in proving the capability to recover the services appropriately. They should, as far as possible, replicate an actual invocation of all stand-by arrangements and should involve external parties if they are planned to be involved in an actual invocation. The tests must not only prove recovery of the IT services but also the recovery of the business processes. It is recommended that an independent observer records all the activities of the tests and the timings of the service recovery. The observer's documentation of the tests will be vital input into the subsequent post mortem review. The full tests may be announced or unannounced. The first test of the plan is likely to be announced and carefully planned, but subsequent tests may be 'sprung' on key players without warning. It is also essential that many different people get involved, including those not very familiar with the IT service and systems, as the people with the most knowledge may not be available when a disaster actually occurs.
- **Partial tests** can also be undertaken where recovery of certain elements of the overall plan is tested, such as single services or servers. These types of tests should be in addition to the full test not instead of the full test. The full test is the best way of testing that all services can be recovered in required timescales and can run together on the recovery systems.
- **Scenario tests** can be used to test reactions and plans to specific conditions, events and scenarios. They can include testing that BCPs and IT Service Continuity Plans interface with each other, as well as interfacing with all other plans involved in the handling and management of a major incident.

All tests need to be undertaken against defined test scenarios, which are described as realistically as possible. It should be noted, however, that even the most comprehensive test does not cover everything. For example, in a service disruption where there has been injury or even death to colleagues, the reaction of staff to a crisis cannot be tested and the plans need to make allowance for this. In addition, tests must have clearly defined objectives and Critical Success Factors, which will be used to determine the success or otherwise of the exercise.

4.5.5.4 Stage 4 – Ongoing operation

This stage consists of the following:

- **Education, awareness and training** – this should cover the organization and, in particular, the IT organization, for service continuity-specific items. This ensures that all staff are aware of the implications of business continuity and of service continuity and consider these as part of their normal working, and that everyone involved in the plan has been trained in how to implement their actions.
- **Review** – regular review of all of the deliverables from the ITSCM process needs to be undertaken to ensure that they remain current.
- **Testing** – following the initial testing, it is necessary to establish a programme of regular testing to ensure that the critical components of the strategy are tested, preferably at least annually, although testing of IT Service Continuity Plans should be arranged in line with business needs and the needs of the BCPs. All plans should also be tested after every major business change. It is important that any changes to the IT technology are also included in the strategy, implemented in an appropriate fashion and tested to ensure that they function correctly within the overall provision of IT following a disaster. The backup and recovery of IT service should also be monitored and tested to ensure that when they are needed during a major incident, they will operate as needed. This aspect is covered more fully in the Service Operation publication
- **Change Management** – the Change Management process should ensure that all changes are assessed for their potential impact on the ITSCM plans. If the planned change will invalidate the plans, then the plan must be updated before the change is implemented, and it should be tested as part of the change testing. The plans themselves must be under very strict Change Management and Configuration Management control. Inaccurate plans and inadequate recovery capabilities may result in the failure of BCPs. Also, on an ongoing basis, whenever there are new services or where services have major changes, it is essential that a BIA and risk assessment is conducted on the new or changed service and the strategy and plans updated accordingly.

Invocation

Invocation is the ultimate test of the Business Continuity and ITSCM Plans. If all the preparatory work has been successfully completed, and plans developed and tested, then an invocation of the Business Continuity Plans should be a straightforward process, but if the plans have not been tested, failures can be expected. It is important that due consideration is given to the design of all invocation processes, to ensure that they are fit for purpose and interface to all other relevant invocation processes.

Invocation is a key component of the plans, which must include the invocation process and guidance. It should be remembered that the decision to invoke, especially if a third-party recovery facility is to be used, should not be taken lightly. Costs will be involved and the process will involve disruption to the business. This decision is typically made by a 'crisis management' team, comprising senior managers from the business and support departments (including IT), using information gathered through damage assessment and other sources.

A disruption could occur at any time of the day or night, so it is essential that guidance on the invocation process is readily available. Plans must be available to key staff in the office and away from the office.

The decision to invoke must be made quickly, as there may be a lead-time involved in establishing facilities at a recovery site. In the case of a serious building fire, the decision may be fairly easy to make. However, in the case of power failure or hardware fault, where a resolution is expected within a short period, a deadline should be set by which time if the incident has not been resolved, invocation will take place. If using external services providers, they should be warned immediately if there is a chance that invocation might take place.

The decision to invoke needs to take into account the:

- Extent of the damage and scope of the potential invocation
- Likely length of the disruption and unavailability of premises and/or services
- Time of day/month/year and the potential business impact. At year-end, the need to invoke may be more pressing, to ensure that year-end processing is completed on time.

Therefore the design of the invocation process must provide guidance on how all of these areas and circumstances should be assessed to assist the person invoking the continuity plan.

The ITSCM Plan should include details of activities that need to be undertaken, including:

- Retrieval of backup tapes or use of data vaulting to retrieve data
- Retrieval of essential documentation, procedures, workstation images, etc. stored off-site
- Mobilization of the appropriate technical personnel to go to the recovery site to commence the recovery of required systems and services
- Contacting and putting on alert telecommunications suppliers, support services, application vendors, etc. who may be required to undertake actions or provide assistance in the recovery process.

The invocation and initial recovery is likely to be a time of high activity, involving long hours for many individuals. This must be recognized and managed by the recovery team leaders to ensure that breaks are provided and prevent 'burn-out'. Planning for shifts and handovers must be undertaken to ensure that the best use is made of the facilities available. It is also vitally important to ensure that the usual business and technology controls remain in place during invocation, recovery and return to normal to ensure that information security is maintained at the correct level and that data protection is preserved.

Once the recovery has been completed, the business should be able to operate from the recovery site at the level determined and agreed in the strategy and relevant SLA. The objective, however, will be to build up the business to normal levels, maintain operation from the recovery site in the short term and vacate the recovery site in the shortest possible time. Details of all these activities need to be contained within the plans. If using external services, there will be a finite contractual period for using the facility. Whatever the period, a return to normal must be carefully planned and undertaken in a controlled fashion. Typically this will be over a weekend and may include some necessary downtime in business hours. It is important that this is managed well and that all personnel involved are aware of their responsibilities to ensure a smooth transition.

4.5.6 Triggers, inputs, outputs and interfaces

Many events may trigger ITSCM activity. These include:

- New or changed business needs, or new or changed services
- New or changed targets within agreements, such as SLRs, SLAs, OLAs or contracts
- The occurrence of a major incident that requires assessment for potential invocation of either Business or IT Continuity Plans

- Periodic activities such as the BIA or Risk Analysis activities, maintenance of Continuity Plans or other reviewing, revising or reporting activities
- Assessment of changes and attendance at Change Advisory Board meetings
- Review and revision of business and IT plans and strategies
- Review and revision of designs and strategies
- Recognition or notification of a change of risk or impact of a business process or VBF, an IT service or component
- Initiation of tests of continuity and recovery plans.

Integration and interfaces exist from ITSCM to all other processes. Important examples are as follows:

- **Change Management** – all changes need to be considered for their impact on the continuity plans, and if amendments are required to the plan, updates to the plan need to be part of the change. The plan itself must be under Change Management control.
- **Incident and Problem Management** – incidents can easily evolve into major incidents or disasters. Clear criteria need to be agreed and documented on for the invocation of the ITSCM plans.
- **Availability Management** – undertaking Risk Analysis and implementing risk responses should be closely coordinated with the availability process to optimize risk mitigation.
- **Service Level Management** – recovery requirements will be agreed and documented in the SLAs. Different service levels could be agreed and documented that could be acceptable in a disaster situation.
- **Capacity Management** – ensuring that there are sufficient resources to enable recovery onto replacement computers following a disaster.
- **Configuration Management** – the CMS documents the components that make up the infrastructure and the relationship between the components. This information is invaluable for all the stages of the ITSCM lifecycle, the maintenance of plans and recovery facilities.
- **Information Security Management** – a very close relationship exists between ITSCM and Information Security Management. A major security breach could be considered a disaster, so when conducting BIA and Risk Analysis, security will be a very important consideration.

4.5.6.1 Inputs

There are many sources of input required by the ITSCM process:

- Business information: from the organization's business strategy, plans and financial plans, and information on their current and future requirements
- IT information: from the IT strategy and plans and current budgets
- A Business Continuity Strategy and a set of Business Continuity Plans: from all areas of the business
- Service information: from the SLM process, with details of the services from the Service Portfolio and the Service Catalogue and service level targets within SLAs and SLRs
- Financial information: from Financial Management, the cost of service provision, the cost of resources and components
- Change information: from the Change Management process, with a Change Schedule and a need to assess all changes for their impact on all ITSCM plans
- CMS: containing information on the relationships between the business, the services, the supporting services and the technology
- Business Continuity Management and Availability Management testing schedules
- IT Service Continuity Plans and test reports from supplier and partners, where appropriate.

4.5.6.2 Outputs

The outputs from the ITSCM process include:

- A revised ITSCM policy and strategy
- A set of ITSCM plans, including all Crisis Management, Emergency Response Plans and Disaster Recovery Plans, together with a set of supporting plans and contracts with recovery service providers
- Business Impact Analysis exercises and reports, in conjunction with BCM and the business
- Risk Analysis and Management reviews and reports, in conjunction with the business, Availability Management and Security Management
- An ITSCM testing schedule
- ITSCM test scenarios
- ITSCM test reports and reviews.

Forecasts and predictive reports are used by all areas to analyse, predict and forecast particular business and IT scenarios and their potential solutions.

4.5.7 Key Performance Indicators

IT services are delivered and can be recovered to meet business objectives:

- Regular audits of the ITSCM Plans to ensure that, at all times, the agreed recovery requirements of the business can be achieved

- All service recovery targets are agreed and documented in SLAs and are achievable within the ITSCM Plans
- Regular and comprehensive testing of ITSCM Plans
- Regular reviews are undertaken, at least annually, of the business and IT continuity plans with the business areas
- Negotiate and manage all necessary ITSCM contracts with third party
- Overall reduction in the risk and impact of possible failure of IT services.

Awareness throughout the organizations of the plans:

- Ensure awareness of business impact, needs and requirements throughout IT
- Ensure that all IT service areas and staff are prepared and able to respond to an invocation of the ITSCM Plans
- Regular communication of the ITSCM objectives and responsibilities within the appropriate business and IT service areas.

4.5.8 Information Management

ITSCM needs to record all of the information necessary to maintain a comprehensive set of ITSCM plans. This information base should include:

- Information from the latest version of the BIA
- Comprehensive information on risk within a Risk Register, including risk assessment and risk responses
- The latest version of the BCM strategy and BCPs
- Details relating to all completed tests and a schedule of all planned tests
- Details of all ITSCM Plans and their contents
- Details of all other plans associated with ITSCM Plans
- Details of all existing recovery facilities, recovery suppliers and partners, recovery agreements and contracts, spare and alternative equipment
- Details of all backup and recovery processes, schedules, systems and media and their respective locations.

All the above information needs to be integrated and aligned with all BCM information and all the other information required by ITSCM. Interfaces to many other processes are required to ensure that this alignment is maintained.

4.5.9 Challenges, Critical Success Factors and risks

One of the major challenges facing ITSCM is to provide appropriate plans when there is no BCM process. If there is no BCM process, then IT is likely to make incorrect assumptions about business criticality of business processes and therefore adopt the wrong continuity strategies and options. Without BCM, expensive ITSCM solutions and plans will be rendered useless by the absence of corresponding plans and arrangements within the business. Also, if BCM is absent, then the business may fail to identify inexpensive non-IT solutions and waste money on ineffective, expensive IT solutions.

