

Information and knowledge systems



Information and knowledge systems

Jihočeská univerzita v Českých Budějovicích

Branišovská 1645/31a

370 05 České Budějovice

IČ: 60076658

CONTENT

Obsah

[1 Introduction 2](#_Toc2536152)

[2 Information and decision making 3](#_Toc2536153)

[2.1 Data 3](#_Toc2536154)

[2.2 Information 3](#_Toc2536155)

[2.3 Knowledge 4](#_Toc2536156)

[2.4 Building knowledge – learning 5](#_Toc2536157)

[2.5 Learning – from the individual to the organisation 6](#_Toc2536158)

[2.6 Forms of information 6](#_Toc2536159)

[2.7 Information as an aid to decision making 8](#_Toc2536160)

[2.8 Levels of management decisions 10](#_Toc2536161)

[2.9 Getting the right information to make decisions 11](#_Toc2536162)

[2.10 Formal information systems for management support 12](#_Toc2536163)

[2.11 Management information systems (MIS) 12](#_Toc2536164)

[2.12 Decision support systems (DSS) 13](#_Toc2536165)

[2.13 Executive support systems (ESS) 14](#_Toc2536166)

[2.14 Other systems 15](#_Toc2536167)

[2.15 Using the Web as an information resource 15](#_Toc2536168)

[3 Evaluating information 20](#_Toc2536169)

[3.1 Information overload 20](#_Toc2536170)

[4 Communicating information 25](#_Toc2536171)

[5 Information systems 28](#_Toc2536172)

[6 Knowledge management 38](#_Toc2536173)

[7 Semantic Web 50](#_Toc2536174)

[8 References 68](#_Toc2536175)

# Introduction

Information management underpins the key activities of planning, analysis, action and, above all, learning and development. How to make information useful

Organizations need to manage information well and consistently in order to be responsive to the needs of their customers. This book approaches information management from two key perspectives: How you as a manager use and manage information. The information management process and how it impacts on decision making and organizational performance. It looks at information in five themes, starting with the sourcing of information and culminating in an exploration of the ways in which organizations manage information and knowledge:

* Finding information to meet your needs – finding good sources of information
* Managing your incoming information – reducing the overload
* Managing your outgoing information – the way you communicate information
* How organizations manage information and knowledge – the systems
* How organizations manage information and knowledge – the content

Objectives of knowledge management are to:

* Identify sources of information relevant to your needs inside and outside of your organization
* Evaluate and improve the quality of your information sources
* Learn how to manage information overload
* Describe key principles for communicating effectively in writing
* Identify the principles behind information system design and management
* Explain the features of knowledge management.

# Information and decision making

People need information to plan their work, meet their deadlines and achieve their goals. They need it to analyse problems and make decisions. Information is certainly not in short supply these days, but not all of it is useful or reliable. This first theme explores your needs for information and asks you to consider how they are served by the sources of information that are available to you.

In this chapter, you will:

Consider the differences between data, information and knowledge;

Identify and evaluate the sources of information that you use;

Assess whether information flows effectively within your team and identify areas for improvement;

Analyse how effectively you use the Internet as an information source.

From data to information to knowledge and learning

H D Clifton (1990) wrote that ‘one man’s information is another man’s data’, and certainly the definitions are blurred. However, it is now generally agreed that ‘data’ is pure and unprocessed – facts and figures without any added interpretation or analysis. Depending on the context, data can be highly significant. Think of a cricket or football score, your name and address. Since it provides the raw material to build information, it also has to be accurate. Any inaccuracies within the initial raw data will magnify as they aggregate upwards, and will seriously corrupt the validity of any conclusions you draw from it or decisions you base upon it.

## Data

In a business context, data is associated with the operational aspects of the business and its day-to-day running. As such, it is often entered into a system and stored in large quantities, for example payroll data and sales figures. Such input data goes to create a data ‘set’ – names and addresses for a mail-merge file, an index to an online product database. It has to be structured correctly – all systems have some kind of validation process to check for obvious technical errors and missing data. To be reliable, the content needs to be accurate, not simply in terms of the correct number and type of characters per data field, but what the data actually represents in terms of meaning. This needs human intervention. Another aspect that affects accuracy is where the data comes from. You may be able to check your own in-house sources – for example, for internally generated data such as the payroll – but have to depend on trust (or the reputation of the supplier) for data received from outside, for example customer credit card details.

## Information

So how does ‘data’ (whether internal or external) become ‘information’? When it is applied to some purpose and is adding value which has meaning for the recipient, for example taking sets of sales figures (data) and producing a sales report on them (information).

Data

Information

**applied for**

**a purpose**

Figure 1.1 From data to information

Of course, the same set of data can be used to produce different kinds of information, depending on how it is applied and who applies it. The same sales figures that you use to produce a market sector report might be used by someone else to justify adding to or reducing the size of the sales team. Such information can be used to manage a department, and for short and medium-term planning. Data can move to information and be turned to practical advantage very quickly – in 1815 the London Stock Market rapidly took advantage of the news brought by carrier pigeon of Wellington’s victory at Waterloo, which arrived two days before the human messenger arrived.

Information produced inside the organisation can be supplemented by a wealth of business information produced outside – market analyses, reports and case studies, for example.

Put briefly, information by itself is only of use if it is:

The right information (fit for the purpose);

At the right time;

In the right format at the right price.

## Knowledge

Just as the words ‘data’ and ‘information’ are used interchangeably, there is considerable blurring and confusion between the terms ‘information’ and ‘knowledge’. It is helpful to think of knowledge as being of two types: the instinctive, subconscious, tacit or hidden knowledge, and the more formal, explicit or publicly available knowledge. An everyday example of these might be the knowledge that you use when driving a car (tacit), compared with the knowledge available from a driving manual or the Highway Code (explicit).

Theme 5 looks at knowledge in more detail and how it can be managed within organisations.



Figure 1.2 From data to information to knowledge

In a business context, knowledge is often linked to strategic levels of management and long-term business planning, where it is associated with having a head for business or business flair. However, knowledge vital to an organisation’s success can come from any level within it, and needs to be recognised as an important part of organisational assets. It combines information, experience and insight into a mix that is unique to every employee. It is this mix of understandings, based on personal knowledge at a tacit level, that creates the strengths and at times the vulnerability of organisations. It is important for organisations to recognise that holding knowledge at the tacit or hidden level can only have value where people are isolated from everyone else in their decision making. This is neither realistic nor good business practice.

Let’s sum up data–information–knowledge with an everyday example. Assume that you’re trying to decide on a specialist holiday for photography enthusiasts. Here, very broadly, are the stages you will go through:

Stage 1: collect lots of brochures on photography holidays. This is your basic data store.

Stage 2: work through the brochures, filtering out what you don’t want by applying your own criteria to them. Some will be in places you don’t want to go to, or at the wrong time of year, or the programmes may be at the wrong level of expertise (you may be looking for some advanced tuition, and many of the holidays are geared to beginners). You can now apply your information and make a decision on where to go on your holiday.

Stage 3: you go on your holiday and build your knowledge from testing your actual experience of the holiday against the information you had when you booked it. This knowledge (which you can use next time you want a similar holiday) can be kept to yourself (tacit) or you can share it by reporting back to your local photography club (explicit).

Capitalising on knowledge by making the tacit explicit, and identifying and managing the processes that nurture it, is a thread that runs through this text.

## Building knowledge – learning

So how do we collect, process and build our knowledge? Kolb (1985) believes that there are four stages we all go through as part of the learning cycle:

Learning from feeling (through specific experience and relations with other people);

Learning by watching and listening (looking at things from different perspectives, observing carefully and reflecting before making judgements);

Learning by thinking (reflecting on and analysing ideas, drawing up mental maps and planning);

Learning by doing (getting things done, influencing other people, taking risks).

**CONCRETE EXPERIENCE**

(Learning from feeling)

**ABSTRACT CONCEPTUALISATION**

(Learning by thinking)

**REFLECTIVE**

**OBSERVATION**

(Learning by

watching

and listening)

**ACTIVE**

**EXPERIMENTATION**

(Learning by doing)

Figure 1.3 Kolb’s learning cycle Source: Kolb (1985)

We all go through each of these processes to an extent, but different people feel more comfortable with some than with others. For example, an action-oriented person who likes to learn by doing may get very frustrated in a learning-by-watching situation or in one that requires reflection and analysis. It is useful for managers to be aware of their own and their staff’s learning styles, since these provide valuable insights into making most effective use of different methods of training.

Argyris and Schön (1974) argue that people act in accordance with a set of mental maps that they themselves have created. It is these subconscious maps (or private, self-generated theories) that guide people’s actions. They called these theories that are implicit in what we do theories-in-use: these are what govern our actual behaviour. The words we use to describe that behaviour to others – how we like to justify our actions to other people, or what we would like them to think – can be quite different. This is called espoused theory. It may sound cynical, but if someone asks you how you would behave in a particular set of circumstances, the answer you will give will almost certainly be espoused theory: the public rather than the private set of principles.

Argyris and Schön’s view is that real effectiveness results from developing congruence between theory-in-use and espoused theory: creating harmony between your inner and outer self.

## Learning – from the individual to the organisation

People learn by seeking out information when faced with a new situation, and using this information to draw conclusions and form mental models which they use as the basis for their action. If these mental models are confirmed and reinforced by our experience in reality, then over time they become so familiar that they become routine, used automatically and with no conscious effort.

This applies to the presenter who always opens up proceedings with a joke. It also applies to the air traffic controller at an international airport, but in this case we expect the knowledge to be embedded and made explicit through a series of rules and procedures that are recognised and shared by everyone else.

Organisations use routines, rules and procedures as a way of sharing knowledge and creating standardised processes throughout the organisation. These are the systems we use to do our work. Such systems existed before the desktop computer, but computerisation has led to sophisticated information technology (IT) systems for accessing, inputting, processing and sharing information that can be used widely and quickly across the organisation.

The problem for organisations is that routines become old learning and so embedded into our systems that they stifle creativity and the flexibility to respond to changing circumstances. This flexibility – the ability to change and learn – is essential to organisations if they are to survive and grow. The way organisations seek to encourage learning and the sharing of information and knowledge are important aspects of information management.

## Forms of information

Forms of information include the following:

Information need not be written down or be verbalised to be valuable

Internal and external – information generated inside the organisation and information generated outside. External intelligence and research may be incorporated into internal reports, and issues arising from internal reports may stimulate external market research.

Electronic and hard copy (paper-based), and spoken. At Sun Microsystems, employees receive, on average, 100 e-mails each day, but few people work in a paperless office. Most people also use conversation with others for information.

Hard and soft – or quantitative and qualitative. Hard information is often derived from large quantities of precise factual data, such as figures, that lends themselves to statistical analysis. Soft information, on the other hand, tends to come from few sources and depends on opinions, feelings, impressions and judgements.

Formal and informal. This is worth exploring in more depth.

**Formal and informal**

Some of the formal information sources you might use every day include:

newspapers or electronic newsfeeds;

magazine articles;

management reports;

staff disciplinary procedures;

videos of product presentations layouts, maps, blueprints.

You will also use a number of informal information sources – so informal that you might not even recognise them as such! They can include:

a chat with the managing director’s personal assistant whilst queuing for lunch

checking out a problem with a colleague

meeting up with colleagues from the same trade or professional association at the annual conference

informal contacts with suppliers and customers.

Some of the most useful of these sources will be information gatekeepers – people who routinely collect, evaluate and disseminate information in an informal way which may have nothing to do with their job role. These people are well aware of the way information flows around their local environment, and can exercise an influence that goes well beyond their notional status within the organisation.

If you think about it, information need not even be written down or verbalised to be valuable. You can learn a lot about an organisation and its culture simply by walking about and keeping your eyes open, observing the way the organisation goes about its business and presents itself to staff and the outside world.

There are some key differences in the characteristics of formal and informal information sources, as shown in Table 1.1.

Table 1.1 Characteristics of formal and informal information sources

|  |  |
| --- | --- |
| *Formal* | *Informal* |
| Available to more than one person | May be an interchange between just two people |
| Information captured has been recorded in some way, | The information is transient – not stored or retrievable |
| so can be reused  The information used is selected by the recipient – | The information is selected by the provider |
| for example, you decide which newspaper reports you are going to read  Information tends to be static | Information is interactive |
| Information is likely to conform to the organisation’s | Information is more likely to be ‘private’ and although |
| promoted self-image – it is likely to be ‘espoused theory’ | partial, is likely to be closer to theory-in-use than formal information sources |

There are several reasons why managers prefer informal to formal methods of information transfer:

The response and feedback is instant. The whole process is quicker and so is perceived as more efficient (even if the information is only patchy or actually inaccurate).

Being personal, it is targeted at the recipient, so some initial filtering will have been carried out (but is this the half of the picture you want and need...?).

They might not know what useful formal information is available, or how to access it.

Cultural reasons: decisions are often made on the basis of experience and judgement, not painstaking fact finding.

In practice, it makes sense to use a mix of formal and informal, hard and soft data to get a complete picture.

Table 1.2 shows some typical information needs and the information sources that might meet them.

Table 1.2 Examples of information needs and sources

|  |  |  |
| --- | --- | --- |
| *Need/purpose* | *Types of information* | |
| Produce a report on | Who asked for the report and who will read it | |
| ice-cream sales for June | Projected and actual sales figures  Previous year’s figures  Meteorological data  Report of June launch of new ice-cream product by major competitor | |
| Your awareness of your | Company reports and budgets | Discussions at the coffee machine |
| own organisational | Products and services launched or axed | Share price |
| environment | Internal newsletters and memos | Competitor share price |
| (keeping your finger on | Meetings |  |
| the pulse)  Competitor intelligence | Press reports on company performance | Trade journals |
|  | and activities | News reports |
|  | Market research data/market analysis | Share price |
|  | Company websites | Trends analysis and forecasting |
|  | Company annual reports | Industry gossip |

## Information as an aid to decision making

Much decision making is based on our inbuilt mental models and knowledge base, but this tacit information source can be corroborated and enhanced by formal decision-support mechanisms.

**The decision-making process**

How do you make decisions? Do you assemble all the facts relating to the problem? Rely on your experience and insight? Shut your eyes and hope for the best? Most people do some or all of these things at different times, depending on the nature of the decision. However, the decision-making process shown in Figure 1.4 describes the basic steps involved in consciously making a decision.

**Define the problem**

**Analyse it and evaluate options**

**Make the decision**

**Implement the decision**

**Review the decision**

Figure 1.4 The decision-making process

The key step is the second one: analyse and evaluate options. Whatever the problem – sorting out a production schedule, conducting a staff appraisal, negotiating a deal – you will need either consciously or unconsciously to weigh up the situation and make decisions accordingly.

An increasing difficulty facing managers now is the speed at which these decisions have to be made: there is just no time for detailed investigation. In an age when managers are faced with more and more information, there is less and less time in which to evaluate its usefulness. As a result, decisions are made on the basis of partial information, wrong information – or whatever information is available, rather than appropriate.

The concept of cause and effect is commonly used in the way people argue and reason. In making our choices, it is important to identify the right causes and effects – it is all too easy to focus on the symptoms rather than the root causes. It is also necessary to consider your decision-making criteria – what you want to achieve, within what time frame, with what resources. This does assume, of course, that there is a single ‘right’ decision that you can make to achieve a predictable, successful outcome.

Informix, a software development company, carried out a survey in 1999 to examine how decisions are made in different organisations around the world, and to find out how well the available information, in all its forms, supported the decision-making process. A general finding was that managers, even when they are supported by a multitude of different information sources, find decision making extremely stressful. Most of these managers quoted examples of major decisions that were made incorrectly in the previous six months, and the larger the organisation, the more likely it was to have had a problem.

One of the most important detrimental factors affecting decision making was limited, incorrect or misinterpreted data.

Some key findings of the survey:

32 per cent of the sample had made an important business decision in the past six months based on hope or luck

the single biggest cause of stress in decision making is a lack of information

33 per cent of managers ignore relevant data either when making a decision in the first place or when it becomes apparent that a decision has been incorrect.

Source: Informix (2014)

What happens when it goes wrong? Below are some examples of information disasters, where the information needed to make decisions was unavailable or ignored:

On 19 October 1987, the Dow Jones Industrial Average took its biggest one-day plunge in the history of the US Stock Market. A major factor in this was that information systems malfunctioned and impeded information flows.

In the same month, British meteorologists failed to appreciate the strength of the oncoming winds which led to one of the biggest storms in living memory: they ignored the available information.

Nuclear scientists at Three Mile Island, and later Chernobyl, failed to take account quickly enough of the information coming from their instrumentation to prevent accidents happening.

Catastrophes of all manner can and do ensue because of what the behavioural scientist might call ‘dysfunctional information attitudes and behaviours’. This is a fancy phrase that means that information has been mismanaged somehow, somewhere, by someone, at some time, and often with disastrous consequences in terms of human misery, political misfortune or business failure.

Source: Horton and Lewis (1998)

The kinds of decision making in which managers are involved:

## Levels of management decisions

Management decisions are made at three broad levels within the organisation, and each type of decision has its own characteristics:

Operational decisions: these are the day-to-day decisions affecting the running of the organisation. The decisions tend to be short term (days or weeks) and need to be made quite frequently. For example, a supermarket deciding on whether it needs to order more strawberries to cope with current demand.

Tactical decisions: these have a longer time frame (months or years) and tend to be made by middle managers who are directly involved in implementing the policies of the organisation. For example, a toy shop timing the start of its Christmas promotion.

Strategic decisions: these are made by top management, and since they affect the organisational plans of the whole business, possibly for a number of years, they are not made very frequently. For example, whether to sell off a subsidiary company in response to falling profits.

All these decisions will require information, but the type of information that is needed will be different for each level of decision making. See Figure 1.5.

**Mostly internal information; detailed**

**Internal and external information; less detail, more summary**

**External information in summary form**

**Strategic**

**Tactical**

**Operational**

Figure 1.5 Characteristics of information for management decisions Source: Nickerson (2001)

Operational decisions rely mostly on internal, detailed data: how many strawberries did we sell yesterday, or last weekend? Tactical decisions involve a wider spread of less detailed information: for the past two years, what were the sales figures for the month prior to the Christmas promotion and during the promotion? Strategic decisions may use long-term performance figures from inside the organisation, but also financial forecasts and analyses from the wider marketplace, its own shareholders’ views, and so on.

Cross-functional, in-house systems such as accounting, finance, marketing and human resources (HR) can, of course, support decision making at every level of the organisation, whether operational, tactical or strategic. There are also general types of information systems for management support, which we will look at a little later in this section.

## Getting the right information to make decisions

In an ideal world, getting the right information to make decisions would be very easy. We would just type a question into our PC, or know exactly the right person we need to telephone in order to get an instant, accurate and authoritative answer.

In real life, most of us have to get by without perfect one-stop solutions. Where do we get the information we need to make decisions when our systems are not organisation wide, but are locked into ‘silos’ where we can perhaps drill down to increasing levels of detail, but not across to the vital missing piece of data that is held within another department?

If you think of a decision you have made recently and about where the information came from, you will probably realise that it is a mix of your own knowledge, whatever information was available and maybe a chat to a couple of colleagues who always seem to have an answer or know where to find one. Think a little more broadly: how does your team get the information it needs in order to operate? The model would probably look something like the one shown in Figure 1.6.

**INPUTS**

Manager's instruction

Procedures

Technical manuals

Opinions, ideas, knowledge

**OUTPUTS**

Product or

service

Solution to

problem

Team

knowledge

and

experience

**+**

**=**

Figure 1.6 Information for team operations

In this model, the inputs (your basic raw materials) come from a diverse mix of sources. Some of these will be formal, some very informal – but no less valuable for that. These inputs will be processed by individuals or the team to produce the desired outputs (a specific product or service, or the solution to a problem). Getting it right assumes that the flow of information, both formal and informal, is:

unimpeded – there are no bottlenecks and blockages (human or technical);

able to move upwards, downwards and sideways with equal ease equally accessible to all who need it.

An organisational approach to take some of the luck out of getting the right information for decision making – for making individual knowledge explicit and sharing it across the organisation – is to develop formal information systems to support managers.

## Formal information systems for management support

Computer systems that can store and manipulate information provide a structured and accessible support for management decision making. Here are descriptions of three kinds of systems in common use: management information systems (MIS), decision support systems (DSS) and executive support systems (ESS).

## Management information systems (MIS)

A management information system, or MIS, supports management decisions by providing information in the form of reports and responses to queries to managers at different levels within an organisation. The MIS database that provides the information to the manager comes from both inside and outside the organisation, much of it from the data stored in transaction processing systems – the nuts and bolts of day-to-day operations and processes.

**MIS**

**TPS**

**External**

**data**

**TPS**

**software**

**TPS**

**stored data**

**Users and other Transaction**

**Processing Systems (TPSs)**

**MIS**

**database**

MIS

**software**

**Strategic (top level) managers**

**Tactical (middle level) managers**

-

**Operational (lower-level) managers**

Figure 1.7 Structure of an MIS Source: Nickerson (2001)

## Decision support systems (DSS)

Whereas an MIS provides information from a database with little or no analysis, a decision support system (DSS) helps managers by analysing data from a database and providing them with the results, often in the form of statistical calculations or mathematical models. It is used most often for decisions at tactical and strategic levels. The main system components are the DSS database that contains the data, and the model base which contains the mathematical models and statistical calculation routines that are used to analyse data from the database. Decision support systems are often used in situations where decisions are unstructured or semi-structured, and are good for working through ‘what if’ scenarios to calculate the effects of different decisions on outcomes (what happens if we start the Christmas promotion two weeks earlier?)

**DSS model base**

**DSS database**

**DSS software**

**Users (decision makers)**

Figure 1.8 Structure of a DSS Source: Nickerson (2001)

A variation on this is a group DSS, typically used in a networked environment where several PCs are joined together, in which users can collaborate to reach a group decision.

## Executive support systems (ESS)

Also known as executive information systems, these are designed to support strategic business decisions. Although strategic decisions usually involve summarised information, there is often a need for a specific level of detail to pinpoint a particular problem. For example, executives in an organisation that is thinking of selling off a failing subsidiary might want to try to discover where its failure lies: is it a particular market segment, a region, a product line? This will often require a drilling-down process to get from general information to highly specific data subsets.

The user of an ESS will typically need to access a wide variety of databases: internal, external, those created by the individual user and electronic mailboxes.

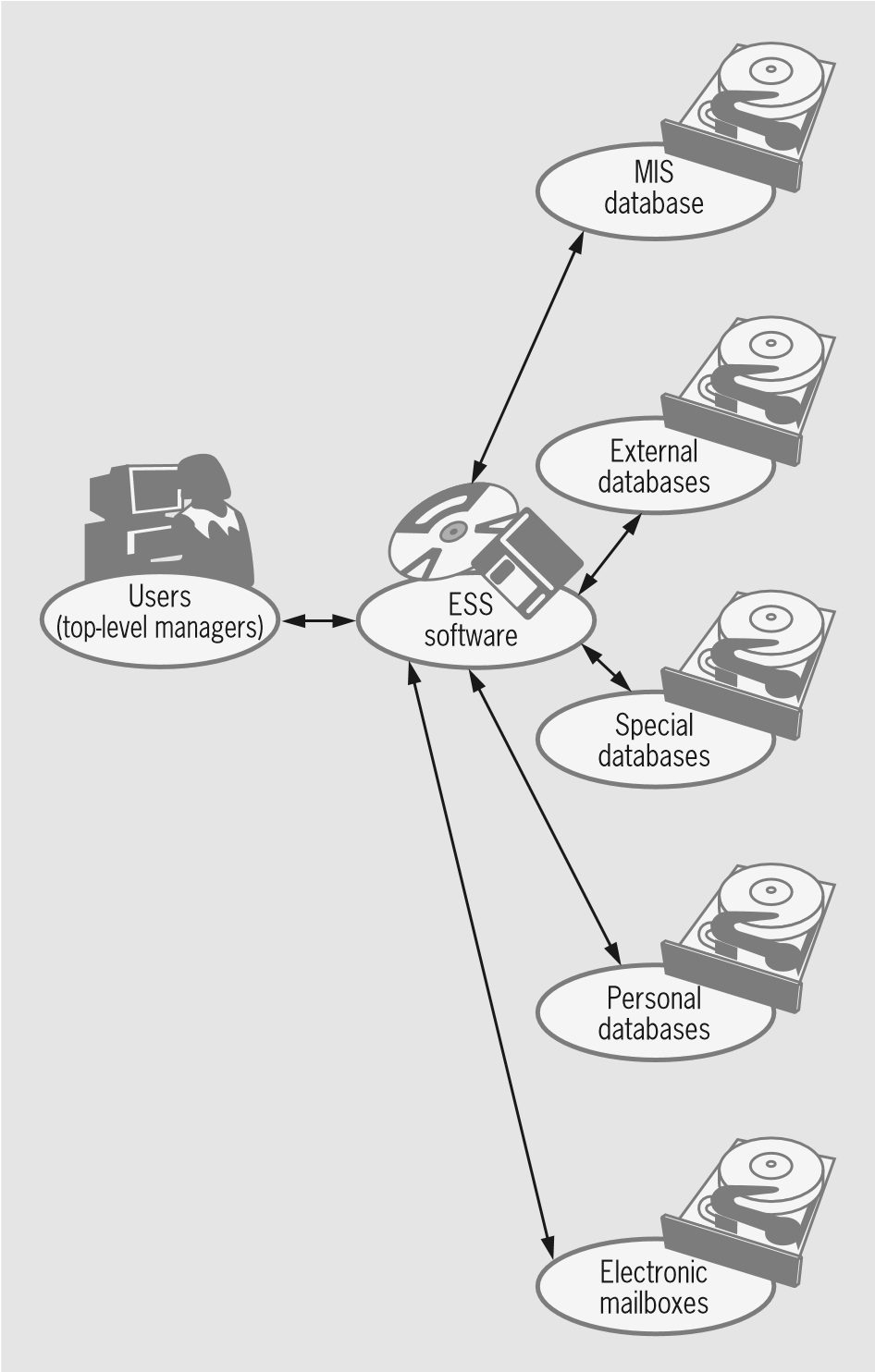


Figure 1.9 Structure of an ESS Source: Nickerson (2001)

## Other systems

Increasingly, managers are looking at more sophisticated methods such as expert systems that mimic the way a human would analyse a situation and recommend a particular course of action, and knowledge management systems that can organise, store and enable shared access to the collective knowledge of the organisation.