In some organizations, the business perception is that continuity is an IT responsibility, and therefore the business assumes that IT will be responsible for disaster recovery and that IT services will continue to run under any circumstances. This is especially true in some outsourced situations where the business may be reluctant to share its BCM information with an external service provider.

If there is a BCM process established, then the challenge becomes one of alignment and integration. ITSCM must ensure that accurate information is obtained from the BCM process on the needs, impact and priorities of the business, and that the ITSCM information and plans are aligned and integrated with those of the business. Having achieved that alignment, the challenge becomes one of keeping them aligned by management and control of business and IT change. It is essential, therefore, that all documents and plans are maintained under strict Change Management and Configuration Management control.

The main CSFs for the ITSCM process are:

- IT services are delivered and can be recovered to meet business objectives
- Awareness throughout the organization of the business and IT Service Continuity Plans.

Some of the major risks associated with ITSCM include:

- Lack of commitment from the business to the ITSCM processes and procedures
- Lack of commitment from the business and a lack of appropriate information on future plans and strategies
- Lack of senior management commitment or a lack of resources and/or budget for the ITSCM process
- The processes focus too much on the technology issues and not enough on the IT services and the needs and priorities of the business

- Risk Analysis and Management are conducted in isolation and not in conjunction with Availability Management and Security Management
- ITSCM plans and information become out-of-date and lose alignment with the information and plans of the business and BCM.

4.6 INFORMATION SECURITY MANAGEMENT

4.6.1 Purpose/goal/objective

> 'The goal of the ISM process is to align IT security with business security and ensure that information security is effectively managed in all service and Service Management activities'.

ISM needs to be considered within the overall corporate governance framework. Corporate governance is the set of responsibilities and practices exercised by the board and executive management with the goal of providing strategic direction, ensuring the objectives are achieved, ascertaining the risks are being managed appropriately and verifying that the enterprise's resources are used effectively.

Information security is a management activity within the corporate governance framework, which provides the strategic direction for security activities and ensures objectives are achieved. It further ensures that the information security risks are appropriately managed and that enterprise information resources are used responsibly. The purpose of ISM is to provide a focus for all aspects of IT security and manage all IT security activities.

The term 'information' is used as a general term and includes data stores, databases and metadata. The objective of information security is to protect the interests of those relying on information, and the systems and communications that deliver the information, from harm resulting from failures of availability, confidentiality and integrity.

For most organizations, the security objective is met when:

- Information is available and usable when required, and the systems that provide it can appropriately resist attacks and recover from or prevent failures (availability)
- Information is observed by or disclosed to only those who have a right to know (confidentiality)
- Information is complete, accurate and protected against unauthorized modification (integrity)
- Business transactions, as well as information exchanges between enterprises, or with partners, can be trusted (authenticity and non-repudiation).

Prioritization of confidentiality, integrity and availability must be considered in the context of business and business processes. The primary guide to defining what must be protected and the level of protection has to come from the business. To be effective, security must address entire business processes from end to end and cover the physical and technical aspects. Only within the context of business needs and risks can management define security.

4.6.2 Scope

The ISM process should be the focal point for all IT security issues, and must ensure that an Information Security Policy is produced, maintained and enforced that covers the use and misuse of all IT systems and services. ISM needs to understand the total IT and business security environment, including the:

- Business Security Policy and plans
- Current business operation and its security requirements
- Future business plans and requirements
- Legislative requirements
- Obligations and responsibilities with regard to security contained within SLAs
- The business and IT risks and their management.

Understanding all of this will enable ISM to ensure that all the current and future security aspects and risks of the business are cost-effectively managed.

The ISM process should include:

- The production, maintenance, distribution and enforcement of an Information Security Policy and supporting security policies
- Understanding the agreed current and future security requirements of the business and the existing Business Security Policy and plans
- Implementation of a set of security controls that support the Information Security Policy and manage risks associated with access to services, information and systems
- Documentation of all security controls, together with the operation and maintenance of the controls and their associated risks
- Management of suppliers and contracts regarding access to systems and services, in conjunction with Supplier Management
- Management of all security breaches and incidents associated with all systems and services
- The proactive improvement of security controls, and security risk management and the reduction of security risks

- Integration of security aspects within all other IT SM processes.

To achieve effective information security governance, management must establish and maintain an Information Security Management System (ISMS) to guide the development and management of a comprehensive information security programme that supports the business objectives.

4.6.3 Value to the business

ISM ensures that an Information Security Policy is maintained and enforced that fulfils the needs of the Business Security Policy and the requirements of corporate governance. ISM raises awareness of the need for security within all IT services and assets throughout the organization, ensuring that the policy is appropriate for the needs of the organization. ISM manages all aspects of IT and information security within all areas of IT and Service Management activity.

ISM provides assurance of business processes by enforcing appropriate security controls in all areas of IT and by managing IT risk in line with business and corporate risk management processes and guidelines.

4.6.4 Policies/principles/basic concepts

Prudent business practices require that IT processes and initiatives align with business processes and objectives. This is critical when it comes to information security, which must be closely aligned with business security and business needs. Additionally all processes within the IT organization must include security considerations.

Executive management is ultimately responsible for the organization's information and is tasked with responding to issues that affect its protection. In addition, boards of directors are expected to make information security an integral part of corporate governance. All IT service provider organizations must therefore ensure that they have a comprehensive ISM policy(s) and the necessary security controls in place to monitor and enforce the policies.

4.6.4.1 Security framework

The Information Security Management process and framework will generally consist of:

- An Information Security Policy and specific security policies that address each aspect of strategy, controls and regulation
- An Information Security Management System (ISMS), containing the standards, management procedures and guidelines supporting the information security policies
- A comprehensive security strategy, closely linked to the business objectives, strategies and plans
- An effective security organizational structure
- A set of security controls to support the policy
- The management of security risks
- Monitoring processes to ensure compliance and provide feedback on effectiveness
- Communications strategy and plan for security
- Training and awareness strategy and plan.

4.6.4.2 The Information Security Policy

Information Security Management activities should be focused on and driven by an overall Information Security Policy and a set of underpinning specific security policies. The ITP should have the full support of top executive IT management and ideally the support and commitment of top executive business management. The policy should cover all areas of security, be appropriate, meet the needs of the business and should include:

- An overall Information Security Policy
- Use and misuse of IT assets policy
- An access control policy
- A password control policy
- An e-mail policy
- An internet policy
- An anti-virus policy
- An information classification policy
- A document classification policy
- A remote access policy
- A policy with regard to supplier access of IT service, information and components
- An asset disposal policy.

These policies should be widely available to all customers and users, and their compliance should be referred to in all SLRs, SLAs, contracts and agreements. The policies should be authorized by top executive management within the business and IT, and compliance to them should be endorsed on a regular basis. All security policies should be reviewed – and, where necessary, revised – on at least an annual basis.

4.6.4.3 The Information Security Management System (ISMS)

The framework or the ISMS in turn provides a basis for the development of a cost-effective information security programme that supports the business objectives. It will involve the Four Ps of People, Process, Products and

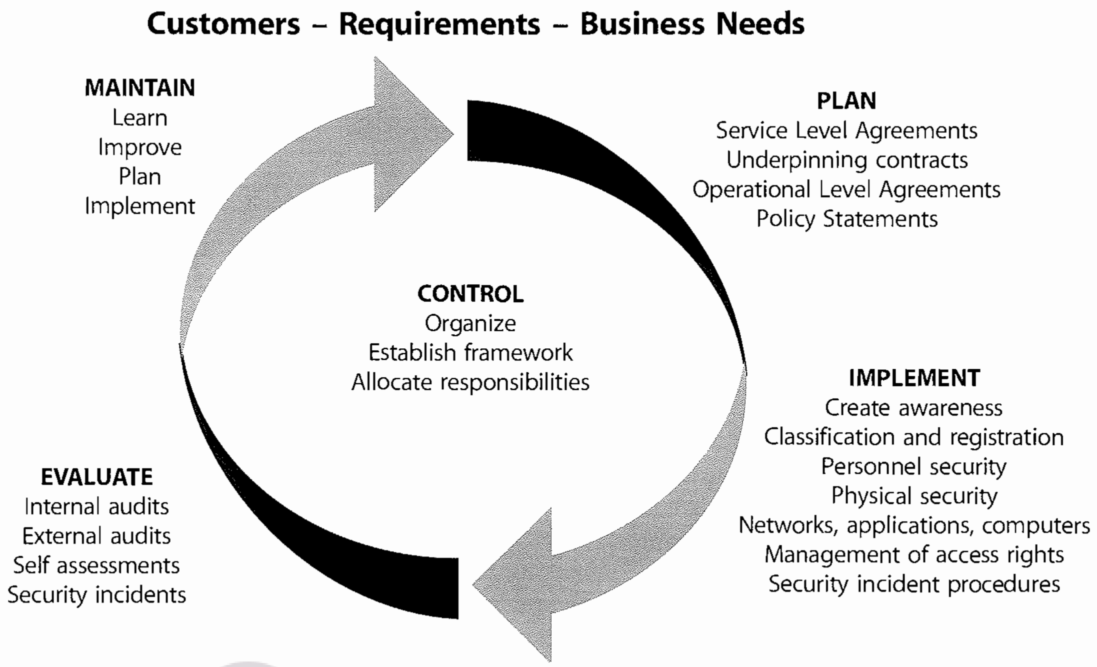


Figure 4.26 Framework for managing IT security

technology as well as Partners and suppliers to ensure high levels of security are in place.

ISO 27001 is the formal standard against which organizations may seek independent certification of their ISMS (meaning their frameworks to design, implement, manage, maintain and enforce information security processes and controls systematically and consistently throughout the organizations). The ISMS shown in Figure 4.26 shows an approach that is widely used and is based on the advice and guidance described in many sources, including ISO 27001.

The five elements within this framework are as follows:

Control

The objectives of the control element of the ISMS are to:

- Establish a management framework to initiate and manage information security in the organization
- Establish an organization structure to prepare, approve and implement the Information Security Policy
- Allocate responsibilities
- Establish and control documentation.

Plan

The objective of the plan element of the ISMS is to devise and recommend the appropriate security measures, based on an understanding of the requirements of the organization.

The requirements will be gathered from such sources as business and service risk, plans and strategies, SLAs and OLAs and the legal, moral and ethical responsibilities for information security. Other factors, such as the amount of funding available and the prevailing organization culture and attitudes to security, must be considered.

The Information Security Policy defines the organization's attitude and stance on security matters. This should be an organization-wide document, not just applicable to the IT service provider. Responsibility for the upkeep of the document rests with the Information Security Manager.

Implement

The objective of the implementation of the ISMS is to ensure that appropriate procedures, tools and controls are in place to underpin the Information Security Policy.

Amongst the measures are:

- **Accountability for assets** – Configuration Management and the CMS are invaluable here
- **Information classification** – information and repositories should be classified according to the sensitivity and the impact of disclosure.

The successful implementation of the security controls and measures is dependent on a number of factors:

- The determination of a clear and agreed policy, integrated with the needs of the business
- Security procedures that are justified, appropriate and supported by senior management
- Effective marketing and education in security requirements
- A mechanism for improvement.

Evaluation

The objectives of the evaluation element of the ISMS are to:

- Supervise and check compliance with the security policy and security requirements in SLAs and OLAs
- Carry out regular audits of the technical security of IT systems
- Provide information to external auditors and regulators, if required.