The point to remember about all of these systems is that the quality of the output is only as good as the quality of the input.

The activities that follow explore the theme of information flows and systems for management support. You will start by looking at your information sources.

## Using the Web as an information resource

Here we provide some very practical guidance on finding the information you need, both to make informed decisions and to build up your own knowledge base in your chosen area. It will not eliminate all the problems involved in finding Web-based material, but should give you some sharper tools to help you along the way.

Search engines

Note that although the World Wide Web is technically only a part of the Internet, it is the one that is most familiar to most people, and the terms ‘Web’ and ‘Internet’ are used interchangeably.

The starting point for all Web searches is a search engine – quite literally, a force that responds to an information request by searching the Web for what it interprets as relevant material. Search engines are also referred to here as indexes as they act like gigantic indexes to selected chunks of the Web. They take an input search word (search term) or phrase, and retrieve a set of results (hits) that relate to that term or phrase from the Web pages that they have identified, collected into a virtual database and indexed. Note the word ‘selected’ – none of them scans absolutely everything, and you will need to learn which search engines are most useful for which purposes.

There are four basic types of search tool:

free text search engines

human-generated indexes

metasearch tools

natural language tools.

As search engines develop, the distinction between the types is becoming more blurred.

**Free text search engines**

Search engines retrieve a set of Web pages (hits) that match a word or phrase input by the user. They do not search the entire Web – only those pages that exist in the index of the search engine. The indexes are compiled by computer robots and can be vast. Google (www.google.com) and Alta Vista (www.altavista.com) are currently the biggest with billions of pages each. Since the indexing method is basically a free text search, the engine will retrieve every instance of the search term, whether it is relevant to your search or not. This means that if you’re a bird enthusiast looking for information on ‘cranes’, you will also retrieve references to heavy lifting gear, maybe crane flies and companies that have crane in their title. On the other hand, these searches may not pick up useful related terms, so a search on ‘boats’ may not select references to ‘yachts’ or ‘ships’. This is an area that is improving all the time.

**Index based search engines**

Some companies also try to catalogue the Web. Whereas search engines use computers to create the search engine index, classified and specialist directories use humans to select and catalogue the Web pages. Yahoo (www.yahoo.com) is one of the most notable. As well as being able to enter search text, the user can also browse through the directory. For example, if you want to find a new movie, you might start with entertainment and then click movies and carry on until you find what you want.

There are numerous specialist directories that act as gateways to specific subjects on the Web. The medical gateway www.omni.ac.uk is an example. For a comprehensive list of what is available, go to www.vlib.org.

Metasearch engines

These are not search engines themselves – more tools that know about other search tools and will submit your query to several search engines at once. Metacrawler (www.metacrawler.com) and Dogpile (www.dogpile.com) are examples.

**Natural language search engines**

Natural language search engines are very appealing, as you can literally type in a question in the way that you would ask it. Ask

Jeeves is probably the best known of these. Inputting: ‘Who won the World Cup in 1998?’ retrieves not only the result but details of many other World Cup and football-related sites.

Table 1.3 summarises the main types of search engines and what they are most useful for.

Table 1.3 Main types of search engines

|  |  |  |
| --- | --- | --- |
| *Type of search engine* | *Example* | *What it’s most useful for* |
| Free text search engines | Google | When you know exactly what you want and can be specific about it.  Good for ‘Mercedes-Benz’; bad for ‘performance cars’ |
| Index-based search engines | Yahoo | An overview of the subject area, structured so that you can narrow down a search or make it broader. For example, from  ‘astrophotography’ you can go up to the broader category ‘astronomy’ or down to the more specific ‘lunar eclipse photography’ |
| Metasearch engines | Dogpile | A broad and comprehensive view of sites in a subject area |
| Natural language search | Ask Jeeves | Good for novice searchers,or if you want a general look around a subject area |
| Specialist indexes | Omni | In-depth access to a highly specific subject area |

A list of selected search engines is provided at the end of this section.

**Getting better results**

Choosing the right kind of search engine for your purpose will go a long way towards getting better search results more quickly. There are a couple of other things you can do too.

**Advanced search**

One of these is to take advantage of any advanced search facilities offered by the search engines. These should make your search more specific, and more likely to retrieve focused results rather than irrelevant hits (false drops).

For some search engines, such as Excite, HotBot and Lycos, this kind of search supports the use of Boolean operators. These sound alarming but are really quite simple, and consist of just three words which you can incorporate into your search phrase: AND, OR and NOT.

You may have noticed AND appearing in the search header when you are running a search. What it means is that only items that have all elements of the search phrase in them will be retrieved. For example, ‘bottling AND canning’ will only produce results where both the terms ‘bottling’ and ‘canning’ appear in the same item.

OR will find all occurrences of the terms in your search phrase whether they are together or not. So ‘bottling OR canning’ will retrieve all items containing ‘bottling’, all items containing ‘canning’ and all items containing both terms. You will see from this that the effect of OR is to broaden your search, leading to a greater number of hits.

NOT is used to narrow a search. In our example, ‘bottling NOT canning’ will retrieve items which relate to ‘bottling’ but will exclude those which contain a reference to ‘canning’.

In search engines such as Google and Yahoo, the Boolean operators have been replaced by signs such as ‘+’ and ‘–’ for words to include or exclude. Some searches can also be restricted by date or other criteria such as language, or expanded by the use of ‘wildcard’ characters. Search results will also be affected by whether you input a phrase in quotes or not (try ‘London Bridge’ and London Bridge). Check out any help notes and search tips offered by a search engine from an advanced search facility to make the most of your time online. They do vary – compare the advanced search facilities offered by Google or Yahoo and Alta Vista.

Don’t forget you can also search for images, audio and video – Google claims to offer the most comprehensive resource of images, with 250 million of them.

**Bookmarking**

Finding the information once is one thing; finding it a second time is another. Make full use of your bookmarking facility to set up topic folders that you can refer to later and update periodically.

Some tips for effective use of the Internet

Be prepared to put in some time at first to surf around and get a good idea of the main websites in your area of interest

Try out different search engines to see how the results compare

Set up folders of bookmarked sites to create your own virtual library – but check the URLs (uniform resource locator, the address of a Web page) from time to time to see if they’re still valid

Take advantage of any online help or advanced search tips offered by the search engines.

You will also find useful advice on Internet searching on the sites of experts such as Phil Bradley (www.philb.com) and Karen Blakeman (www.rba.co.uk) or check out Manchester Metropolitan University at www.hlss.mmu.ac.uk/dic/main/howto.html

**Keeping up to date**

Another important aspect of quality information for decision making is that the information needs to be up to date.

One way of doing this is by joining mailing lists and newsgroups in your area of interest. A good place to start is http://groups.google.com which offers a list of groups to browse in all subject areas.

For the latest news and current affairs, there is a huge selection on offer. All the major search engines will have a news service of some kind (try looking at Yahoo or Northern Light) though some will have a strong US bias. As you might expect, the BBC is an excellent source of business news (www.bbc.co.uk/1/hi/business/default.stm) and the Financial Times is a popular site (www.ft.com). A more personalised offering is Create Your Own Newspaper or CRAYON, which allows you to personalise your own information sources and subject areas (for details see www.crayon.net/using/how.html).

The last few years have seen the rise of robots (bots) or intelligent agents, which can ‘learn’ your requirements and scurry around the Web looking for information on your behalf. They are more effective than average search tools for two reasons (Edmunds and Morris, 2000):

An intelligent agent can make decisions on the basis of the data it acquires without needing direct instruction from the user

Because it is able to learn about individual preferences, it can predict the likelihood of items it comes across being of interest.

For information on these clever little creatures, go to BotSpot

(www.botspot.com) which describes hundreds of them (start with What’s a Bot? or the FAQs section). The site includes a list of bots by category. These include several Update bots which can monitor your favourite websites and report on new developments.

Selected search engines:

* Google (www.google.com)
* Alta Vista (www.altavista.com)
* Ask Jeeves (www.ask.com)
* Dogpile (www.dogpile.com)
* Excite (www.excite.com)
* HotBot (www.hotbot.com)
* Lycos (www.lycos.com)
* Metacrawler (www.metacrawler.com)
* Northern Light (www.northernlight.com)
* Yahoo! (www.yahoo.com)

If you’re looking for non-English language search engines, try Searchengine Colossus (www.searchenginecolossus.com) which covers about 100 countries.

Consider the differences between data, information and knowledge

Information may be:

– electronic, hard copy or verbal

– quantitative or qualitative.

Identify and evaluate the sources of information that you use Information sources may be:

– Internal or external

– Formal or informal.

Managers often prefer informal sources because the data is instantly available, but using a mix of hard and soft information from formal and informal sources is a more reliable foundation for decision making. As a manager, you need to cultivate a range of information sources to meet your needs.

The information you need depends to a large extent on the types of decisions you make. Operational decisions can be made on the basis of internal data whereas strategic decisions require data and analyses from external sources.

Assess whether information flows effectively within your team and identify areas for improvement

Analysing the flow of information into your team may reveal blockages or bottlenecks, or that some people cannot access the information they need. This provides a basis for improving operational processes and decision making.

Computer systems are used to improve the flow of information. The different levels of decision making are supported by three types of system: management information systems, decision support systems and executive support systems.

Analyse how effectively you use the Internet as an information source

The four main types of search tool are: free text search engines, human-generated indexes, metasearch tools and natural language search enagines.

Choosing the right search engine and developing your knowledge of advanced search techniques will increase your efficiency and improve your search results.

# Evaluating information

A constant flow of information is essential for all of us to run our lives and businesses, but what happens when we get too much of it? The concept of information overload isn’t new – indeed it goes back centuries – but the huge expansion in publishing in the 19th century, and radio and TV in the early part of the 20th century, was already causing noticeable problems by the late 1950s. Now, with even greater growth of electronic information, individuals and organisations need to devise a whole new set of strategies to deal with the massive volumes of information that we all encounter day after day.

This will involve the ability to evaluate our information sources so that we can instantly discard what is irrelevant. It will also require us to consider more carefully the information we give out to other people – and the number of people we give it to. Refining the concept of fitness for purpose – the right information at the right time – and making this available across a shared-access system can help to improve the way in which we manage the growing volume of information.

In this chapter we will:

* Identify information overload and assess why it occurs
* Evaluate the information you receive by assessing its quality and value to you
* Reduce your information overload.

## Information overload

Are you subjected to ‘infoglut’ or ‘data smog’? Is it giving you ‘analysis paralysis’? The literature on information overload has invented some striking terms for a condition that many people claim to suffer from.

But what is information overload? Here is one definition:

Information overload occurs when information received becomes a hindrance rather than a help, when the information is potentially useful...

Source: Bawden et al. (1999)

Common feelings associated with information overload are that you feel overwhelmed, in a situation that’s out of your control and faced with more information than anyone can possibly handle. You don’t know which of the many pieces of information are the important ones, or indeed whether there are more pieces of information still to come before you can make an intelligent decision. The resulting feelings of helplessness can have a major impact on the health of employees and their efficiency and effectiveness within organisations.

A survey carried out by Reuters in 1996 revealed that:

* two-thirds of managers surveyed believed that information overload not only caused a loss of job satisfaction but also affected their personal relationships
* half thought it had damaged their health
* nearly half believed that important decisions were delayed and affected by having too much information.

Source: Reuters (1996)

How does it happen?

There are various reasons why managers collect more information than they can handle:

There’s a lot more information around to collect, and it’s increasing all the time

A general increase in unsolicited information (think of all the junk mail you get)

The fear of missing out on some vital piece of information that your colleagues may already know about

The perception that this is what you have to do to be well informed

To justify the decisions you make

To collect information in case it might be useful one day

Having a piled up in-tray and a cluttered desk demonstrates how busy and important you are

You get the same information from a lot of different sources (this includes reports and articles which basically repeat a large percentage of their content)

Cross-checking that the original information you have is accurate

Getting on everyone’s ‘copy-to’ (cc) mail list so that you don’t miss out

Poor information searching skills, so that a lot of irrelevant information is retrieved.

Research published shows that companies risk wasting their investment in technology implemented to manage information because they are failing to tackle the ‘human hurdle’ – up to two thirds of respondents had problems with information overload, employees not having time to share knowledge and reinventing the wheel.

So, we’re getting more information, more rapidly, from more sources, in more formats – and less and less time to deal with it.

**Evaluating information**

For our decision making to be quality decision making, the information that supports it needs to be quality information. But what does ‘quality’ actually mean in this context? How can we recognise and measure it? What criteria or benchmarks can we use? One difficulty is that, over time, the quantitative and qualitative value of information can decay. Also, information quality in this context can be considered from the point of view of function (does it do what it is meant to do and what are the functions it satisfies or supports?) or of form (the image or intangible benefits that accrue from having the information).

What’s the added value?

For some experts, the problem of information overload is to a great extent part of a failure to create ‘quality’ information – that is, information that has real value-added content. This of course assumes that you have a very clear idea of what it is that makes a piece of information add value.