Maintain

The objectives of this maintain element of the ISMS are to:

- Improve security agreements as specified in, for example, SLAs and OLAs
- Improve the implementation of security measures and controls.

This should be achieved using a PDCA (Plan–Do–Check–Act) cycle, which is a formal approach suggested by ISO 27001 for the establishment of the Information Security Management System (ISMS) or framework. This cycle is described in more detail in the Continual Service Improvement publication.

Security governance

Information security governance, when properly implemented, should provide six basic outcomes:

- Strategic alignment:
  - Security requirements should be driven by enterprise requirements
  - Security solutions need to fit enterprise processes
  - Investment in information security should be aligned with the enterprise strategy and agreed-on risk profile.
- Value delivery:
  - A standard set of security practices, i.e. baseline security requirements following best practices
  - Properly prioritized and distributed effort to areas with greatest impact and business benefit
  - Institutionalized and commoditized solutions
  - Complete solutions, covering organization and process as well as technology
  - A culture of continual improvement.
- Risk management:
  - Agreed-on risk profile
  - Understanding of risk exposure
  - Awareness of risk management priorities
  - Risk mitigation
  - Risk acceptance/deference.
- Performance Management:
  - Defined, agreed and meaningful set of metrics
  - Measurement process that will help identify shortcomings and provide feedback on progress made resolving issues
  - Independent assurance.
- Resource management:
  - Knowledge is captured and available
  - Documented security processes and practices
  - Developed security architecture(s) to efficiently utilize infrastructure resources.
- Business process assurance.

4.6.5 Process activities, methods and techniques

The purpose of the ISM process is to ensure that the security aspects with regard to services and all Service Management activities are appropriately managed and controlled in line with business needs and risks:

The key activities within the ISM process are:

- Production, review and revision of an overall Information Security Policy and a set of supporting specific policies
- Communication, implementation and enforcement of the security policies
- Assessment and classification of all information assets and documentation
- Implementation, review, revision and improvement of a set of security controls and risk assessment and responses
- Monitoring and management of all security breaches and major security incidents
- Analysis, reporting and reduction of the volumes and impact of security breaches and incidents
- Schedule and completion of security reviews, audits and penetration tests.

The interactions between these key activities are illustrated in Figure 4.27.

The developed Information Security Management processes, together with the methods, tools and techniques, constitute the security strategy. The security manager should ensure that technologies, products and services are in place and that the overall policy is developed and well published. The security manager is also responsible for security architecture, authentication, authorization, administration and recovery.

The security strategy also needs to consider how it will embed good security practices into every area of the business. Training and awareness are vital in the overall

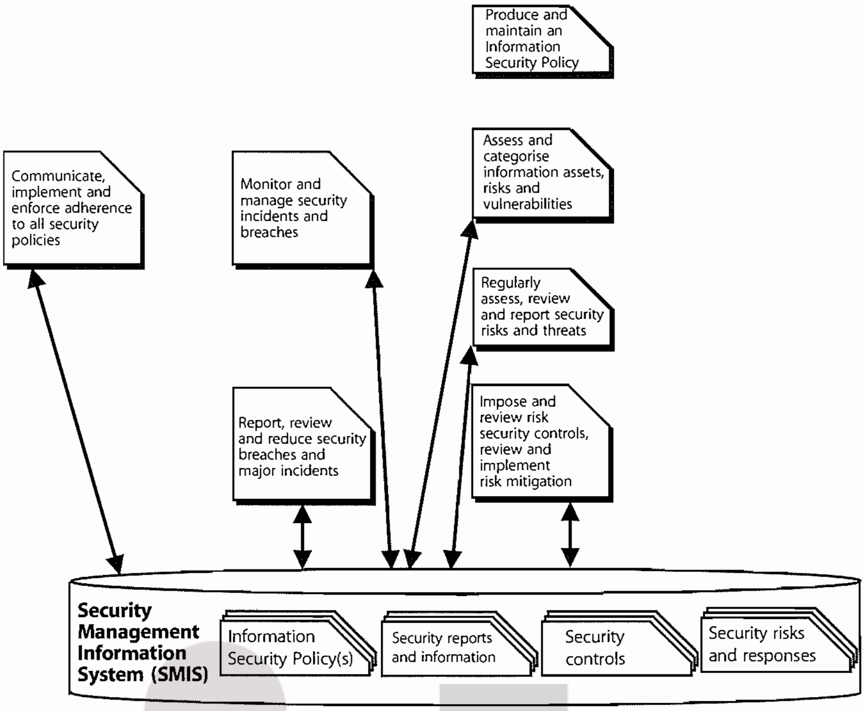

Produce and maintain an Information Security Policy
Communicate, implement and enforce adherence to all security policies
Monitor and manage security incidents and breaches
Assess and categorise information assets, risks and vulnerabilities
Regularly assess, review and report security risks and threats
Report, review and reduce security breaches and major incidents
Impose and review risk security controls, review and implement risk mitigation
Security Management Information System (SMIS)
Information Security Policy(s)
Security reports and information
Security controls
Security risks and responses

Figure 4.27 IT Security Management process

strategy, as security is often weakest at the end-user stage. It is here, as well, that there is a need to develop methods and processes that enable the policies and standards to be more easily followed and implemented.

Resources need to be assigned to track developments in these enabling technologies and the products they support. For example, privacy continues to be important and, increasingly, the focus of government regulation, making privacy compliance technologies an important enabling technology.

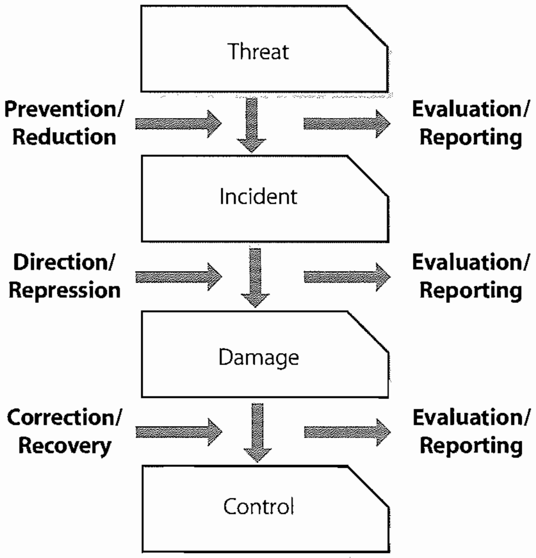


Figure 4.28 Security controls for threats and incidents

4.6.5.1 *Security controls*

The Information Security Manager must understand that security is not a step in the lifecycle of services and systems and that security cannot be solved through technology. Rather, information security must be an integral part of all services and systems and is an ongoing process that needs to be continuously managed using a set of security controls, as shown in Figure 4.28.

The set of security controls should be designed to support and enforce the Information Security Policy and to minimize all recognized and identified threats. The controls will be considerably more cost-effective if included within the design of all services. This will ensure the continued protection of all existing services and that new services and access to them are in line with the policy.

Security measures can be used at a specific stage in the prevention and handling of security incidents, as illustrated in Figure 4.28. Security incidents are not solely caused by technical threats – statistics show that, for example, the large majority stem from human errors (intended or not) or procedural errors, and often have implications in other fields such as safety, legal or health.

The following stages can be identified. At the start there is a risk that a threat will materialize. A threat can be anything that disrupts the business process or has negative impact on the business. When a threat materializes, we speak of a security incident. This security incident may result in damage (to information or to assets) that has to be repaired or otherwise corrected. Suitable measures can be selected for each of these stages. The choice of measures will depend on the importance attached to the information.

- **Preventive:** security measures are used to prevent a security incident from occurring. The best-known example of preventive measures is the allocation of access rights to a limited group of authorized people. The further requirements associated with this measure include the control of access rights (granting, maintenance and withdrawal of rights), authorization (identifying who is allowed access to which information and using which tools), identification and authentication (confirming who is seeking access) and access control (ensuring that only authorized personnel can gain access).
- **Reductive:** further measures can be taken in advance to minimize any possible damage that may occur. These are 'reductive' measures. Familiar examples of reduction measures are making regular backups and the development, testing and maintenance of contingency plans.
- **Detective:** if a security incident occurs, it is important to discover it as soon as possible – detection. A familiar example of this is monitoring, linked to an alert procedure. Another example is virus-checking software.
- **Repressive:** measures are then used to counteract any continuation or repetition of the security incident. For example, an account or network address is temporarily blocked after numerous failed attempts to log on or the retention of a card when multiple attempts are made with a wrong PIN number.
- **Corrective:** The damage is repaired as far as possible using corrective measures. For example, corrective measures include restoring the backup, or returning to a previous stable situation (roll-back, back-out). Fallback can also been seen as a corrective measure.

The documentation of all controls should be maintained to reflect accurately their operation, maintenance and their method of operation.

4.6.5.2 Management of security breaches and incidents

In the case of serious security breaches or incidents, an evaluation is necessary in due course, to determine what went wrong, what caused it and how it can be prevented in the future. However, this process should not be limited to serious security incidents. All breaches of security and security incidents need to be studied in order to gain a full picture of the effectiveness of the security measures as a whole. A reporting procedure for security incidents is required to be able to evaluate the effectiveness and efficiency of the present security measures based on an insight into all security incidents. This is facilitated by the maintenance of log files and audit files and, of course, the incident records of the Service Desk function. The analysis of these statistics on security issues should lead to improvement actions focused on the reduction of the impact and volume of all security breaches and incidents, in conjunction with Problem Management.

4.6.6 Triggers, inputs, outputs and interfaces

ISM activity can be triggered by many events. These include:

- New or changed corporate governance guidelines
- New or changed Business Security Policy
- New or changed corporate risk management processes and guidelines
- New or changed business needs or new or changed services
- New or changed requirements within agreements, such as SLRs, SLAs, OLAs or contracts
- Review and revision of business and IT plans and strategies
- Review and revision of designs and strategies
- Service or component security breaches or warnings, events and alerts, including threshold events, exception reports
- Periodic activities, such as reviewing, revising or reporting, including review and revision of ISM policies, reports and plans
- Recognition or notification of a change of risk or impact of a business process or VBF, an IT service or component
- Requests from other areas, particularly SLM for assistance with security issues.

The effective and efficient implementation of an Information Security Policy within an organization will, to a large extent, be dependent on good Service Management processes. Indeed, the effective implementation of some processes can be seen as a pre-requisite for effective security control. The key interfaces that ISM has with other processes are as follows:

- Incident and Problem Management: in providing assistance with the resolution and subsequent justification and correction of security incidents and problems. The Incident Management process must include the ability to identify and deal with security incidents. Service Desk and Service Operations staff must 'recognize' a security incident.

- ITSCM: with the assessment of business impact and risk, and the provision of resilience, fail-over and recovery mechanisms. Security is a major issue when continuity plans are tested or invoked. A working ITSCM plan is a mandatory requirement for ISO 27001.
- SLM: assistance with the determining of security requirements and responsibilities and their inclusion within SLRs and SLAs, together with the investigation and resolution of service and component security breaches.
- Change Management: ISM should assist with the assessment of every change for impact on security and security controls. Also ISM can provide information on unauthorized changes.
- Legal and HR issues must be considered when investigating security issues.
- Configuration Management will give the ability to provide accurate asset information to assist with security classifications. Having an accurate CMS is therefore an extremely useful ISM input.
- Security is often seen as an element of Availability Management, with Confidentiality Integrity and Availability (CIA) being the essence of Availability and ISM. Also, ISM should work with both Availability Management and ITSCM to conduct integrated Risk Analysis and Management exercises.
- Capacity Management must consider security implications when selecting and introducing new technology. Security is an important consideration when procuring any new technology or software.
- Financial Management should provide adequate funds to finance security requirements.
- Supplier Management should assist with the joint management of suppliers and their access to services and systems, and the terms and conditions to be included within contracts concerning supplier responsibilities.

4.6.6.1 Inputs

Information Security Management will need to obtain input from many areas, including:

- Business information: from the organization's business strategy, plans and financial plans, and information on their current and future requirements.
- Corporate governance and business security policies and guidelines, security plans, Risk Analysis and responses
- IT information: from the IT strategy and plans and current budgets
- Service information: from the SLM process with details of the services from the Service Portfolio and the Service Catalogue and service level targets within SLAs and SLRs, and possibly from the monitoring of SLAs, service reviews and breaches of the SLAs
- Risk Analysis processes and reports: from ISM, Availability Management and ITSCM
- Details of all security events and breaches: from all areas of IT and SM, especially Incident Management and Problem Management
- Change information: from the Change Management process with a Change Schedule and a need to assess all changes for their impact on all security policies, plans and controls
- CMS: containing information on the relationships between the business, the services, supporting services and the technology
- Details of partner and supplier access: from Supplier Management and Availability Management on external access to services and systems.

4.6.6.2 Outputs

The outputs produced by the Information Security Management process are used in all areas and should include:

- An overall Information Security Management Policy, together with a set of specific security policies
- A Security Management Information System (SMIS), containing all the information relating to ISM
- Revised security risk assessment processes and reports
- A set of security controls, together with details of the operation and maintenance and their associated risks
- Security audits and audit reports
- Security test schedules and plans, including security penetration tests and other security tests and reports
- A set of security classifications and a set of classified information assets
- Reviews and reports of security breaches and major incidents
- Policies, processes and procedures for managing partners and suppliers and their access to services and information.

4.6.7 Key Performance Indicators

Many KPIs and metrics can be used to assess the effectiveness and efficiency of the ISM process and activities. These metrics need to be developed from the service, customer and business perspective such as:

- Business protected against security violations:
  - Percentage decrease in security breaches reported to the Service Desk
  - Percentage decrease in the impact of security breaches and incidents
  - Percentage increase in SLA conformance to security clauses.
- The determination of a clear and agreed policy, integrated with the needs of the business: decrease in the number of non-conformances of the ISM process with the business security policy and process.
- Security procedures that are justified, appropriate and supported by senior management:
  - Increase in the acceptance and conformance of security procedures
  - Increased support and commitment of senior management.
- A mechanism for improvement:
  - The number of suggested improvements to security procedures and controls
  - Decrease in the number of security non-conformance detected during audits and security testing.
- Information security is an integral part of all IT services and all ITSM processes: increase in the number of services and processes conformant with security procedures and controls.
- Effective marketing and education in security requirements, IT staff awareness of the technology supporting the services:
  - Increased awareness of the security policy and its contents, throughout the organization
  - Percentage increase in completeness of the technical Service Catalogue against IT components supporting the services
  - Service Desk supporting all services.

4.6.8 Information Management

All the information required by ISM should be contained within the SMIS. This should include all security controls, risks, breaches, processes and reports necessary to support and maintain the Information Security Policy and the ISMS. This information should cover all IT services and components and needs to be integrated and maintained in alignment with all other IT information management systems, particularly the Service Portfolio and the CMS. The SMIS will also provide the input to security audits and reviews and to the continual improvement activities so important to all ISMSs. The SMIS will also provide invaluable input to the design of new systems and services.

4.6.9 Challenges, Critical Success Factors and risks

ISM faces many challenges in establishing an appropriate Information Security Policy with an effective supporting process and controls. One of the biggest challenges is to ensure that there is adequate support from the business, business security and senior management. If these are not available, it will be impossible to establish an effective ISM process. If there is senior IT management support, but there is no support from the business, IT security controls and risk assessment will be severely limited in what they can achieve because of this lack of support from the business. It is pointless implementing security policies, procedures and controls in IT if these cannot be enforced throughout the business. The major use of IT services and assets is outside of IT, and so are the majority of security threats and risks.

In some organizations the business perception is that security is an IT responsibility, and therefore the business assumes that IT will be responsible for all aspects of IT security and that IT services will be adequately protected. However, without the commitment and support of the business and business personnel, money invested in expensive security controls and procedures will be largely wasted and they will mostly be ineffective.

If there is a business security process established, then the challenge becomes one of alignment and integration. ISM must ensure that accurate information is obtained from the business security process on the needs, risks, impact and priorities of the business and that the ISM policies, information and plans are aligned and integrated with those of the business. Having achieved that alignment, the challenge becomes one of keeping them aligned by management and control of business and IT change using strict Change Management and Configuration Management control. Again, this requires support and commitment from the business and senior management.

The main CSFs for the ISM process are:

- Business protected against security violations
- The determination of a clear and agreed policy, integrated with the needs of the business

- Security procedures that are justified, appropriate and supported by senior management
- Effective marketing and education in security requirements
- A mechanism for improvement
- Information security is an integral part of all IT services and all ITSM processes
- The availability of services is not compromised by security incidents
- Clear ownership and awareness of the security policies amongst the customer community.

Information systems can generate many direct and indirect benefits, and as many direct and indirect risks. These risks have led to a gap between the need to protect systems and services and the degree of protection applied. The gap is caused by internal and external factors, including the widespread use of technology, increasing dependence of the business on IT, increasing complexity and interconnectivity of systems, disappearance of the traditional organizational boundaries and increasingly onerous regulatory requirements.

This means that there are new risk areas that could have a significant impact on critical business operations, such as:

- Increasing requirements for availability and robustness
- Growing potential for misuse and abuse of information systems affecting privacy and ethical values
- External dangers from hackers, leading to denial-of-service and virus attacks, extortion, industrial espionage and leakage of organizational information or private data.

Because new technology provides the potential for dramatically enhanced business performance, improved and demonstrated information security can add real value to the organization by contributing to interaction with trading partners, closer customer relationships, improved competitive advantage and protected reputation. It can also enable new and easier ways to process electronic transactions and generate trust. In today's competitive global economy, if an organization wants to do business, it may well be asked to present details of its security posture and results of its past performance in terms of tests conducted to ensure security of its information resources.

Other areas of major risks associated with ISM include:

- A lack of commitment from the business to the ISM processes and procedures
- Lack of commitment from the business and a lack of appropriate information on future plans and strategies
- A lack of senior management commitment or a lack of resources and/or budget for the ISM process
- The processes focus too much on the technology issues and not enough on the IT services and the needs and priorities of the business
- Risk assessment and management is conducted in isolation and not in conjunction with Availability Management and ITSCM
- ISM policies, plans, risks and information become out-of-date and lose alignment with the corresponding relevant information and plans of the business and business security.

4.7 SUPPLIER MANAGEMENT

4.7.1 Purpose/goal/objective

> 'The goal of the Supplier Management process is to manage suppliers and the services they supply, to provide seamless quality of IT service to the business, ensuring value for money is obtained.'

The Supplier Management process ensures that suppliers and the services they provide are managed to support IT service targets and business expectations. The aim of this section is to raise awareness of the business context of working with partners and suppliers, and how this work can best be directed toward realising business benefit for the organization.

It is essential that Supplier Management processes and planning are involved in all stages of the Service Lifecycle, from strategy and design, through transition and operation, to improvement. The complex business demands require the complete breadth of skills and capability to support provision of a comprehensive set of IT services to a business, therefore the use of value networks and the suppliers and the services they provide are an integral part of any end-to-end solution. Suppliers and the management of suppliers and partners are essential to the provision of quality IT services.

The purpose of the Supplier Management process is to obtain value for money from suppliers and to ensure that suppliers perform to the targets contained within their contracts and agreements, while conforming to all of the terms and conditions.

The main objectives of the Supplier Management process are to:

- Obtain value for money from supplier and contracts
- Ensure that underpinning contracts and agreements with suppliers are aligned to business needs, and

support and align with agreed targets in SLRs and SLAs, in conjunction with SLM

- Manage relationships with suppliers
- Manage supplier performance
- Negotiate and agree contracts with suppliers and manage them through their lifecycle
- Maintain a supplier policy and a supporting Supplier and Contract Database (SCD).

4.7.2 Scope

The Supplier Management process should include the management of all suppliers and contracts needed to support the provision of IT services to the business. Each service provider should have formal processes for the management of all suppliers and contracts. However, the processes should adapt to cater for the importance of the supplier and/or the contract and the potential business impact on the provision of services. Many suppliers provide support services and products that independently have a relatively minor, and fairly indirect, role in value generation, but collectively make a direct and important contribution to value generation and the implementation of the overall business strategy. The greater the contribution the supplier makes to business value, the more effort the service provider should put into the management of the supplier and the more that supplier should be involved in the development and realization of the business strategy. The smaller the supplier's value contribution, the more likely it is that the relationship will be managed mainly at an operational level, with limited interaction with the business. It may be appropriate in some organizations, particularly large ones, to manage internal teams and suppliers, where different business units may provide support of key elements.

The Supplier Management process should include:

- Implementation and enforcement of the supplier policy
- Maintenance of a Supplier and Contract Database (SCD)
- Supplier and contract categorization and risk assessment
- Supplier and contract evaluation and selection
- Development, negotiation and agreement of contracts
- Contract review, renewal and termination
- Management of suppliers and supplier performance
- Agreement and implementation of service and supplier improvement plans
- Maintenance of standard contracts, terms and conditions

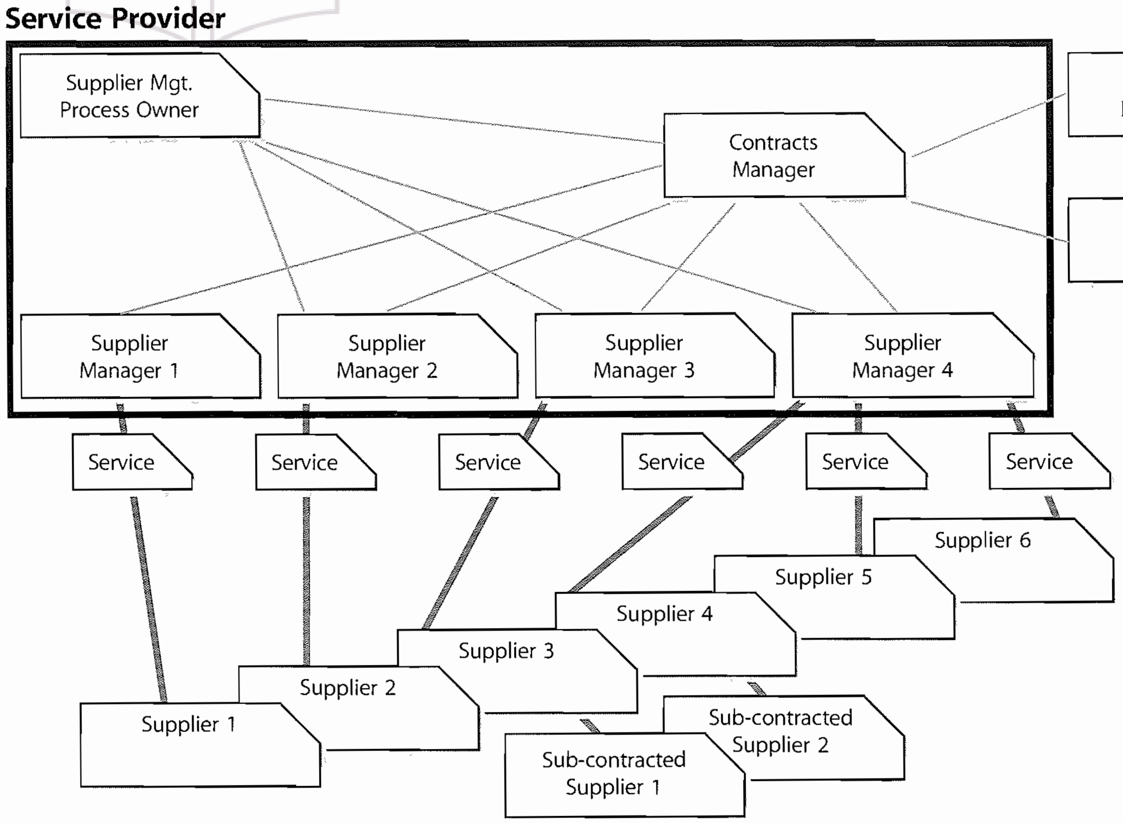


Figure 4.29 Supplier Management – roles and interfaces

- Management of contractual dispute resolution
- Management of sub-contracted suppliers.