Simpson and Prusak (1995) believe that the value of information can only be measured in terms of the benefit you get from using it. But how do you measure this ‘benefit in use’? Many approaches have been tried but none has stood the test of real-time business practice, and yet individuals try to add value in their own communications all the time! Think back to your last couple of conversations. Were any of the people involved:

trying to verify the information given (‘Well, you say that, but is it actually true?’)

drawing conclusions from the information (‘It looks as if what we’ve got here is...’)

challenging something that didn’t sound quite right (‘Hang on a minute...that can’t be right...’)?

During these kinds of conversations, although you may not be aware of it, you are evaluating the information you received so that you can work out what to do about it.

From this kind of instinctive evaluation, Simpson and Prusak have devised a model that proposes five universal elements of value in information, shown in Figure 2.1.

**Truth**

**Weight**

**Scarcity**

**Guidance**

**Accessibility**

Figure 2.1 Adding value to information Source: Simpson and Prusak (1995)

**This five universal elements:**

**Truth.** How confident are you about the validity of the information? In operational or systems terms, its validity or otherwise may be obvious; in management processes, where you are dealing with a mix of factual information, inferences and subjective judgement, things may be less clear-cut. But you will still need to have a high degree of confidence in the information if you are to make a decision based on it.

**Scarcity**. This refers to the value of information which is ‘new’ or not freely available to competitor organisations. There may be little that is new at first glance in your day-to-day operating data. However, using freely available information combined in a new way, or with a flash of insight, can produce something that provides your organisation with a real advantage over the competition.

Guidance. This is the extent to which information points the way to what action needs to be taken in a certain situation or set of circumstances. This is most obvious in sets of procedures or instructions, but information can also give guidance where you are highlighting a problem or diagnosing the causes of a problem.

**Accessibility.** This is the availability of information to its potential users when it is needed and in a form that they can use. Information is only of use if people know about it, can get at it and can understand it. An important element of this is the way the information is presented – does it encourage and help understanding by the user?

**Weight.** This is what prompts recipients to treat the information so seriously that they will act on it. It will incorporate things like relevance to the user’s situation, timeliness and accuracy. It sometimes has very little to do with the information itself, but is closely tied up with the credibility or authority of the person providing the information.

**Evaluating information from the Internet**

These five universal elements can be applied to information generated within or outside the organisation. They can be used to evaluate information received over the Internet. You can use the following checklist to evaluate sites that are dedicated to a particular organisation or based around a specific subject area. Each question is followed by the most relevant universal element of evaluation:

Checklist for evaluating websites

Is the purpose of the site clear? (Guidance)

Are contact details and basic information easy to find?

(Accessibility)

Is the coverage of the site appropriate for its purpose?

(Accessibility)

Does the institution responsible for the site have an established reputation and expertise? (Weight)

Is the information likely to be accurate? (Truth)

Is the information current? Are there details of when it was last updated? (Weight)

Is a site map provided? Is navigation clear and straightforward? (Guidance)

Is the information well presented and arranged?

(Accessibility)

Does the site compare well with those of similar organisations or those in the same subject area? (Weight)

Are there good help facilities? (Guidance)

Does the site provide links to other sites, or supporting materials? (Guidance)

Are these other sites or supporting materials useful? (Truth)

Good practice for reducing overload

There is no single tool or technique that will provide a magic answer to all your information overload problems, but there are techniques that can help. Bawden et al. (1999) divide these techniques into managerial and technical.

**Managerial techniques**

On the managerial side, a lot of the techniques come under the general heading of time management. You can re-take control of your information by managing your time more effectively, using some of the following techniques:

Structure your information searching more intelligently, and link it directly to your goal: why are you looking for this information, and how can you best find it? This is likely to be quicker and more effective than just surfing in a random way, hoping that something useful will turn up.

Follow the classic time-management recommendation and ‘handle a piece of paper only once’ (the same applies to electronic messages). Take action on it immediately or delete/ bin it.

Be very selective about the newsgroups and mailing lists you join – they can generate a lot of irrelevant information that is timeconsuming to read through.

Delete irrelevant e-mails without reading them.

Only file material when you know it will be difficult to find it again.

Improve your own information literacy – your ability to retrieve, evaluate, organise and use information from a variety of sources. This will include effective management of both paper files and e-mail folders.

All this lies within your own hands. If you are looking at reducing organisation-wide information overload, putting out some sensible rules for e-mail etiquette is a good start. The European Forum for Electronic Business has developed a code of practice to help organisations use e-mail more effectively. Here are some of its main points:

**Guidelines for using e-mail**

Do you need to e-mail at all? Sometimes it’s quicker to telephone.

* Give your messages a meaningful title – not ‘Meeting’ but ‘Team meeting 29 April’.
* For clarity (and to save other people’s time) restrict action requests to one recipient only, and copy to (cc) anyone else who needs to know.
* Keep your messages brief.
* Don’t mail or cc more people that absolutely necessary.
* Think very carefully before putting a message on a distribution list for general use.
* Using the ‘BCC’ (Blanket Carbon Copy) field instead of ‘To’ for messages to several people will reduce message size.
* Currency symbols can be changed in transmission. If your e-mails are likely to contain references to different currencies, it’s better to use an agreed alpha abbreviation like GBP for sterling and USD for US dollars.
* Use the ‘Urgent’ flag sparingly, or its impact will be lost.

**Technical techniques**

On the technical side, there are systems for ranking and filtering e-mail and other messages; check for details with your information technology (IT) department. As an individual trying to reduce overload on the Internet, your best approach, as indicated earlier, is to make as much use as you can of personal software agents and any customisation offered by the major search engines.

Identify information overload and assess why it occurs

When the amount of information received exceeds that desired or needed by a user, it becomes a hindrance and a potential cause of stress, and the user experiences information overload.

Assessing the extent to which you contribute towards your own information overload is a good first step in improving the way you manage and use information.

Evaluate the information you receive by assessing its quality and value to you

Simpson and Prusak (1995) propose that you can evaluate the quality of information available to you using five criteria: weight or importance, truth or validity, the extent to which you rely on the information for guidance, accessibility and scarcity.

If you are receiving information that is of poor quality, then communicate your needs to your information source to see whether it can be improved.

Reduce your information overload

You can reduce information overload by becoming more selective about the information – including e-mails – that you access and read, and by developing systems for effectively managing hard and soft information.

# Communicating information

People handle astonishing quantities of written text on a daily basis, both consciously and unconsciously: newspapers at breakfast, advertising hoardings and shop fronts on the way to work, reports, memos and e-mail on the desktop. They may read a book on the train home, or pick up a few text messages from friends. It is worth pausing for a moment to think about the different ways in which such channels of communication get their message across to you – and what influences how receptive you are to what they are trying to tell you or persuade you to do. How often do you stop reading – simply switch off your attention – from something that is longwinded, difficult to follow, boring or full of errors?

There are lots of very practical reasons why everyone should aim to communicate clearly:

* It makes it easier for the recipient to understand the message, which saves time
* Written instructions that are clear and unambiguous are easy to follow and act upon
* A good written case can be a powerful aid to influencing
* In the case of a dispute (for example a disciplinary case) your written reports may be produced as evidence in an employment tribunal or a court of law
* What you record now may be a precedent that will need to be referred to for guidance in the future
* If the messages aren’t understood by the reader, do they count as communication or just a waste of your time and everyone else’s?

Many successful business leaders have recognised that the ability to write persuasively – getting people to take their message on board or do what they want them to do – is a key skill. The most effective documents, whether long or short, are those where the author has taken the trouble to ensure maximum impact.

In this chapter we will:

* Identify the features of clear written communication Evaluate your writing style Plan an effective presentation
* Develop notes and visual aids to support your presentation.
* Planning and structuring your document

You may think that a logical approach is the best way to do this. Surely the facts will speak for themselves? Or perhaps you think that if you really want to get your own way quickly, a bit of coercion will do the trick? Andrew Leigh (1997) believes that learning to develop a persuasive writing style is the best way of encouraging people to accept and endorse what you have to say.

Purpose

An important aspect of this is the purpose of the document, which you need to be absolutely clear about:

What do you want your readers to do?

What outcome do you want to achieve?

Readers

An essential step in planning a document is to put yourself in the reader’s shoes, and to try to predict how they will understand and react to it. Consider the following questions:

What do your readers expect to gain by reading your document?

What length of document will be appropriate for the purpose and the recipient? Are they likely to want a one-page summary or a 10-page analysis?

How much time will they have to read it?

What is their likely standpoint on the topic, and how can you counter resistance?

What questions are they likely to raise?

Since your aim is to communicate with people and persuade them rather than to antagonise them at the outset, it is always useful to start by establishing some common ground and getting across that you understand and respect their position. This is important, even if you then go on to provide evidence that their position is no longer tenable and that they will have to consider changing it.

Structure

You can strengthen whatever case you are making (and this applies to the shortest e-mail or longest report) if you structure a document carefully so that:

the information in it is prioritised, with the most important information coming first – this will mean giving the ‘headlines’ and main conclusions first, not leaving the punchline until the end

you have carefully selected what to include and what to omit – this involves thinking about what information your readers need

you give some indication of what should or will happen next – in other words your recommendations.

Politeness and clarity will get you a long way. However, there are other means of ensuring that your message is received positively, such as style and tone.

**Style and tone**

You may not have much leeway with style, as a corporate house style may exist that you have to adhere to. If you are not restricted by a corporate house style, you can make your style more interesting by using active language rather than passive.

Think about the difference between:

‘It is generally recognised within the company that ...’ and

‘As you know...’

Or the difference between:

‘The project outline was put together by the author of this document’ and

‘I put the project outline together’.

Keep the language simple and straightforward by avoiding features such as:

double negatives, for example, ‘it was not impossible to foresee the consequences...’

long words

a complicated sentence construction

technical jargon and other features that you may think look professional but in fact just get in the way of understanding and actually lessen the impact of your message.

The aim is to ensure that your reader progresses smoothly through the document, without having to stop and puzzle out what you are trying to say. As well as keeping it simple, it is useful to keep it short, and this may require some discipline and firm editing. Think about how long it takes to unravel a sentence such as the following:

It is our opinion that, in the circumstances, and with all things considered, the best way forward will be to talk initially to HQ, RSB and GRE staff, then JPU, EN and PNU staff about the new procedures and make sure that they are up to date.

If the author of this example is so unconfident about the way ahead, why should you, as the recipient, be convinced? Also, as the recipient, will you instantly recognise all the staff modules referred to? Is it the staff or the procedures that need to be up to date?

Style tips checklist

Avoid long sentences or paragraphs

Use simple, active language

Avoid double negatives

Avoid jargon or overuse of acronyms and abbreviations

Use bullet points and numbered lists to break up the text

Edit ruthlessly – don’t hang on to a nice phrase that adds nothing just because you thought of it and like it.

In verbal communication, what you say is often not as important as the way you say it, and the message communicated to the recipient may have little to do with the actual words used. In written communications too, the tone you use may be so inappropriate as to be unacceptable, even if the facts it contains are true. Common errors are being:

* aggressive rather than assertive – ‘I want this revised and on my desk by 8.30am tomorrow – or else.’
* patronising – ‘I realise that your experience of this process isn’t as extensive as mine. However...’
* dismissive – ‘This is too trivial to comment on. Just go away and sort it.’
* critical – ‘That was really stupid.’

Before you send a document, check it by putting yourself in the receiver’s shoes. How would you react to being on the receiving end of it?

**Other good practice**

If the message is clearly set out, does it matter if you make minor errors, break the odd grammatical rule or misspell words? Think about how you would feel if your bank got your name wrong or a brochure for a smart hotel contained basic typing errors. You may feel that these small mistakes undermine the message the company is trying to convey. The key issue here is reliability. The message you receive is that if these people can’t take the trouble to get the basics right, what else can’t they be bothered to do, and what does that say for their levels of customer care?

These days, most typing and some grammatical errors can be picked up automatically by your spellchecker, but text will still benefit from

Politeness and clarity will get you a long way.

proofreading to ensure that errors such as ‘their’ instead of ‘there’ are corrected, and that all personal names in the document are spelled correctly.

You also need to keep a lookout for discriminatory language. This pitfall has been around long enough for acceptable alternatives to become current, for example ‘workforce’ for ‘manpower’.

Be careful how you use numerical data in written communications. Incorrect numbers, or statistics provided out of context, can completely undermine your otherwise convincing case.

# Information systems

So far in this book we have focused on how you can improve your own management of information. But you are working within an infrastructure of organisational information and knowledge systems.

This theme explores the key issues in systems development and reviews how the Internet is transforming corporate communication systems. The massive growth in computer systems and the use of Web-based technology have attracted a corresponding rise in the number and variety of threats to security. With widespread desktop access to e-mail and the Internet, all managers – indeed all staff – need to be alert to the dangers of unauthorised access to an organisation’s systems.

In this chapter we will:

Identify the key stages in the system development life cycle and your contribution towards it Identify the benefits of an corporate intranet

Assess how well your organisation manages data security.

Key issues in systems development

It’s a fair guess that many of your working hours are spent in front of a computer, using the information system in different ways: reading, inputting, organising and sending out data. Typically, there will be times when the system won’t do what you want it to do, or you think, ‘Why did they design it like this? Why can’t I just go straight to... It’s useless...’ However, the way your system was designed probably originally depended (at least to some extent) on the way local users and managers described the jobs they wanted the system to perform. This legacy will have a crucial impact on the way you manage your own incoming and outgoing information and, by extension, the extent to which the organisation as a whole manages and makes accessible its information and knowledge resources.