IT Supplier Management often has to comply with organizational or corporate standards, guidelines and requirements, particularly those of corporate legal, finance and purchasing, as illustrated in Figure 4.29.

In order to ensure that suppliers provide value for money and meet their service targets, the relationship between each supplier should be owned by an individual within the service provider organization. However, a single individual may own the relationship for one or many suppliers, as illustrated in Figure 4.29. To ensure that relationships are developed in a consistent manner and that suppliers' performance is appropriately reviewed and managed, roles need to be established for a Supplier Management process owner and a Contracts Manager. In smaller organizations, these separate roles may be combined into a single responsibility.

4.7.3 Value to the business

The main objectives of the Supplier Management process are to provide value for money from suppliers and contracts and to ensure that all targets in underpinning supplier contracts and agreements are aligned to business needs and agreed targets within SLAs. This is to ensure the delivery to the business of end-to-end, seamless, quality IT services that are aligned to the business's expectation. The Supplier Management process should align with all corporate requirements and the requirements of all other IT and SM processes, particularly ISM and ITSCM. This ensures that the business obtains value from supporting supplier services and that they are aligned with business needs.

4.7.4 Policies/principles/basic concepts

The Supplier Management process attempts to ensure that suppliers meet the terms, conditions and targets of their contracts and agreements, whilst trying to increase the value for money obtained from suppliers and the services they provide. All Supplier Management process activity should be driven by a supplier strategy and policy from Service Strategy. In order to achieve consistency and effectiveness in the implementation of the policy, a Supplier and Contracts Database (SCD) should be established, as illustrated in Figure 4.30, together with clearly defined roles and responsibilities.

Ideally the SCD should form an integrated element of a comprehensive CMS or SKMS, recording all supplier and contract details, together with details of the type of service(s) or product(s) provided by each supplier, and all other information and relationships with other associated CIs. The services provided by suppliers will also form a key part of the Service Portfolio and the Service Catalogue. The relationship between the supporting services and the IT and business services they support are key to providing quality IT services.

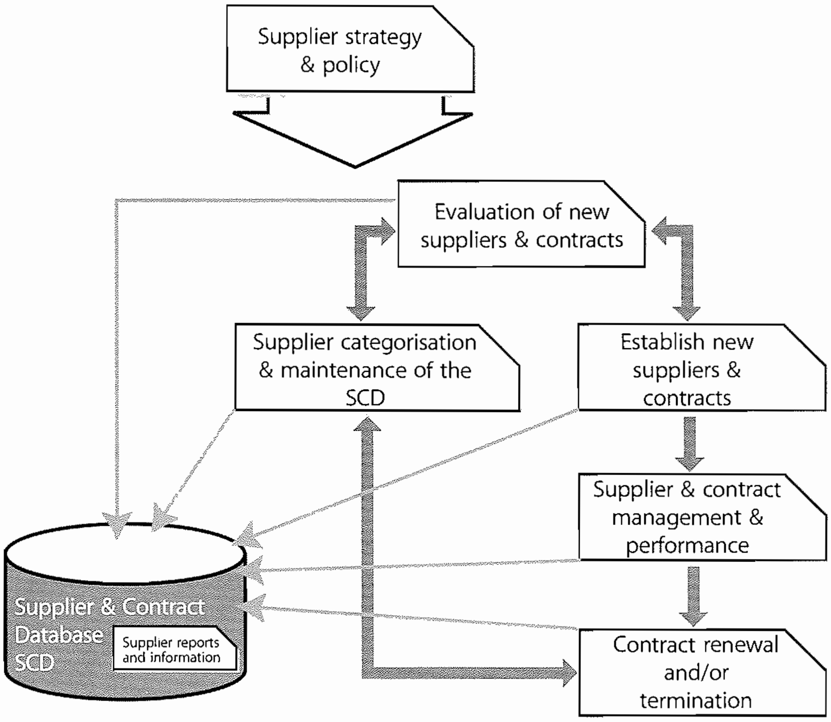

Supplier strategy & policy
Evaluation of new suppliers & contracts
Supplier categorisation & maintenance of the SCD
Establish new suppliers & contracts
Supplier & contract management & performance
Contract renewal and/or termination
Supplier & Contract Database SCD
Supplier reports and information

Figure 4.30 Supplier Management process

This information within the SCD will provide a complete set of reference information for all Supplier Management procedures and activities:

- Supplier categorization and maintenance of the Supplier and Contracts Database (SCD)
- Evaluation and set-up of new suppliers and contracts
- Establishing new suppliers
- Supplier and Contract Management and performance
- Contract renewal and termination.

The first two elements within the above list are covered within the Service Design stage. The third element is part of Service Transition, and the last two are part of the Service Operation stage and are covered in more detail in those publications.

4.7.5 Process activities, methods and techniques

This section provides more detail on the Supplier Management process, its sub-processes and activities, and the management of the contract lifecycle.

When dealing with external suppliers, it is strongly recommended that a formal contract with clearly defined, agreed and documented responsibilities and targets is established and managed through the stages of its lifecycle, from the identification of the business need to the operation and cessation of the contract:

- Identification of business need and preparation of the business case:
  - Produce a Statement of Requirement (SOR) and/or Invitation To Tender (ITT)
  - Ensure conformance to strategy/policy
  - Prepare the initial business case, including options (internal and external), costs, timescales, targets, benefits, risk assessment.
- Evaluation and procurement of new contracts and suppliers:
  - Identify method of purchase or procurement
  - Establish evaluation criteria – for example, services, capability (both personnel and organization), quality and cost
  - Evaluate alternative options
  - Select
  - Negotiate contracts, targets and the terms and conditions, including responsibilities, closure, renewal, extension, dispute, transfer
  - Agree and award the contract.
- Establish new suppliers and contracts:
  - Set up the supplier service and contract, within the SCD and any other associated corporate systems
  - Transition of service
  - Establish contacts and relationships.
- Supplier and contract categorization:
  - Assessment or reassessment of the supplier and contract
  - Ensure changes progressed through Service Transition
  - Categorization of the supplier
  - Update of SCD
  - Ongoing maintenance of the SCD.
- Manage the supplier and contract performance:
  - Management and control of the operation and delivery of service/products
  - Monitor and report (service, quality and costs)
  - Review and improve (service, quality and costs)
  - Management of the supplier and the relationship (communication, risks, changes, failures, improvements, contacts, interfaces)
  - Review, at least annually, service scope against business need, targets and agreements
  - Plan for possible closure/renewal/extension.
- End of term:
  - Review (determine benefits delivered, ongoing requirement)
  - Renegotiate and renew or terminate and/or transfer.

The business, IT, finance, purchasing and procurement need to work together to ensure that all stages of the contract lifecycle are managed effectively. All areas need to be jointly involved in selecting the solution and managing the ongoing performance of the supplier, with each area taking responsibility for the interests of their own area, whilst being aware of the implications on the organization as a whole. The processes involved in the stages of the contract lifecycle are explained in detail in the following sections.

4.7.5.1 Evaluation of new suppliers and contracts

The activities associated with the identification of business needs and the subsequent evaluation of new suppliers and contracts are part of the Service Design process. The outputs from this area provide the inputs to all other stages of the contract lifecycle. IT is vital to the ongoing success of the contract and the relationship that the business is closely involved in all aspects of these activities. Every organization should have templates and a formal method for the production of business cases and their approval and sign-off. The detailing of the business needs and the content of the business case should be agreed, approved and signed off by both the business and IT.

When selecting a new supplier or contract, a number of factors need to be taken into consideration, including track record, capability, references, credit rating and size relative to the business being placed. In addition, depending on the type of supplier relationship, there may be personnel issues that need to be considered. Each organization should have processes and procedures for establishing new suppliers and contracts.

While it is recognized that factors may exist that influence the decision on type of relationship or choice of supplier (e.g. politics within the organization, existing relationships), it is essential that in such cases the reasoning is identified and the impact fully assessed to ensure costly mistakes are avoided.

Services may be sourced from a single supplier or multi-sourced. Services are most likely to be sourced from two or more competing suppliers where the requirement is for standard services or products that are readily available 'off-the-shelf'. Multi-sourcing is most likely to be used where cost is the prime determinant, and requirements for developing variants of the services are low, but may also be undertaken to spread risk. Suppliers on a multi-source list may be designated with 'Preferred Supplier' status within the organization, limiting or removing scope for use of other suppliers.

Partnering relationships are established at an executive level and are dependent on a willingness to exchange strategic information to align business strategies. Many strategically important supplier relationships are now positioned as partnering relationships. This reflects a move away from traditionally hierarchical relationships, where the supplier acts subordinately to the customer organization, to one characterized by:

- **Strategic alignment:** good alignment of culture, values and objectives, leading to an alignment of business strategies
- **Integration:** a close integration of the processes of the two organizations
- **Information flow:** good communication and information exchange at all levels, especially at the strategic level, leading to close understanding
- **Mutual trust:** a relationship built on mutual trust between the organizations and their individuals
- **Openness:** when reporting on service performance, costs and Risk Analysis
- **Collective responsibility:** joint partnership teams taking collective responsibility for current performance and future development of the relationship
- **Shared risk and reward:** e.g. agreeing how investment costs and resultant efficiency benefits are shared, or how risks and rewards from fluctuations in material costs are shared.

Both parties derive benefits from partnering. An organization derives progressively more value from a supplier relationship as the supplier's understanding of the organization as a whole increases, from its IT inventory architectures through to its corporate culture, values and business objectives. With time, the supplier is able to respond more quickly and more appropriately to the organization's needs. The supplier benefits from a longer-term commitment from the organization, providing it with greater financial stability, and enabling it to finance longer-term investments, which benefit its customers.

A partnership makes it possible for the parties to align their IT infrastructures. Joint architecture and risk control agreements allow the partners to implement a range of compatible solutions from security, networking, data/information interchange, to workflow and application processing systems. This integration can provide service improvements and lowered costs. Such moves also reduce risks and costs associated with one-off tactical solutions, put in place to bridge a supplier's IT with that of the organization.

The key to a successful partnering relationship is being absolutely clear about the benefits and costs such a relationship will deliver before entering into it. Both parties then know what is expected of them at the outset. The success of the partnership may involve agreeing the transfer of staff to the partner or outsourcing organization as part of the agreement and relationship.

Service provider organizations should have documented and formal processes for evaluating and selecting suppliers based on:

- Importance and impact: the importance of the service to the business, provided by the supplier
- Risk: the risks associated with using the service
- Costs: the cost of the service and its provision.