At some point in your career you will be involved in providing input to a major system upgrade or replacement, even if you have not yet done so. Understanding how and why systems are developed, and the possible pitfalls, provides important lessons for managers involved in future systems development.

**Systems model and life cycle**

Computers have been with us for a long time, but their development has been surprisingly unpredictable. Up until the 1970s there were few attempts to produce a coherent view of computer operations. One of the first models was developed by Richard Anthony (Mason and Willcocks, 1994). This was actually a model of organisational behaviour, which put forward the view that there are three basic types of decisions made within organisations:

* Strategic: these involve setting overall goals and objectives and determining how to meet them
* Control: making sure that the organisation’s functions are carried out efficiently and resources are used effectively
* Operational: relating to day-to-day operations, ensuring that tasks are done properly, in the right order, at the right time.

As a model, Anthony’s pyramid (see Figure 4.1) has had a huge influence on management thinking – and will still be recognisable in your own organisation today. This hierarchical view of functions within the organisation was mirrored by the systems managers’ approach, which was geared to the belief that the logical starting point for introducing computer applications was at the operational level, working upwards from there to the rarefied heights of supporting strategic business decisions.

**Information system**

**Function organisation**

Lead

Control

Activity

Strategy

Tactics

Operations

DSS

MIS

TPS\*

\*Transaction Processing Systems

Figure 4.1 Anthony’s pyramid Source: Mason and Willcocks (2001)

From the late 1960s it was realised that systems development actually consisted of well-defined stages, and a ‘life cycle’ view of systems emerged that formed the basis of many different methodologies for systems development (Galliers et al. 1999). Even so, it took a long time to realise that the life cycle was not linear, with a neat start and end point, but needed to be viewed as a continuing process in order to:

review and correct earlier errors and misconceptions

revisit and retune the original specification in the light of changing requirements

deal adequately with the problem of a growing number of systems involving increasing amounts of maintenance.

Figure 4.2 gives a graphical view of a system life cycle.

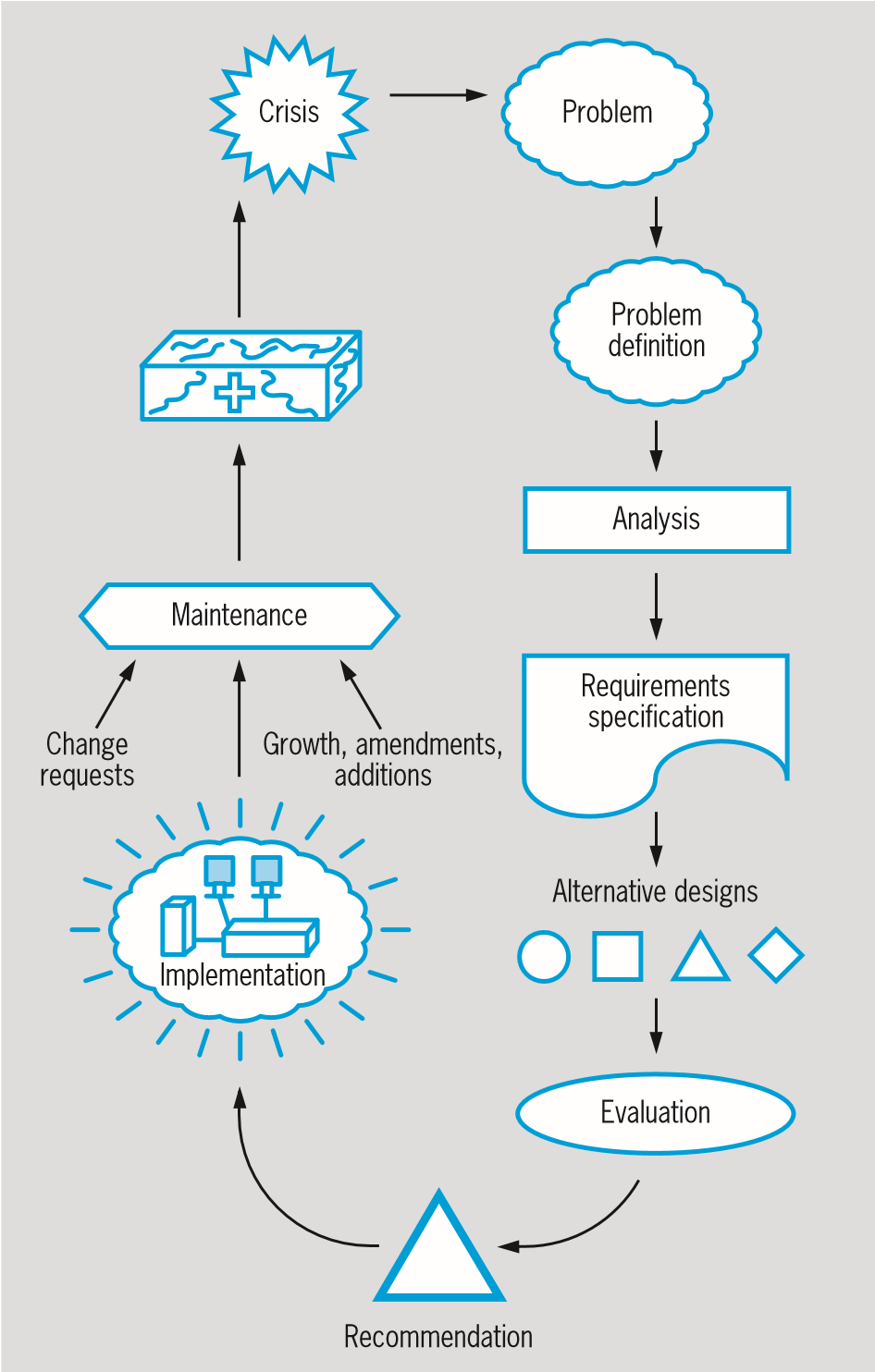


Figure 4.2 The system life cycle

**Roles in developing information systems**

If you are not an IT manager, where do you fit into this process? You may not realise how important you are. Recent years have seen the growth of user involvement at every stage of systems development. As the impact of systems development has become visible and organisation wide, organisations have belatedly realised that it has a human dimension, and that ignoring this can wipe out all the advantages of your expensive new system.

Many organisations are now battling with the problem of so-called legacy systems. This is a system that was developed 20-odd years ago which was designed to solve the problems the organisation faced at that time. The trouble is that 20 years on, they are out of step with evolving business needs and can hold back organisations that want to apply a more up-to-date set of routines. They are now seen to be inflexible, expensive to maintain and even more expensive to replace. This has all lent urgency to the need for genuine user involvement in systems design, and few would dispute the necessity for this. But mistakes still happen. Lytle (1991) devised an information systems development disaster menu, shown in Figure 4.3, that still holds good. As you can see, it shows all the things you shouldn’t do when developing computer systems. If you do, then you’re heading for disaster.

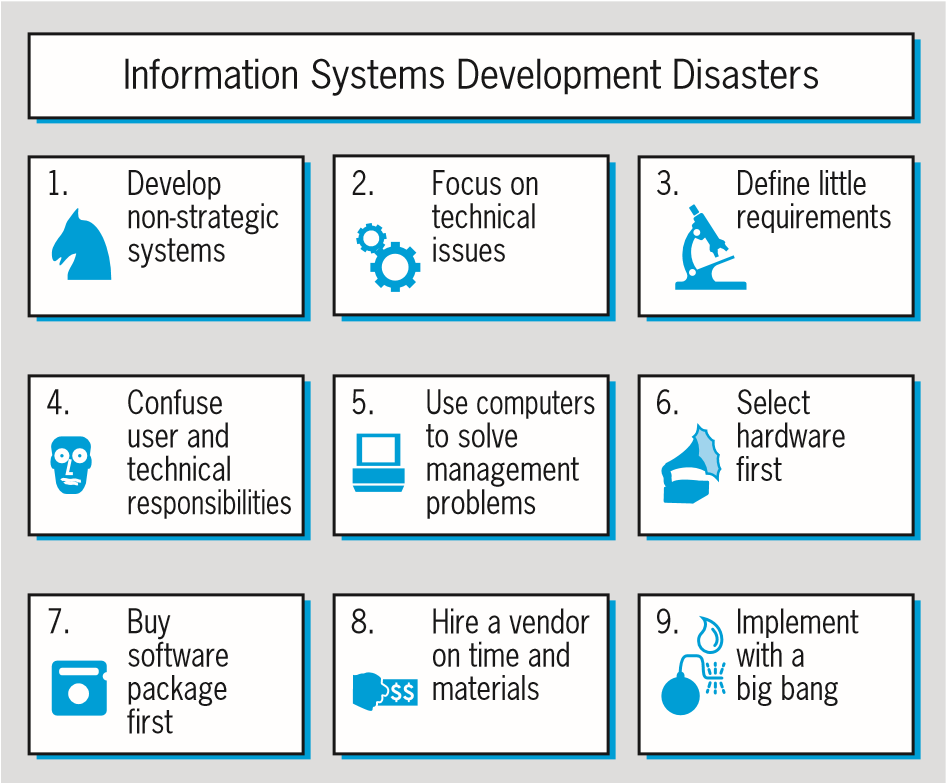


Figure 4.3 An information systems disaster menu

Let’s turn the negatives from Figure 4.3 around and see what happens.

1. Develop strategic systems. What are the key strategic information areas for your business? These are your key business critical systems.
2. Don’t focus on technical issues. Systems are not a matter of hardware and software; they are a matter of the right hardware and the right software, selected on the basis of user needs and organisational critical success factors.
3. Take time to define requirements at ‘big picture’ and operational levels. This is a key area of a manager’s job. Welcome it as a real opportunity to examine quite critically what you are doing now. Do you still need to do it at all? Are there other, better ways of achieving your objectives? This will involve you in the following activities:

Analysing all the business processes that you manage. For example, if you’re an HR manager, this will include such areas as recruitment, appraisal and reward, workforce planning and job analysis and design. How do the functions that you manage fit into the wider organisational picture?

Documenting the type of information you need to carry out your various tasks. Are you receiving everything you need, in the most efficient and timely way? Are you getting too much information, information you don’t need at all or need less often?

Drawing a picture of the data that flows in and out of your unit (don’t forget to include informal data sources). What, and from where, are your data feeds? Who else receives the same information? What do you do with the information you receive? If you process it in some way, how is this done and what are the outputs? Who are they delivered to?

1. Be clear about user and technical responsibilities. Make sure that responsibilities are clearly defined at the outset, with staff allocated the roles that they are best qualified to do. That way you can build up co-operation and mutual respect, not mutual antagonism.
2. Use management, not computers, to solve management problems. Too often, problems that are actually related to poor management are conveniently blamed on ‘the system’. Computers can do lots of things to improve your data management and information flow, but they can’t resolve problems of organisational culture or personality clashes. Do what you can to get these issues resolved before your user specification gets underway.
3. 6, 7 and 8 Select hardware and software to fit the requirement, and be specific about any customisation required. Adopt the motto ‘focus on functionality’. Vendors are experts at showing off their systems to their best advantage, but will the system do what you want it to do? Can the software be customised, and will the vendor do this? Will they need to involve third-party suppliers? If customisation is required, get this specified in terms of activities and costs. Hiring on a time and materials basis is a recipe for a long drawn out, expensive and increasingly sour relationship.

Find out when the next version of the software is due. What do the licensing arrangements really mean in terms of multiple, realtime access? It is important to check out the vendor’s financial stability as a standard procedure, and it may also be worth checking the business press to see whether the company (or its parent) is involved in merger discussions or is about to be swallowed by a giant competitor.

9 Beware ‘big bangs’. It is rare now for a complete system to be developed in full before live operation, and for good reason. It takes time to develop a complete system, and while this is happening there will inevitably be evolution and change in user requirements. These need to be incorporated into the developing system and checked by the user to see if they work. It is much better to develop a prototype system that can be piloted (tested, reviewed and improved) and used to inform the final development.

**Intranets and extranets**

The idea of getting computers to communicate with each other, either on a one-to-one basis or via a network, has been around for a long time. Networks are of two main types, determined by the size of area that they cover:

Local area network (LAN) which can link computers in a single room, one building or several buildings that are geographically close (for example on a university campus)

Wide area network (WAN) which, as its name implies, can link computers that can be hundreds or even thousands of miles apart.

The widespread take-up of the Internet in recent years has transformed the way that networks are used, and added a whole new dimension (with new opportunities and new problems) to the way in which organisations communicate internally and externally.

An intranet is a network (LAN or WAN) that utilises Internet technology. However, unlike the Internet, access to an intranet is restricted to specific individuals, and the data it holds will be secured behind stringent data security systems or firewalls.

An extranet uses Internet technology to link together intranets in different locations. In contrast to intranet transactions, extranet transmissions take place over the Internet, and so are not secure. This necessitates strengthening the security of the connecting portions of the Internet. This can be done by creating ‘tunnels’ of secured data flows. The Internet with such tunnelling technology is known as a virtual private network (VPN) – see Figure 4.4.

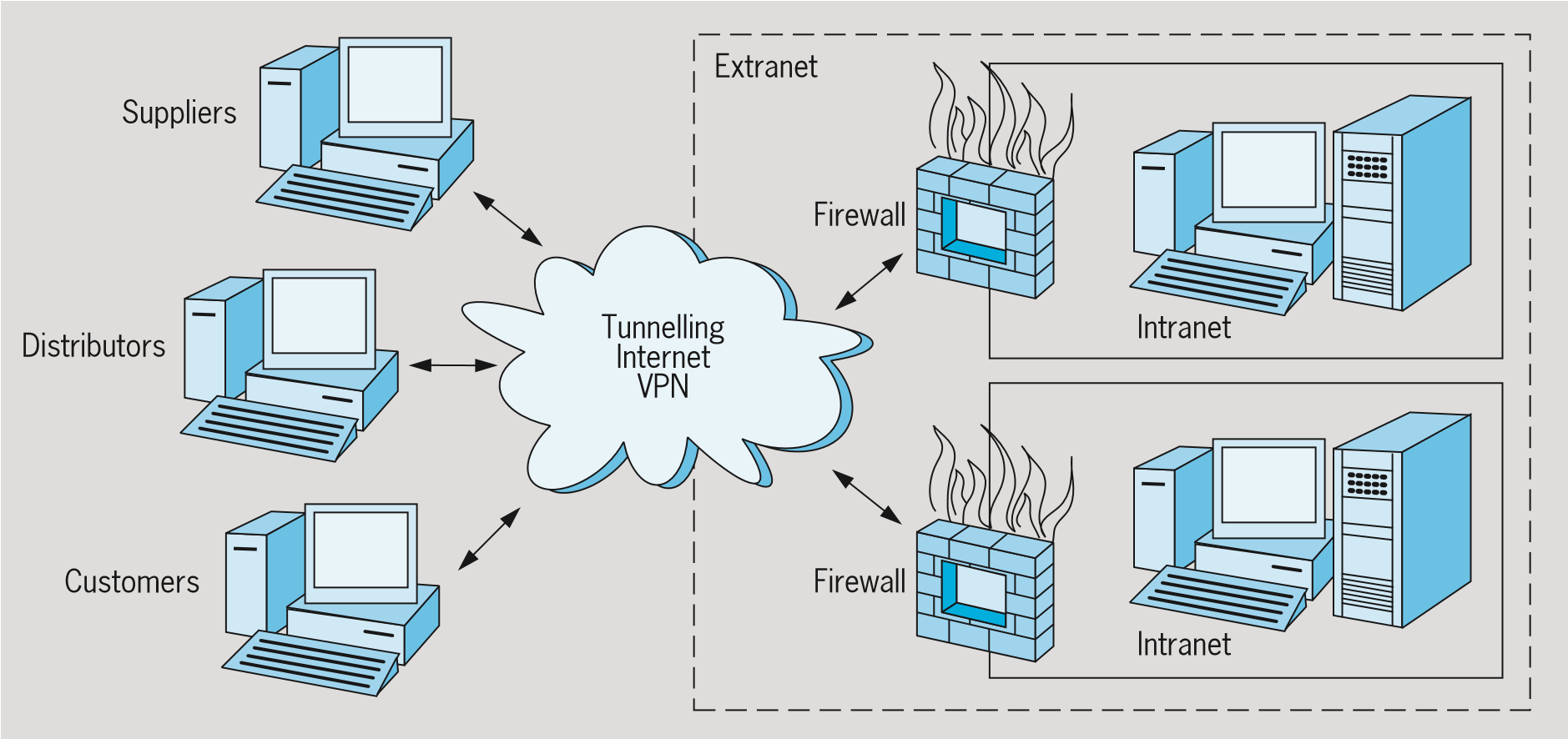


Figure 4.4 Diagrammatic contrast of the Internet, intranet and extranet

Both systems and departmental managers have been quick to seize the advantages of a corporate intranet:

With businesses under significant pressure to empower employees and to better leverage internal information resources, intranets furnish a very effective communications platform – one that is timely and extensive. A basic intranet can be set up in days and can eventually act as an ‘information hub’ for the whole company... Intranets can provide the following features:

* easy navigation (internal home page provides links to information)
* ability to integrate a distributed computing strategy (localised web servers residing near the content author)
* rapid prototyping (can be measured in days or even hours in some cases)
* accessible via most computing platforms
* scaleable (start small, build as requirements dictate)
* extensible to many media types (video, audio, interactive applications)
* can be tied to ‘legacy’ information sources (databases, existing word processed documents, groupware [software designed for group communication and shared group use]).

The potential business benefits of intranets are numerous:

* improved information flows between employees, customers and suppliers
* reduced geographical constraints: worldwide organisations can now communicate as a logical whole
* easy access to information through a common single interface
* better access to information = quicker and better decision making = reduced cost
* reduced cost of IT operations: Internet-derived technology is a cheap way of improving communication and data flows
* increasing an organisation’s profile on an international scale: selected components of an intranet can be shared via a public interface on the Internet.
* It can be used for a huge range of practical applications, from making corporate information available to all employees to providing specific information to a salesforce in the field or external stakeholder groups.

Case study: Albion Oil

Having secured a contract to assist with exploration and exploitation of natural resources in North Africa, Albion Oil needed a means of handling its rapidly growing information assets and communicating efficiently with all members of a project team scattered across Europe, North Africa and North America. A technical review of the existing infrastructure revealed a mix of Macs, PCs and UNIX systems. What to do? These are the steps they took:

Produced a project initiation document defining key business objectives

Restricted the scope of the project to users involved in the North African exploration

Defined key deliverables, including the establishment of a user group

Defined key success criteria – e.g. providing users with reliable and secure access to information and round-the-clock support and training

Documentation, clarification of roles and responsibilities and mechanisms for addressing security, reliability, contingency and other issues were all established as necessary targets

Clear targets were identified in terms of ‘publishing’ information (authoring, formats, ownership), locating information and Newsgroup ‘netiquette’

The project was controlled by a small tightly focused team

Beginning with an online telephone directory of staff as a pilot, Albion moved to adopt the same approach to other project related data. The results have been a system that matched management and user requirements in which the technology was clearly focused and not just applied for its own sake.

Source: Fishenden (1997)

**Data security**

Security threats can present themselves in direct form, through hackers (and as far back as 1997 it was estimated that the Internet is hacked into every 20 seconds) and through indirect information systems penetration (Mitchell et al. 1999). These indirect threats occur in four major types:

Worms: a worm is a program that, once established, can spread copies of itself throughout a network

Trojan horses: these are also programs that appear to be carrying out a non-malicious activity which, when activated, reveal their true destructive intent

Logic bombs: these are programs activated by a specific event, for example St Valentine’s Day

Viruses: like a medical virus, these ‘infect’ other programs.

A popular route in for these invaders is via e-mail – and they don’t always come in as attachments. The header message is usually friendly and intriguing, encouraging the user to believe that it is a message from a friend or admirer.

The results of these attacks can range from the irritating and embarrassing to the devastating, and can include the destruction of data or its modification, interception or fabrication by unauthorised personnel.

The Melissa virus

Melissa was an e-mailed virus that emerged from nowhere to overwhelm commercial, government and military computer systems, leading the FBI to launch the biggest Internet manhunt ever.

Melissa affects Word 97 and Word 2000 documents. If launched, this virus will attempt to start Microsoft Outlook to send copies of the infected document to up to 50 people in Outlook’s address book as an attachment.

The e-mail subject line reads: Important message from [username]

While the message reads:

Here is that document you asked for ... don’t show anyone else. ; – )

Source: MelissaVirus.com (www)

Viruses often spawn ever more dangerous variants. The ‘I Love You’ virus, which appeared in Spring 2000, had 50 variants by October that year.

The growth of e-commerce has seen a surge in opportunities for business fraud and other security issues.

KPMG survey

The management consultancy firm KPMG has produced some worrying findings from its 2001 Global e-fr@ud Survey:

E-fraud is a growing problem for companies around the world.

Although credit card numbers and personal information are of prime concern to customers, less than 35 per cent of companies surveyed have had security audits performed on their e-commerce systems.

50 per cent of businesses identified hackers and poor implementation of security policies as the greatest threats to their e-commerce systems. However, the company is at greater risk of being the victim of an internal security breach.

83 per cent of respondents feel that the public perceives the traditional ‘bricks and mortar’ business as more secure than e-commerce-based dot.coms.

Source: Adapted from KPMG (2011)

Methods of data security

There is a range of methods of varying complexity that organisations can use to protect themselves from unauthorised access. See Table 4.1.

Table 4.1 Methods of data security Source: Hawkins et al. (2000)

|  |  |
| --- | --- |
| *Method* | *Description* |
| Firewalls | The first line of defence from the outside. Acts as a security guard for the company’s internal network, filtering all incoming traffic from the Internet. A good tool for networks connected to the Internet |
| User authentication | Verifies the identity of the user. Could also be used to restrict access to certain resources within the network. A requirement for any user accessing a corporate network |
| Data encryption | Scrambles the data before and during transmission. Use this method when data protection is important |
| Key management | Acts like a ‘key’ to access encrypted data. Maximum protection to protect data from unauthorised parties. Use in conjunction with data encryption |
| Digital certificate | Like a watermark on a bank cheque – this is an electronic ID card that establishes your credentials when doing business on the Web |
| Intrusion detection system (IDS) | Scans the network for abnormal activity and security breaches. A minimal requirement for any corporate network |
| Virus detection | Scans the network data for viruses, providing both prevention and cure if updated regularly. One of the best defences for data protection |
| Virtual private networks (VPN) | A secure private data network developed on a public data network like the  Internet |
| Extranets | A secure private data network that uses a public data network like the Internet to extend a company’s network to suppliers, vendors, partners, etc. A company can minimise its overheads by exchanging data through an extranet via electronic data interchange (EDI) |

**Data protection**

Another aspect of information security is data protection. One of the effects of increasing globalisation of business activities and cross-border data transactions has been to raise awareness of the need to safeguard personal details which are held in either manual or electronic systems. Several basic principles of data protection have now been established and codified in law. For example, in the UK, anyone processing personal data must comply with the enforceable principles of good practice. These are that personal data (which includes facts and opinions, and information regarding the intentions of the holder of the data towards the individual) must be:

* fairly and lawfully processed
* processed for limited purposes (for example legitimate business purposes)
* adequate, relevant and not excessive
* accurate
* not kept longer than necessary
* processed in accordance with the data subject’s rights
* secure
* not transferred to other countries without adequate protection.

To find out more about data protection requirements, you could do an Internet search for your country using key words ‘data protection’. A useful site for UK-based managers is www.dataprotection.gov.uk, where these principles can be found.

Every organisation, whether government, public or private sector, needs to devise its own specific security arrangements. The following provides a useful checklist of good practice in information security management.

Good practice in information security management

Draw up a security policy document

Allocate specific security responsibilities

Institute security awareness and training programmes for staff

Have a formal reporting procedure for security incidents, and make sure that staff are aware of it

Implement good antivirus controls, updated daily (a problem here can be with mobile staff who spend a lot of time on the road and forget to update)

Identify risks to business operations and develop disaster plans

Control proprietary software copying – make sure that only software developed by or licensed to the company is used

Safeguard organisational records to protect them from loss or falsification

Comply with your country’s data protection legislation and ensure that the information you record is only used for general business purposes

Monitor compliance with security policy throughout the organisation and review arrangements periodically.

Source: British Standards Institution (2018)

Recommendation

Identify the key stages in the system development life cycle and your contribution towards it

Systems pass through a series of stages during their development; problem definition and analysis, specification of requirements, design and evaluation of options, recommendation and implementation.

Systems development should be seem as cyclical rather than linear. Maintenance is required on an ongoing basis to manage change requests, growth, amendments and additions.

User involvement is critical at each stage of the life cycle if the system is to meet its purpose and be fit for use.

Identify the benefits of a corporate intranet

Intranets utilise Internet technology and have become a very popular means of improving information flow and communication through an organisation.

Access is restricted to authorised individuals and data is secured behind firewalls, making intranets a safe and cost-effective approach to networking.

An effective intranet should have directories and search engines that make it easy for users to find and retrieve the information that they need. Achieving this level of user friendliness requires careful planning.

Assess how well your organisation manages data security

Security threats present themselves directly from hackers and indirect threats of four major types: worms, viruses, Trojan horses and logic bombs.

Data can be made more secure through the use of firewalls, user authentication, data encryption, key management, digital certificates, intrusion detection systems, virus detection software, virtual private networks and extranets.

Organisations should minimise the risk to their data by designing and implementing data security and management policies.

**Use standards for Information Security Management System (ISMS) – ISO/IEC 27000 – 27005.**

# Knowledge management

Knowledge management has been hyped as a must-have business solution for a number of years now. However, quite what it is and how you are supposed to manage something so intangible is still a source of much confusion.

In this chapter we you will:

Define knowledge management and its relationship to

* learning processes
* Identify the barriers to knowledge management
* Identify the critical success factors in knowledge management
* Mobilise knowledge management in your organisation.
* How do you manage knowledge?

One of the problems in trying to define knowledge management is that it is sometimes difficult to see how it differs from information management. Swan et al. (2000) see the two as being very closely associated, with each interacting with the other. See Figure 5.1.

**Facilitates the**

**development of**

**Information**

**Knowledge**

**Facilitates the**

**creation of new**

Figure 5.1 The dynamic relationship between information and knowledge

As explored in Theme 1, knowledge combines information, experience and insight into something that is unique to every individual. But what is knowledge management? Here are two definitions:

Knowledge Management is the explicit and systematic management of vital knowledge and its associated processes of creating, gathering, organising, diffusion, use and exploitation. It requires turning personal knowledge into corporate knowledge that can be widely shared throughout an organisation and appropriately applied.