Often other areas of the service provider organization, such as Legal, Finance and Purchasing, will get involved with this aspect of the process. Service provider organizations should have processes covering:

- Production of business case documents
- Production of SoR and Invitations to Tender or proposal documents
- Formal evaluation and selection of suppliers and contracts
- The inclusion of standard clauses, terms and conditions within contracts, including early termination, benchmarking, exit or transfer of contracts, dispute resolution, management of sub-contracted suppliers and normal termination
- Transitioning of new contracts and suppliers.

These processes may, and should be, different, based on the type, size and category of the supplier and the contract.

The nature and extent of an agreement depends on the relationship type and an assessment of the risks involved. A pre-agreement Risk Analysis is a vital stage in establishing any external supplier agreement. For each party, it exposes the risks that need to be addressed and needs to be as comprehensive as practical, covering a wide variety of risks, including financial, business reputation, operational, regulatory and legal.

A comprehensive agreement minimizes the risk of disputes arising from a difference of expectations. A flexible agreement, which adequately caters for its adaptation across the term of the agreement, is maintainable and supports change with a minimum amount of renegotiation.

The contents of a basic underpinning contract or service agreement are as follows:

- **Basic terms and conditions:** the term (duration) of the contract, the parties, locations, scope, definitions and commercial basis.
- **Service description and scope:** the functionality of the services being provided and its extent, along with constraints on the service delivery, such as performance, availability, capacity, technical interface and security. Service functionality may be explicitly defined, or in the case of well-established services, included by reference to other established documents, such as the Service Portfolio and the Service Catalogue.
- **Service standards:** the service measures and the minimum levels that constitute acceptable performance and quality, e.g. IT may have a performance requirement to respond to a request for a new desktop system in 24 hours, with acceptable service deemed to have occurred where this performance requirement is met in 95% of cases. Service levels must be realistic, measurable and aligned to the organization's business priorities and underpin the agreed targets within SLRs and SLAs.
- **Workload ranges:** the volume ranges within which service standards apply, or for which particular pricing regimes apply.
- **Management Information (MI):** the data that must be reported by the supplier on operational performance – take care to ensure that MI is focused on the most important or headline reporting measures on which the relationship will be assessed. Key Performance Indicators (KPIs) and Balanced Scorecards (BSCs) may form the core of reported performance data.
- **Responsibilities and dependencies:** description of the obligations of the organization (in supporting the supplier in the service delivery efforts) and of the supplier (in its provision of the service), including communication, contacts and escalation.

An extended service agreement may also contain:

- Service debit and credit regime (incentives and penalties)
- Additional performance criteria.

The following gives a limited sample of the legal and commercial topics typically covered by a service or contractual agreement:

- Scope of services to be provided
- Service performance requirements
- Division and agreement of responsibilities
- Contact points, communication and reporting frequency and content
- Contract review and dispute resolution processes
- Price structure
- Payment terms
- Commitments to change and investment
- Agreement change process
- Confidentiality and announcements
- Intellectual property rights and copyright
- Liability limitations
- Termination rights of each party
- Obligations at termination and beyond.

The final form of an agreement, and some of the terminology, may be dictated by the views and preferences of the procurement and legal departments, or by specialist legal firms.

Tip

Seek legal advice when formalizing external supply agreements.

Formal contracts

Formal contracts are appropriate for external supply arrangements that make a significant contribution to the delivery and development of the business. Contracts provide for binding legal commitments between customer and supplier, and cover the obligations each organization has to the other from the first day of the contract, often extending beyond its termination. A contract is used as the basis for external supplier agreements where an enforceable commitment is required. High-value and/or strategic relationships are underpinned by a formal contract. The formality and binding nature of a contract are not at odds with the culture of a partnering agreement, but rather form the basis on which trust in the relationship may be founded.

A contract is likely to be structured with a main body containing the commercial and legal clauses, and with the elements of a service agreement, as described earlier, attached as schedules. Contracts may also include a number of other related documents as schedules, for example:

- Security requirements
- Business continuity requirements
- Mandated technical standards
- Migration plans (agreed pre-scheduled change)
- Disclosure agreements.

Most large organizations have procurement and legal departments specializing in sourcing contracts. Specialist legal firms may be employed to support the internal procurement and legal function when establishing significant formal contracts.

Underpinning agreements

In ITIL an SLA is defined as a 'written agreement between a service provider and the customer(s) that documents agreed service levels for a service'. Service providers should be aware that SLAs are widely used to formalize service-based relationships, both internally and externally, and that while conforming to the definition above, these agreements vary considerably in the detail covered.

Key message

The views of some organizations, such as the Chartered Institute of Purchase and Supply (CIPS) and various specialist lawyers, are that SLAs ought not to be used to manage external relationships unless they form part of an underlying contract. The Complete Guide to Preparing and Implementing Service Level Agreements (2001) emphasizes that a stand-alone SLA may not be legally enforceable but instead 'represents the goodwill and faith of the parties signing it'. Therefore it is in service providers' and suppliers' interests to ensure that SLAs are incorporated into an appropriate contractual framework that meets the ITIL objective that SLAs are binding agreements.

SLAs, underpinning agreements and contracts should be reviewed on a regular basis to ensure performance conforms to the service levels that have been agreed.

The organization is likely to be dependent on its own internal support groups to some extent. To be able to achieve SLA targets, it is advisable to have formal arrangements in place with these groups. Operational Level Agreements (OLAs) ensure that underpinning services support the business/IT SLA targets. OLAs focus on the operational requirements that the services need to meet. This is a non-contractual, service-oriented document describing services and service standards, with responsibilities and obligations where appropriate.

Just as with SLAs, it is important that OLAs are monitored to highlight potential problems. The Service Level Manager has the overall responsibility to review performance against targets so that action can be taken to remedy, and prevent future recurrence of, any OLA breaches. Depending on the size of the organization and variety of services, e.g. SLAs and OLAs, a Service Level Manager should take responsibility for their service or set of services.

4.7.5.2 Supplier categorization and maintenance of the Supplier and Contracts Database (SCD)

The Supplier Management process should be adaptive and spend more time and effort managing key suppliers than less important suppliers. This means that some form of categorization process should exist within the Supplier Management process to categorize the supplier and their importance to the service provider and the services provided to the business. Suppliers can be categorized in many ways, but one of the best methods for categorizing

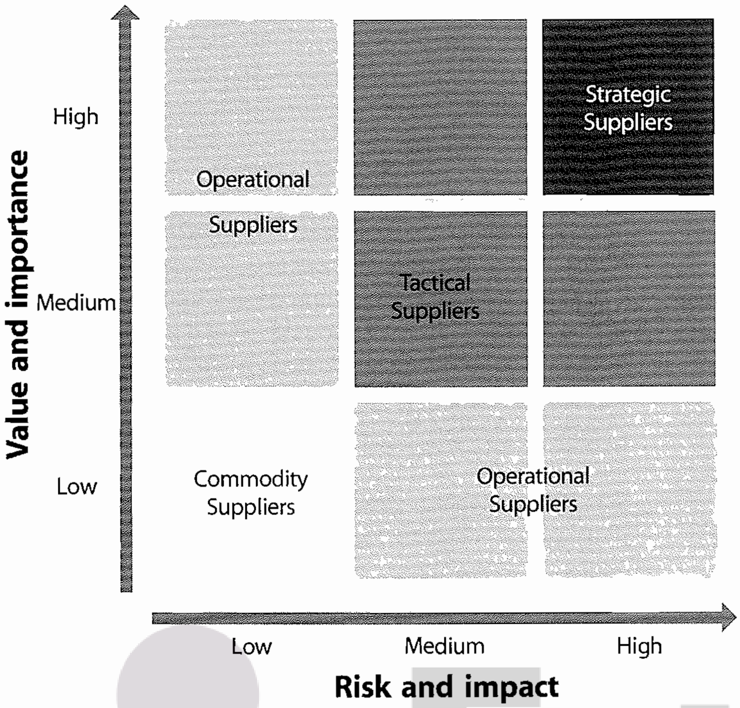


Figure 4.31 Supplier categorization

suppliers is based on assessing the risk and impact associated with using the supplier, and the value and importance of the supplier and their services to the business, as illustrated in Figure 4.31.

The amount of time and effort spent managing the supplier and the relationship can then be appropriate to its categorization:

- **Strategic:** for significant 'partnering' relationships that involve senior managers sharing confidential strategic information to facilitate long-term plans. These relationships would normally be managed and owned at a senior management level within the service provider organization, and would involve regular and frequent contact and performance reviews. These relationships would probably require involvement of Service Strategy and Service Design resources, and would include ongoing specific improvement programmes (e.g. a network service provider, supplying worldwide networks service and their support).
- **Tactical:** for relationships involving significant commercial activity and business interaction. These relationships would normally be managed by middle management and would involve regular contact and performance reviews, often including ongoing improvement programmes (e.g. a hardware maintenance organization providing resolution of server hardware failures).
- **Operational:** for suppliers of operational products or services. These relationships would normally be managed by junior operational management and would involve infrequent but regular contact and performance reviews (e.g. an internet hosting service provider, supplying hosting space for a low-usage, low-impact website or internally used IT service).
- **Commodity:** for suppliers that provide low-value and/or readily available products and services, which could be alternatively sourced relatively easily (e.g. paper or printer cartridge suppliers).

Strategically important supplier relationships are given the greatest focus. It is in these cases that Supplier Managers have to ensure that the culture of the service provider organization is extended into the supplier domain so that the relationship works beyond the initial contract. The rise in popularity of external sourcing, and the increase in the scope and complexity of some sourcing arrangements, has resulted in a diversification of types of supplier relationship. At a strategic level, it is important to understand the options that are available so that the most suitable type of supplier relationship can be established to gain maximum business benefit and evolves in line with business needs.

Tip

To successfully select the most appropriate type of supplier relationship, there needs to be a clear understanding of the business objectives that are to be achieved.

A number of factors, from the nature of the service to the overall cost, determine the importance of a supplier from a business perspective. As shown later, the greater the business significance of a supplier relationship, the more the business needs to be involved in the management and development of a relationship. A formal categorization approach can help to establish this importance.

The business value, measured as the contribution made to the business value chain, provides a more business-aligned assessment than pure contract price. Also, the more standard the services being procured, the lower the dependence the organization has on the supplier, and the more readily the supplier could be replaced (if necessary). Standardized services support the business through minimal time to market when deploying new or changed business services, and in pursuing cost-reduction strategies. More information on this subject can be found in the Service Strategy publication.

The more customized those services are, the greater the difficulty in moving to an alternative supplier. Customization may benefit the business, contributing to competitive advantage through differentiated service, or may be the result of operational evolution.

Tailored services increase the dependence on the supplier, increase risk and can result in increased cost. From a supplier perspective, tailored services may decrease their ability to achieve economies of scale through common operations, resulting in narrowed margins, and reduced capital available for future investment.

Standard products and services are the preferred approach unless a clear business advantage exists, in which case a strategic supplier delivers the tailored service.

Tip

High-value or high-dependence relationships involve greater risks for the organization. These relationships need comprehensive contracts and active relationship management.

Having established the type of supplier, the relationship then needs to be formalized. In the discussion below, the term 'agreement' is used generically to refer to any formalization of a relationship between customer and supplier organizations, and may range from the informal to comprehensive legally binding contracts. Simple, low-value relationships may be covered by a supplier's standard terms and conditions, and be managed wholly by IT. A relationship of strategic importance to the business, on the other hand, requires a comprehensive contract that ensures that the supplier supports evolving business needs throughout the life of the contract. A contract needs to be managed and developed in conjunction with procurement and legal departments and business stakeholders.