Knowledge Management...has been described as ‘knowing what you know, knowing what you don’t know, learning what you need to know and sharing it.’

It is worth being aware that different disciplines are concerned with recognising, valuing, capturing and measuring the knowledge and expertise within organisations, and adopt a range of terminology. Intellectual capital, for example, is a term that is often used alongside knowledge management. It has a broader definition than knowledge and comprises employees’ talent and knowledge, customer loyalty, the value of brands, patents and copyrights and research. In this theme we focus on the concept of knowledge management given in the definitions from Skyrme and Newing above.

The concept of knowledge management grew in the early 1990s from a study of how Japanese companies create knowledge within the organisation, disseminate it and embody it in new products and services (Nonaka and Takeuchi, 1995). The Nonaka and Takeuchi model classified human knowledge into two kinds:

Explicit knowledge: this is formal, easily identifiable and general knowledge, the sort you find in mathematical expressions, or specifications and manuals. Because it is explicit and obvious, it can easily be transmitted between individuals.

Tacit knowledge: this is difficult to articulate, as it is personal, ‘hidden’ knowledge, embedded in an individual’s experience and coloured by their personal beliefs and values.

These are the two basic building blocks of knowledge creation. The assumption is that knowledge is created through the dynamic interaction between explicit and tacit knowledge. For organisations to succeed, they need to find ways to make explicit and share the wealth of tacit knowledge that is locked up within individual employees’ experience. Nonaka and Takeuchi saw the explicit/tacit relationship as a spiral process, in which interaction takes place repeatedly. Willard (1999) reworked and simplified their original spiral (see Figure 5.2), and sees the sequence in this way:

Someone has a bright idea, and finds a way (sometimes easily, sometimes with great difficulty) of expressing that idea. This means that the idea moves from the tacit (personal knowledge and experience) to the explicit – expressed in a way that everyone can understand.

The person who had the idea combines this with other known elements to form some kind of context (so we have explicit added to explicit).

This is then communicated to colleagues, who begin to ‘get the picture’ and start to think about it on the basis of their own knowledge (so the explicit idea moves to tacit reflection and analysis).

Through discussion the idea grows and develops, and colleagues all contribute to the implicit understanding that builds up (adding tacit to tacit).

New ways are found to express the idea, more people are informed and the idea is increasingly combined to present a bigger idea (and so on).

The new understanding is now institutionalised – turned into a working procedure or implemented as a working practice or rule.

**Organisation**

**Team**

**Individual**

**Explicit**

**Tacit**

**Externalisation**

**Internalisation**

**Combination**

**Externalisation**

**Internalisation**

**Combination**

Socialisation

Figure 5.2 The knowledge spiral

Much of the literature about knowledge management relates to the technology: the systems for sharing and exploiting the newly explicit knowledge. However, knowledge management is actually about people and their interaction, rather than technology – though technology is a powerful enabler.

Knowledge management and learning

An important aspect of knowledge management is the way people

(and organisations) learn and how they approach problem solving. This is a good point to revisit Argyris and Schön’s theories-in-use (the private, self-generated theories that govern our behaviour).

Argyris and Schön (1974) built a model of the processes involved in the theory in practice (see Figure 5.3) that has three elements:

Governing variables (or values): there are likely to be a number of these and any action taken is likely to impact on them.

Action strategies: what people do to keep their governing values within an acceptable range.

Consequences: what happens as the result of an action.

Consequences can be intended or unintended.

**Governing variable**

**Action strategy**

**Consequences**

Figure 5.3 Theory-in-use model

Where the consequences of your action strategy are in accord with your governing values, the theory-in-use is confirmed. But what happens if the consequences work against your governing values?

Argyris and Schön suggest that there are two responses to this mismatch, which they describe as single-loop learning and doubleloop learning. When something goes wrong, a common response is to look for another strategy that will work better, but still within the framework of existing governing variables or values – the plans, goals or rules of behaviour that we are familiar with. This is singleloop learning. A more radical approach is to examine critically the governing variables or values themselves, to test how valid they still are. This in turn can lead to a change in the whole framework in which the action strategies and consequences are developed – a double-loop (see Figure 5.4).

How does this translate into organisational learning and behaviour? Looked at in organisational terms, error and correction in a singleloop learning environment will work within the organisation’s existing policies and objectives, but otherwise carry on with these unchanged. Double-loop learning will occur when errors are corrected in ways that involve the modification of the organisation’s underlying norms, policies and objectives. Argyris and Schön argue that double-loop learning must be maximised if organisations are to make informed decisions in rapidly changing contexts. It is an approach which accords very well with the underlying values of knowledge management.

Governing variable

Action strategy

Consequences

**Double-loop learning**

**Single-loop learning**

Figure 5.4 Double-loop learning

**Case study**

A specific example of this corporate amnesia can be found at Ford, where new car developers wanted to replicate the success of the original Taurus design team. But no one remembered, or had recorded, what was so special about that effort... The assumption that technology can replace human knowledge or create its equivalent has proven false time and again.

Source: Davenport and Prusak (1998)

Business benefits of knowledge management

Several business benefits have been identified as accruing from knowledge management (Newing, 2000):

* identifying new markets from high-level intelligence gathering and pooling of knowledge by experts
* more responsiveness to market needs by harnessing external knowledge
* using customer knowledge to improve existing products and create innovative new ones
* faster time to market
* better quality products
* reusing knowledge gained in other parts of the world for other customers with similar problems
* continuous learning and development of best practice
* reducing costs associated with finding and reinventing knowledge by quickly retrieving explicit knowledge already stored
* improving customer service by applying knowledge at the point of first interaction with the customer
* reduction of risk by using wider expertise.

Of all the initiatives we’ve undertaken at Chevron during the 1990s, few have been as important or as rewarding as our efforts to build a learning organization by sharing and managing knowledge throughout our company.

Accepting the theory, and acknowledging the benefits, is a good starting point. But there can be considerable challenges and problems, which we will look at next.

Challenges and critical success factors

As a concept, knowledge management can involve some fundamental rethinking about the value of individual knowledge, and how the retention or sharing of knowledge by individuals is perceived and rewarded by the employing organisation. The argument runs: ‘If knowledge is power, why should I diminish (or eliminate) my power base by sharing it?’ This is one of many challenges that management faces in introducing knowledge management. Here we examine some of the key issues and the critical success factors.

**Barriers to knowledge management**

Even if your organisation has taken on board the message that using your corporate knowledge more intelligently can be a vital component in competing in the marketplace, it may well face a number of basic problems before it can get underway (Bonaventura, 1997). There may be, for example:

* no model for knowledge creation and dissemination within the organisation: you’ve never done it before so where do you start?
* no processes or systems focused on supporting those activities – they weren’t part of the original systems specification so where do they fit in now?
* no systems able to measure or evaluate how well you are creating and disseminating knowledge
* no means of evaluating the effectiveness of the knowledge creation and dissemination activities that you are carrying out.

Von Krogh et al. (2000) believe that managers ought to be supporting knowledge rather than trying to manage it, as it is basically unmanageable and not amenable to traditional management techniques. Individual staff may be reluctant to accept new lessons, insights and ideas, and many organisations can be quite challenging places for people learning to overcome the barriers of sharing knowledge with others. Individual barriers can include the following:

* People approach new experiences on the basis of their experience and beliefs about the world. There will be some situations which are so new and different that they will not have developed a response to them, and will find them too challenging.
* Some people will see new knowledge as a threat to their selfimage, and respond negatively to it.
* The organisation itself may contain its own barriers:
* New ideas will have to be made explicit in a ‘language’ that people in the organisation can understand
* The organisational memory and understanding of how things work can be good for bonding people together – but also make it more difficult for an individual to disagree with a ‘party line’
* Organisational procedures may make cross-functional interactions difficult.

**Key questions for management**

Bonaventura (1997) describes how, where there are no existing models on which knowledge creation processes and systems can be based, management will typically just issue a general call for more ‘learning’. This will fail, because the organisation will not have in place the reporting structures, compensation mechanisms or procedures that are necessary to support it, or because they don’t understand the knowledge creation and knowledge dissemination process, and think it can be managed just like everything else. Bonaventura puts forward some questions for the management of any organisation (particularly those in the knowledge intensive sectors of the global economy) to ask of itself:

What do our culture and our actions as managers say about the value of knowledge in the organisation?

How is knowledge created, embodied and disseminated? What is the relationship between knowledge and the kind of innovation that we need to achieve our objectives?

What commercial benefits do we expect to gain from more effective knowledge management?

Where are we in terms of the maturity of our knowledge systems?

What role does IT play in our knowledge management programme?

Skyrme (1998) has identified a number of recurring characteristics within organisations that demonstrate best practice in knowledge innovation, which also translate into key questions for management:

1. Is your knowledge strategy separate or clearly linked to your business strategy?
2. How much is knowledge discussed in your organisation, and how well is it understood? Is it a key element in your plans and budgets?
3. Is the knowledge facet of your business articulated as a real, compelling vision? Is there a framework that guides management decisions?
4. Are there knowledge champions throughout your business? Does your chief executive officer (CEO) link the importance of your organisational knowledge to your business success?
5. Do you have systematic processes for capturing, organising and sharing knowledge throughout your organisation?
6. Are people and information readily accessible through your computer and communication networks? Do these networks extend to customers, suppliers and experts?
7. Do you measure the contribution of knowledge to your organisation’s performance?
8. Only when an organisation has realistic answers to these questions can it start to develop real advantage from its knowledge assets.

**The employee perspective**

So much for management – but what about the staff? As Morling (2000) points out, asking or telling employees to ‘share knowledge’ is a waste of time. It may look good in the strategic plan, but won’t on its own satisfy the needs of staff to feel valued and recognised for their contribution. For an employee, the benefits of knowledge management relate less to organisational performance than to the general human desire for interaction with other people and some kind of shared interest or expertise within quite small groups, where opinions and ideas are freely exchanged, respected and trusted.

These small groups of shared interest or expertise (knowledge ‘communities’) can fulfil a useful role by creating a balance between the enterprise on the one hand and the individual on the other. A common mistake that management makes in trying to implement knowledge management is to focus on the individual as the source of knowledge – somebody who can deposit a ‘knowledge package’ of their expertise and experience into the system for use by all. This fails to recognise the role of collaboration in developing ideas, or that for many people the real reward of sharing knowledge is the immediate response from the people they have shared it with.

Morling’s company, has developed a system of ‘communities’, defined as:

...a group of people sharing a common interest or practice, whose purpose is to share knowledge and experiences, help each other to learn, act as a support network and sometimes as an informal centre of excellence.

The knowledge and skills gained in the community can be applied directly to the needs and issues of the organisation through the members’ formal organisational roles and participation in teams. For example, they can be used as a specific project team, or within a geographical sales division. The point is that they become a seeding ground for new ideas, competencies and skills. Here is a summary of the basic principles:

The seven Cs of communities

1. Context – the community sets the context for knowledge sharing and creation
2. Contribution – members know where and how to contribute
3. Creation – knowledge is created through the interaction of members
4. Collaboration – members collaborate to build on each other’s ideas
5. Consensus – members agree on best practices and how to take ideas forward
6. Content – knowledge is captured as reusable content
7. Capitalisation – the organisation can exploit new knowledge and best practices.

Communities do need a certain amount of nurturing to develop. If you’ve identified a potentially useful knowledge community within your organisation, here are some tips to get it working to good effect:

Identify a core group of people with the motivation and commitment to get it started

Make sure that the group has a leader who can manage the context of the community, so that when the context shifts the community can respond like a team would

Hold regular meetings

Enlist management support

Ensure access to appropriate technology

Give the group self-governance

Pay attention to team building, especially by helping new members of the community to feel welcome and comfortable and able to participate actively and constructively.

**Knowledge management in practice**

Theory is one thing, practice quite another. How do you get from a grand vision to operational reality? How do you start to mobilise the static knowledge held within your organisation? Some of the world’s most successful companies have adopted a knowledge management approach as the best way to maximise the value of their own knowledge assets and to build good strategic partnerships and customer relations. Many of them have been more than ready to share their experiences via documented case studies. Some of these are included in this section. Here we set out some practical stages in getting underway when setting up a knowledge management programme.

**Approaches to knowledge management**

Skyrme (1998) reports that in analysing the way that over 100 organisations apply knowledge management, two main approaches have been identified:

Sharing existing knowledge so that organisations don’t reinvent the wheel because the knowledge they need is available but not known to them. For example, a department of the US giant AT &T spent US$79,449 to obtain information that could have been found in a publicly available document from their associate company Bell Research for US$13.

Creating new knowledge and converting it into new products, services and processes, enabling better, faster innovation. Jaguar is one company that has focused on this. By codifying how the best engineers design a particular car body panel, engineers can develop detailed designs in hours rather than weeks. This meta-knowledge – analysing how professionals go about their work – is becoming an important facet of a firm’s knowledge that needs to be captured and shared.