Tips

The agreement is the foundation for the relationship. The more suitable and complete the agreement, the more likely it is that the relationship will deliver business benefit to both parties.

The quality of the relationship between the service provider and their supplier(s) is often dependent on the individuals involved from both sides. It is therefore vital that individuals with the right attributes, skills, competences and personalities are selected to be involved in these relationships.

A business service may depend on a number of internal and/or external suppliers for its delivery. These may include a mixture of strategic suppliers and commodity suppliers. Some suppliers supply directly to the organization; others are indirect or sub-contracted suppliers working via another supplier. Direct suppliers are directly managed by the service provider; indirect or sub-contracted suppliers are managed by the leading supplier. Any one supplier may provide products or services used to support a number of different business services.

Supply chain analysis shows the mapping between business services and supplier services. Analysis of business processes will reveal the suppliers involved in each process and the points of hand-off between them. Management of the supply chain ensures that functional boundaries and performance requirements are clearly established for each supplier to ensure that overall business service levels are achieved. Business services are most likely to meet their targets consistently where there are a small number of suppliers in the supply chain, and where the interfaces between the suppliers in the chain are limited, simple and well-defined.

Reducing the number of direct suppliers reduces the number of relationships that need to be managed, the number of peer-to-peer supplier issues that need to be resolved, and reduces the complexity of the Supplier Management activities. Some organizations may successfully reduce or collapse the whole supply chain around a single service provider, often referred to as a 'prime' supplier. Facilities management is often outsourced to a single specialist partner or supplier, who may in turn subcontract restaurant services, vending machine maintenance and cleaning.

Outsourcing entire business services to a single 'prime supplier' may run additional risks. For these reasons, organizations need to consider carefully their supply chain strategies ahead of major outsourcing activity. The scope of outsourced services needs to be considered to reduce the number of suppliers, whilst ensuring that risk is managed and it fits with typical competencies in the supply market.

The SCD is a database containing details of the organization's suppliers, together with details of the products and services that they provide to the business (e.g. e-mail service, PC supply and installation, Service Desk), together with details of the contracts. The SCD contains supplier details, a summary of each product/service (including support arrangements), information on the ordering process and, where applicable, contract details. Ideally the SCD should be contained within the overall CMS.

SCDs are beneficial because they can be used to promote preferred suppliers and to prevent purchasing of unapproved or incompatible items. By coordinating and controlling the buying activity, the organization is more likely to be able to negotiate preferential rates.

4.7.5.3 Establishing new suppliers and contracts

Adding new suppliers or contracts to the SCD needs to be handled via the Change Management process, to ensure that any impact is assessed and understood. In most organizations, the SCD is owned by the Supplier Management process or the procurement or purchasing department. The SCD provides a single, central focal set of information for the management of all suppliers and contracts.

Risk management, working with suppliers, centres on assessing vulnerabilities in each supplier arrangement or contract that pose threats to any aspect of the business, including business impact, probability, customer satisfaction, brand image, market share, profitability, share price or regulatory impacts or penalties (in some industries).

The nature of the relationship affects the degree of risk to the business. Risks associated with an outsourced or strategic supplier are likely to be greater in number, and more complex to manage, than with internal supply. It is rarely possible to 'outsource' risk, although sometimes some of the risk may be transferred to the outsourcing organization. Blaming a supplier does not impress customers or internal users affected by a security incident or a lengthy system failure. New risks arising from the relationship need to be identified and managed, with communication and escalation as appropriate.

A substantial risk assessment should have been undertaken pre-contract, but this needs to be maintained in the light of changing business needs, changes to the contract scope, or changes in the operational environment.

The service provider organization and the supplier must consider the threats posed by the relationship to their own assets, and have their own risk profile. Each must identify their respective risk owners. In a well-functioning relationship, it is possible for much or all of the assessment to be openly shared with the other party. By involving supplier experts in risk assessments, especially in Operational Risk Assessments (ORAs), the organization may gain valuable insights into how best to mitigate risks, as well as improving the coverage of the assessment.

When evaluating risks of disruption to business services or functions, the business may have different priorities for service/function restoration. Business Impact Analysis (BIA) is a method used to assess the impacts on different areas of the business, resulting from a loss of service. Risk assessment and BIA activities relating to suppliers and contracts should be performed in close conjunction with Service Continuity Management, Availability Management and Information Security Management, with a view to reducing the impact and probability of service failure as a result of supplier or supplier service failure.

Once these activities have been completed and the supplier and contract information has been input into the SCD, including the nominated individuals responsible for managing the new supplier and/or contracts, frequency of service/supplier review meetings and contractual review meetings needs to be established, with appropriate break points, automated thresholds and warnings in place. The introduction of new suppliers and contracts should be handled as major changes through transition and into operation. This will ensure that appropriate contacts and communication points are established.

4.7.5.4 Supplier and Contract Management and performance

At an operational level, integrated processes need to be in place between an organization and its suppliers to ensure efficient day-to-day working practices. For example:

- Is the supplier expected to conform to the organization's Change Management process or any other processes?
- How does the Service Desk notify the supplier of incidents?
- How is CMS information updated when CIs change as a result of supplier actions? Who is responsible?

There may be a conflict of interest between the service provider organization and their supplier, especially with regard to the Change Management, Incident Management, Problem Management and Configuration Management processes. The supplier may want to use their processes and systems, whereas the service provider organization will want to use their own processes and systems. If this is the case, clear responsibilities and interfaces will need to be defined and agreed.

These and many other areas need to be addressed to ensure smooth and effective working at an operational level. To do so, all touch points and contacts need to be identified and procedures put in place so that everyone understands their roles and responsibilities. This should include identification of the single, nominated individual responsible for ownership of each supplier and contract. However, an organization should take care not to automatically impose its own processes, but to take the opportunity to learn from its suppliers.

Example

A contract had been awarded for a customized Stores Control System for which the organization's IT department had developed processes to support the live service once it was installed. This included procedures for recording and documenting work done on the service by field engineers (e.g. changes, repairs, enhancement and reconfigurations). At a project progress meeting, the supplier confirmed that they had looked at the procedures and could follow them if required. However, having been in this situation many times before, they had already developed a set of procedures to deal with such events. These procedures were considerably more elegant, effective and easier to follow than those developed and proposed by the organization.

In addition to process interfaces, it is essential to identify how issues are handled at an operational level. By having clearly defined and communicated escalation routes, issues are likely to be identified and resolved earlier, minimizing the impact. Both the organization and the supplier benefit from the early capture and resolution of issues.

Both sides should strive to establish good communication links. The supplier learns more about the organization's business, its requirements and its plans, helping the supplier to understand and meet the organization's needs. In turn, the organization benefits from a more responsive supplier who is aware of the business drivers and any issues, and is therefore more able to provide appropriate solutions. Close day-to-day links can help each party to be aware of the other's culture and ways of working, resulting in fewer misunderstandings and leading to a more successful and long-lasting relationship.

Two levels of formal review need to take place throughout the contract lifecycle to minimize risk and ensure the business realizes maximum benefit from the contract:

- **Service/supplier performance reviews:** reports on performance should be produced on a regular basis, based on the category of supplier, and should form the basis of service review meetings. The more important the supplier, the more frequent and extensive the reports and reviews should be
- **Service, service scope and contract reviews:** these should also be conducted on a regular basis, at least annually for all major suppliers. The objective of these should be to review the service, overall performance, service scope and targets and the contract, together with any associated agreements. This should be compared with the original business needs and the current business needs to ensure that supplier and contracts remain aligned to business needs and continue to deliver value for money.

Formal performance review meetings must be held on a regular basis to review the supplier's performance against service levels, at a detailed operational level. These meetings provide an opportunity to check that the ongoing service performance management remains focused on supporting business needs. Typical topics include:

- Service performance against targets
- Incident and problem reviews, including any escalated issues
- Business and customer feedback

- Expected major changes that will (or may) affect service during the next service period, as well as failed changes and changes that caused incidents
- Key business events over the next service period that need particular attention from the supplier (e.g. quarter-end processing)
- Best practice and Service Improvement Programmes (SIPs).

Major service improvement initiatives and actions are controlled through SIPs with each supplier, including any actions for dealing with any failures or weaknesses. Progress of existing SIPs, or the need for a new initiative, is reviewed at service review meetings. Proactive or forward-thinking organizations not only use SIPs to deal with failures but also to improve a consistently achieved service. It is important that a contract provides suitable incentives to both parties to invest in service improvement. These aspects are covered in more detail in the Continual Service Improvement publication.

The governance mechanisms for suppliers and contracts are drawn from the needs of appropriate stakeholders at different levels within each organization, and are structured so that the organization's representatives face-off to their counterparts in the supplier's organization. Defining the responsibilities for each representative, meeting forums and processes ensure that each person is involved at the right time in influencing or directing the right activities.

The scale and importance of the service and/or supplier influence the governance arrangements needed. The more significant the dependency, the greater the commitment and effort involved in managing the relationship. The effort needed on the service provider side to govern an outsourcing contract should not be underestimated, especially in closely regulated industries, such as the finance and pharmaceutical sectors.

A key objective for Supplier Management is to ensure that the value of a supplier to the organization is fully realized. Value is realized through all aspects of the relationship, from operational performance assurance, responsiveness to change requests and demand fluctuations, through to contribution of knowledge and experience to the organization's capability. The service provider must also ensure that the supplier's priorities match the business's priorities. The supplier must understand which of its service levels are most significant to the business.

Example

A large multi-national company had software agreements in place with the same supplier in no less than 24 countries. By arranging a single global licensing deal with the supplier, the company made annual savings of £5,000,000.

To ensure that all activities and contacts for a supplier are consistent and coordinated, each supplier relationship should have a single nominated individual accountable for all aspects of the relationship.

Example

A nationwide retail organization had an overall individual owning the management of their major network services supplier. However, services, contracts and billing were managed by several individuals spread throughout the organization. The individual owner put forward a business case for single ownership of the supplier and all the various contracts, together with consolidation of all the individual invoices into a single quarterly bill. The estimated cost savings to the organization were in excess of £600,000 per annum.

Satisfaction surveys also play an important role in revealing how well supplier service levels are aligned to business needs. A survey may reveal instances where there is dissatisfaction with the service, yet the supplier is apparently performing well against its targets (and vice versa). This may happen where service levels are inappropriately defined and should result in a review of the contracts, agreements and targets. Some service providers publish supplier league tables based on their survey results, stimulating competition between suppliers.

For those significant supplier relationships in which the business has a direct interest, both the business (in conjunction with the procurement department) and IT will have established their objectives for the relationship, and defined the benefits they expect to realize. This forms a major part of the business case for entering into the relationship.

These benefits must be linked and complementary, and must be measured and managed. Where the business is seeking improvements in customer service, IT supplier relationships contributing to those customer services must be able to demonstrate improved service in their own domain, and how much this has contributed to improved customer service.

For benefits assessments to remain valid during the life of the contract, changes in circumstances that have occurred since the original benefits case was prepared must be taken into account. A supplier may have been selected on its ability to deliver a 5% saving of annual operational cost compared with other options, but after two years has delivered no savings. However, where this is due to changes to contract, or general industry costs that would have also affected the other options, it is likely that a relative cost saving is still being realized. A maintained benefits case shows that saving.

Benefits assessments often receive lower priority than cost-saving initiatives, and are given less priority in performance reports than issues and problem summaries, but it is important to the long-term relationship that achievements are recognized. A benefits report must make objective assessments against the original objectives, but may also include morale-boosting anecdotal evidence of achievements and added value.

Tip

It is important for both organizations, and for the longevity of the relationship, that the benefits being derived from the relationship are regularly reviewed and reported.

An assessment of the success of a supplier relationship, from a business perspective, is likely to be substantially based on financial performance. Even where a service is performing well, it may not be meeting one or both parties' financial targets. It is important that both parties continue to benefit financially from the relationship. A contract that squeezes the margins of a supplier too tightly may lead to under-investment by the supplier, resulting in a gradual degradation of service, or even threaten the viability of supplier. In either case this may result in adverse business impacts to the organization.

The key to the successful long-term Financial Management of the contract is a joint effort directed towards maintaining the financial equilibrium, rather than a confrontational relationship delivering short-term benefits to only one party.

Building relationships takes time and effort. As a result, the organization may only be able to build long-term relationships with a few key suppliers. The experience, culture and commitment of those involved in running a supplier relationship are at least as important as having a good contract and governance regime. The right people with the right attitudes in the relationship team can make a poor contract work, but a good contract does not ensure that a poor relationship team delivers.

A considerable amount of time and money is normally invested in negotiating major supplier deals, with more again at risk for both parties if the relationship is not successful. Both organizations must ensure that they invest suitably in the human resources allocated to managing the relationship. The personality, behaviours and culture of the relationship representatives all influence the relationship. For a partnering relationship, all those involved need to be able to respect and work closely and productively with their opposite numbers.

4.7.5.5 Contract renewal and/or termination

Contract reviews must be undertaken on a regular basis to ensure the contract is continuing to meet business needs. Contract reviews assess the contract operation holistically and at a more senior level than the service reviews that are undertaken at an operational level. These reviews should consider:

- How well the contract is working and its relevance for the future
- Whether changes are needed: services, products, contracts, agreements, targets
- What is the future outlook for the relationship – growth, shrinkage, change, termination, transfer, etc?
- Commercial performance of the contract, reviews against benchmarks or market assessments, suitability of the pricing structure and charging arrangements
- Guidance on future contract direction and ensuring best practice management processes are established
- Supplier and contract governance.

For high-value, lengthy or complex supply arrangements, the period of contract negotiation and agreement can be lengthy, costly and may involve a protracted period of negotiation. It can be a natural inclination to wish to avoid further changes to a contract for as long as possible. However, for the business to derive full value from the supplier relationship, the contract must be able to be regularly and quickly amended to allow the business to benefit from service developments.

Benchmarking provides an assessment against the marketplace. The supplier may be committed by the contract to maintaining charges against a market rate. To maintain the same margin, the supplier is obliged to improve its operational efficiency in line with its competitors. Collectively, these methods help provide an assessment of an improving or deteriorating efficiency.

The point of responsibility within the organization for deciding to change a supplier relationship is likely to depend on the type of relationship. The service provider may have identified a need to change supplier, based on the existing supplier's performance, but for a contractual relationship the decision needs to be taken in conjunction with the organization's procurement and legal departments.

The organization should take careful steps to:

- Perform a thorough impact and Risk Analysis of a change of supplier on the organization and its business, especially during a period of transition. This could be particularly significant in the case of a strategic relationship.
- Make a commercial assessment of the exit costs. This may include contractual termination costs if supplier liability is not clear, but the largest costs are likely to be associated with a transition project. For any significant-sized relationship, this typically includes a period of dual-supply as services are migrated. Any change associated with a change in supplier will increase costs, either immediately as fixed costs, or over time where borne by the supplier and reflected back in service charges.
- Take legal advice on termination terms, applicable notice period and mechanisms, and any other consequences, particularly if the contract is to be terminated early.
- Reassess the market to identify potential benefits in changing supplier.

A prudent organization undertakes most of these steps at the time the original contract is established, to ensure the right provisions and clauses are included, but this review activity needs to be reassessed when a change of supplier is being considered.

4.7.6 Triggers, inputs, outputs and interfaces

There are many events that could trigger Supplier Management activity. These include:

- New or changed corporate governance guidelines
- New or changed business and IT strategies, policies or plans
- New or changed business needs or new or changed services
- New or changed requirements within agreements, such as SLRs, SLAs, OLAs or contracts
- Review and revision of designs and strategies
- Periodic activities such as reviewing, revising or reporting, including review and revision of Supplier Management policies, reports and plans
- Requests from other areas, particularly SLM and Security Management, for assistance with supplier issues
- Requirements for new contracts, contract renewal or contract termination
- Re-categorization of suppliers and/or contracts.

The key interfaces that Supplier Management has with other processes are:

- ITSCM: with regard to the management of continuity service suppliers.
- SLM: assistance with the determining of targets, requirements and responsibilities and their inclusion within underpinning agreements and contracts to ensure that they support all SLR and SLA targets. Also the investigation of SLA and SLR breaches caused by poor supplier performance.
- ISM: in the management of suppliers and their access to services and systems, and their responsibilities with regard to conformance to ISM policies and requirements.
- Financial Management: to provide adequate funds to finance Supplier Management requirements and contracts and to provide advice and guidance on purchase and procurement matters.
- Service Portfolio Management: to ensure that all supporting services and their details and relationships are accurately reflected within the Service Portfolio.

4.7.6.1 Inputs

- **Business information:** from the organization's business strategy, plans and financial plans, and information on their current and future requirements
- **Supplier and contracts strategy:** this covers the sourcing policy of the service provider and the types of suppliers and contracts used. It is produced by the Service Strategy processes
- **Supplier plans and strategies:** details of the business plans and strategies of suppliers, together with details of their technology developments and plans and statements and information on their current financial status and projected business viability
- **Supplier contracts, agreements and targets:** of both existing and new contracts and agreements from suppliers
- **Supplier and contract performance information:** of both existing and new contracts and suppliers
- **IT information:** from the IT strategy and plans and current budgets

- **Performance issues:** the Incident and Problem Management processes, with incidents and problems relating to poor contract or supplier performance
- **Financial information:** from Financial Management, the cost of supplier service(s) and service provision, the cost of contracts and the resultant business benefit and the financial plans and budgets, together with the costs associated with service and supplier failure
- **Service information:** from the SLM process, with details of the services from the Service Portfolio and the Service Catalogue, service level targets within SLAs and SLRs, and possibly from the monitoring of SLAs, service reviews and breaches of the SLAs. Also customer satisfaction data on service quality
- **CMS:** containing information on the relationships between the business, the services, the supporting services and the technology.

4.7.6.2 Outputs

The outputs of Supplier Management are used within all other parts of the process, by many other processes and by other parts of the organization. Often this information is supplied as electronic reports or displays on shared areas or as pages on intranet servers to ensure the most up-to-date information is always used. The information provided is as follows:

- **The Supplier and Contracts Database (SCD):** holds the information needed by all sub-processes within Supplier Management – for example, the data monitored and collected as part of Supplier Management. This is then invariably used as an input to all other parts of the Supplier Management process.
- **Supplier and contract performance information and reports:** used as input to supplier and contract review meetings to manage the quality of service provided by suppliers and partners. This should include information on shared risk where appropriate.
- **Supplier and contract review meeting minutes:** produced to record the minutes and actions of all review meetings with suppliers.
- **Supplier Service Improvement Plans (SIPs):** used to record all improvement actions and plans agreed between service providers and their suppliers, wherever they are needed, and should be used to manage the progress of agreed improvement actions, including risk reduction measures.
- **Supplier survey reports:** often many people within a service provider organization have dealings with suppliers. Feedback from these individuals should be collated to ensure that consistency in the quality of service provided by suppliers is provided in all areas. These can be published as league tables to encourage competition between suppliers.

4.7.7 Key Performance Indicators

Many KPIs and metrics can be used to assess the effectiveness and efficiency of the Supplier Management process and activities. These metrics need to be developed from the service, customer and business perspective, such as:

- Business protected from poor supplier performance or disruption:
  - Increase in the number of suppliers meeting the targets within the contract
  - Reduction in the number of breaches of contractual targets.
- Supporting services and their targets align with business needs and targets:
  - Increase in the number of service and contractual reviews held with suppliers
  - Increase in the number of supplier and contractual targets aligned with SLA and SLR targets.
- Availability of services is not compromised by supplier performance:
  - Reduction in the number of service breaches caused by suppliers
  - Reduction in the number of threatened service breaches caused by suppliers.
- Clear ownership and awareness of supplier and contractual issues:
  - Increase in the number of suppliers with nominated supplier managers
  - Increase in the number of contracts with nominated contract managers.

4.7.8 Information Management

All the information required by Supplier Management should be contained within the SCD. This should include all information relating to suppliers and contracts, as well as all the information relating to the operation of the supporting services provided by suppliers. Information relating to these supporting services should also be contained within the Service Portfolio, together with the relationships to all other services and components. This information should be integrated and maintained in alignment with all other IT information management systems, particularly the Service Portfolio and the CMS.

4.7.9 Challenges, Critical Success Factors and risks

Supplier Management faces many challenges, which could include some of the following:

- Continually changing business and IT needs and managing significant change in parallel with delivering existing service
- Working with an imposed non-ideal contract, a contract that has poor targets or terms and conditions, or poor or non-existent definition of service or supplier performance targets
- Legacy issues, especially with services recently outsourced
- Insufficient expertise retained within the organization
- Being tied into long-term contracts, with no possibility of improvement, which have punitive penalty charges for early exit
- Situations where the supplier depends on the organization in fulfilling the service delivery (e.g. for a data feed) can lead to issues over accountability for poor service performance
- Disputes over charges
- Interference by either party in the running of the other's operation
- Being caught in a daily fire-fighting mode, losing the proactive approach
- Communication – not interacting often enough or quick enough or focusing on the right issues
- Personality conflicts
- One party using the contract to the detriment of the other party, resulting in win-lose changes rather than joint win-win changes
- Losing the strategic perspective, focusing on operational issues, causing a lack of focus on strategic relationship objectives and issues.

Key elements that can help to avoid the above issues are:

- Clearly written, well-defined and well-managed contract
- Mutually beneficial relationship
- Clearly defined (and communicated) roles and responsibilities on both sides
- Good interfaces and communications between the parties
- Well-defined Service Management processes on both sides
- Selecting suppliers who have achieved certification against internationally recognized certifications, such as ISO 9001, ISO/IEC 20000, etc.

The main CSFs for the Supplier Management process are:

- Business protected from poor supplier performance or disruption
- Supporting services and their targets align with business needs and targets
- Availability of services is not compromised by supplier performance
- Clear ownership and awareness of supplier and contractual issues.

The major areas of risk associated with Supplier Management include:

- Lack of commitment from the business and senior management to the Supplier Management process and procedures
- Lack of appropriate information on future business and IT policies, plans and strategies
- Lack of resources and/or budget for the ISM process
- Legacy of badly written and agreed contracts that don't underpin or support business needs or SLA and SLR targets
- Suppliers agree to targets and service levels within contracts that are impossible to meet, or suppliers fail or are incapable of meeting the terms and conditions of the contract
- Supplier personnel or organizational culture are not aligned to that of the service provider or the business
- Suppliers are not cooperative and are not willing to partake in and support the required Supplier Management process
- Suppliers are taken over and relationships, personnel and contracts are changed
- The demands of corporate supplier and contract procedures are excessive and bureaucratic
- Poor corporate financial processes, such as procurement and purchasing, do not support good Supplier Management.