Skyrme (1998) found that in both approaches, organisations tend to focus on a few knowledge ‘levers’ to strengthen their knowledge-building efforts:

Customer knowledge – develop deep knowledge-sharing relationships, and understand the needs of your customers’ customers

Stakeholder relationships – improve knowledge flows between suppliers, employees, shareholders

Business environment insights – systematically scan your political, economic environment, etc. and monitor what your competitors are doing

Organisational memory – share knowledge through best practice databases, directories of expertise, intranets

Knowledge in processes – embed knowledge into business processes and management decision making

Knowledge in products and services – surround products with knowledge, for example in user guides and knowledge-intensive services

Knowledge in people – set up innovation workshops, learning networks, communities of knowledge practice.

The knowledge created within an organisation must add value (Tissen et al. 1998). If the creation of knowledge is to be successfully directed, then the people involved in it must be too. Here are some examples:

Pharmaceuticals company Hoffman-LaRoche is using the approach of knowledge domains and knowledge links in order to reduce its time to market. It has calculated that each day gained in market availability represents a monetary gain of US$1m.

CIGNA Property & Casualty, an insurance company, has created an ‘upward value spiral’ for know-how to be shared through the company. Employees’ information and knowledge is processed by ‘knowledge editors’ and distributed throughout the organisation.

Case study

Procter & Gamble is a consumer-products giant with nearly 110,000 employees spread in locations across the world. Aware that its success depended in part on the knowledge locked in the minds of its employees, P&G used knowledge-sharing software to transform departmental experts into tangible information resources for the whole company. In particular, scientists and engineers working in numerous locations across the world could benefit from collaborating and sharing information and expertise.

The company’s intranet was found to be ‘doing a good job connecting people to knowledge that was documented and published, but not as good a job in connecting them to experts’, commented Mike Telljohann, associate director at P&G’s technical centre in Cincinnati. He explained that it was clear people did not know where to go with questions. They suspected that there was knowledge out there that they couldn’t access.

In response to this feedback the company introduced an integrated system from AskMe Enterprises that forms a directory listing of individuals noted as subject-matter experts who can be called on to lend advice or collaboration for problem solving and product development. It provided a single knowledge base in the company. But a key advantage of the system was that it was able to reward active participants – ‘the more active you are in a particular area, the software highlights you as a featured expert. People in the innovation area enjoy being seen as an expert – it gives a lot of personal satisfaction.’ Telljohann and his team spent a lot of time marketing the benefits of the system. But return on investment from the pilot project was enough to persuade the company to invest in large-scale implementation.

Telljohann sums up the benefits: ‘I think the experts feel like they can make more of an impact. They typically have close circles they share experiences and knowledge with; this broadens their ability to share what they know, and the people with questions have a place to go.’

Here are some other practical projects that organisations can undertake to improve their knowledge management:

Create knowledge databases of best practice, expertise, client profiles, legislative developments

Create a knowledge map (a visual representation of information and relationships)

Actively manage processes for collecting, classifying, storing and disseminating information

Develop knowledge centres that are focal points for specific knowledge, and knowledge webs – networks of experts

Introduce collaborative technologies like intranets or groupware

Appoint a senior executive to be responsible for the knowledge initiative.

**Steps to knowledge mobilisation**

To summarise, let’s try and pull together all the different aspects of making the most of knowledge assets. Paul Miller (1998) provides a good overview of the process which he calls ‘knowledge mobilisation’, shown in Figure 5.5.

**Start with a business strategy**

**Create a knowledge-sharing culture**

**Get the right structure**

**Create a dedicated team**

**Help people to feel secure**

**Reward those who share knowledge**

**Ensure commitment from the top**

**Capture soft knowledge**

**Handle technology with care**

**Maximise employee know-how through active learning**

**Persist and measure**

**Share with other companies and get a win-win strategy**

Figure 5.5 The 12 steps to knowledge mobilisation

Let’s look at these steps in more detail, with examples provided by Miller (1998) that illustrate how some organisations have put theory into practice:

1. Start with a business strategy – before you start changing things, you must understand the business strategy to which knowledge mobilisation can contribute.
2. Create a knowledge-sharing culture – easier said than done! But if your culture is one that says ‘knowledge is power’ and your pay and benefits system rewards the hoarding of knowledge, then the system has to change. Remploy, which specialises in employing disabled staff, has introduced critical paths to make its employees more proactive in knowledge sharing.
3. Get the right structure – flatter organisational structures (rather than many hierarchical levels) encourage knowledge sharing.
4. Create a dedicated team – changing culture is a major initiative and needs a dedicated team to push it forward. This team will raise the knowledge profile and publicise the organisation’s commitment to knowledge sharing. Ernst & Young recognised that it had cultural obstacles to overcome, arising from its traditional culture of not sharing knowledge in order to protect client confidentiality. It set up a large team of knowledge workers to drive initiatives forward and monitor the results.
5. Help people to feel secure – employees don’t willingly share knowledge if they feel their jobs are under threat, and organisations must acknowledge this.
6. Reward those who share knowledge – knowledge
7. contributions need to be recognised in the pay structure. At Boston Consulting Group, a part of each consultant’s compensation depends on knowledge mobilisation activities.
8. Ensure commitment from the top – the role of the CEO is essential to culture change. Employees need to see top management promoting knowledge sharing if they are expected to do so.
9. Capture soft knowledge – connect people with people, and create multilevel networks to capture soft knowledge. Sun Microsystems brings together its employees worldwide by job function on a regular basis to train them in developments in the company and the market.
10. Handle technology with care – technology can enable knowledge mobilisation, but it’s useless without the necessary culture.
11. Maximise employee know-how through active learning – cascade learning throughout the organisation – cross-functional knowledge sharing increases awareness of roles and responsibilities. Allen & Overy’s junior lawyers present their experiences and areas of expertise to new entrants; partners give presentations at other levels. There is a continuous process of learning.
12. Persist and measure – when people leave, they take their training and know-how with them. Organisations need to capture that knowledge.
13. Share with other companies and get a win-win strategy – be open to the idea of sharing non-sensitive information with other companies.

**Recommendation**

Define knowledge management and its relationship to learning processes

Knowledge management is a technique for codifying tacit knowledge, and for making it widely available in the organisation.

Despite being seen as a technology-based technique, knowledge management depends for its success on the ability of people in the organisation to acquire new knowledge through learning.

Double-loop learning supports dissemination of knowledge. It occurs when an error is detected and corrected in a way that involves modifying the organisation’s underlying norms, policies and objectives, enabling the organisation as a whole to benefit from the learning.

Identify the barriers to knowledge management

There are challenging barriers to knowledge management.

Individuals may see knowledge management as a threat to their powerbase or they may lack the experience or context that enables them to acquire new knowledge in a particular situation.

Organisations may not recognise the value of their knowledge assets or may not have the culture, language or processes in place to support the creation and dissemination of knowledge.

Identify the critical success factors in knowledge management

Knowledge management is a strategic process that needs to be aligned with other elements of business strategy, compensation mechanisms, the reporting structure and the technology processes that are necessary to capture and disseminate knowledge.

Collaboration through knowledge communities is an effective way to encourage sharing and development of personal knowledge.

Mobilise knowledge management in your organisation

# Semantic Web

http://graphdb.ontotext.com/documentation/free/introduction-to-semantic-web.html

The Semantic Web represents a broad range of ideas and technologies that attempt to bring meaning to the vast amount of information available via the Web. The intention is to provide information in a structured form so that it can be processed automatically by machines. The combination of structured data and inferencing can yield much information not explicitly stated.

The aim of the Semantic Web is to solve the most problematic issues that come with the growth of the non-semantic (HTML-based or similar) Web that results in a high level of human effort for finding, retrieving and exploiting information. For example, contemporary search engines are extremely fast, but tend to be very poor at producing relevant results. Of the thousands of matches typically returned, only a few point to truly relevant content and some of this content may be buried deep within the identified pages. Such issues dramatically reduce the value of the information discovered as well as the ability to automate the consumption of such data. Other problems related to classification and generalisation of identifiers further confuse the landscape.

The Semantic Web solves such issues by adopting unique identifiers for concepts and the relationships between them. These identifiers, called Universal Resource Identifiers (URIs) (a “resource” is any ‘thing’ or ‘concept’) are similar to Web page URLs, but do not necessarily identify documents from the Web. Their sole purpose is to uniquely identify objects or concepts and the relationships between them.

The use of URIs removes much of the ambiguity from information, but the Semantic Web goes further by allowing concepts to be associated with hierarchies of classifications, thus making it possible to infer new information based on an individual’s classification and relationship to other concepts. This is achieved by making use of ontologies – hierarchical structures of concepts – to classify individual concepts.

Resource Description Framework (RDF)

The World-Wide Web has grown rapidly and contains huge amounts of information that cannot be interpreted by machines. Machines cannot understand meaning, therefore they cannot understand Web content. For this reason, most attempts to retrieve some useful pieces of information from the Web require a high degree of user involvement – manually retrieving information from multiple sources (different Web pages), ‘digging’ through multiple search engine results (where useful pieces of data are often buried many pages deep), comparing differently structured result sets (most of them incomplete), and so on.

For the machine interpretation of semantic content to become possible, there are two prerequisites:

Every concept should be uniquely identified. (For example, if a particular person owns a web site, authors articles on other sites, gives an interview on another site and has profiles in a couple of social media sites such as Facebook and LinkedIn, then the occurrences of his name/identifier in all these places should be related to exact same identifier.)

There must be a unified system for conveying and interpreting meaning that all automated search agents and data storage applications should use.

One approach for attaching semantic information to Web content is to embed the necessary machine-processable information through the use of special meta-descriptors (meta-tagging) in addition to the existing meta-tags that mainly concern the layout.

Within these meta tags, the resources (the pieces of useful information) can be uniquely identified in the same manner in which Web pages are uniquely identified, i.e., by extending the existing URL system into something more universal – a URI (Uniform Resource Identifier). In addition, conventions can be devised, so that resources can be described in terms of properties and values (resources can have properties and properties have values). The concrete implementations of these conventions (or vocabularies) can be embedded into Web pages (through meta-descriptors again) thus effectively ‘telling’ the processing machines things like:

[resource] John Doe has a [property] web site which is [value] www.johndoesite.com

The Resource Description Framework (RDF) developed by the World Wide Web Consortium (W3C) makes possible the automated semantic processing of information, by structuring information using individual statements that consist of: Subject, Predicate, Object. Although frequently referred to as a ‘language’, RDF is mainly a data model. It is based on the idea that the things being described have properties, which have values, and that resources can be described by making statements. RDF prescribes how to make statements about resources, in particular, Web resources, in the form of subject-predicate-object expressions. The ‘John Doe’ example above is precisely this kind of statement. The statements are also referred to as triples, because they always have the subject-predicate-object structure.

The basic RDF components include statements, Uniform Resource Identifiers, properties, blank nodes and literals. They are discussed in the topics that follow.

Uniform Resource Identifiers (URIs)

A unique Uniform Resource Identifier (URI) is assigned to any resource or thing that needs to be described. Resources can be authors, books, publishers, places, people, hotels, goods, articles, search queries, and so on. In the Semantic Web, every resource has a URI. A URI can be a URL or some other kind of unique identifier. Unlike URLs, URIs do not necessarily enable access to the resource they describe, i.e, in most cases they do not represent actual web pages. For example, the string http://www.johndoesite.com/aboutme.htm, if used as a URL (Web link) is expected to take us to a Web page of the site providing information about the site owner, the person John Doe. The same string can however be used simply to identify that person on the Web (URI) irrespective of whether such a page exists or not.

Thus URI schemes can be used not only for Web locations, but also for such diverse objects as telephone numbers, ISBN numbers, and geographic locations. In general, we assume that a URI is the identifier of a resource and can be used as either the subject or the object of a statement. Once the subject is assigned a URI, it can be treated as a resource and further statements can be made about it.

This idea of using URIs to identify ‘things’ and the relations between them is important. This approach goes some way towards a global, unique naming scheme. The use of such a scheme greatly reduces the homonym problem that has plagued distributed data representation in the past.

Statements: Subject-Predicate-Object Triples

To make the information in the following sentence

“The web site www.johndoesite.com is created by John Doe.”

machine-accessible, it should be expressed in the form of an RDF statement, i.e., a subject-predicate-object triple:

“[subject] the web site www.johndoesite.com [predicate] has a creator [object] called John Doe.”

This statement emphasises the fact that in order to describe something, there has to be a way to name or identify a number of things:

the thing the statement describes (Web site “www.johndoesite.com”);

a specific property (“creator”) of the thing the statement describes;

the thing the statement says is the value of this property (who the owner is).

The respective RDF terms for the various parts of the statement are:

the subject is the URL “www.johndoesite.com”;

the predicate is the expression “has creator”;

the object is the name of the creator, which has the value “John Doe”.

Next, each member of the subject-predicate-object triple should be identified using its URI, e.g.:

the subject is http://www.johndoesite.com;

the predicate is http://purl.org/dc/elements/1.1/creator (this is according to a particular RDF Schema, namely, the Dublin Core Metadata Element Set);

the object is http://www.johndoesite.com/aboutme (which may not be an actual web page).

Note that in this version of the statement, instead of identifying the creator of the web site by the character string “John Doe”, we used a URI, namely http://www.johndoesite.com/aboutme. An advantage of using a URI is that the identification of the statement’s subject can be more precise, i.e., the creator of the page is neither the character string “John Doe”, nor any of the thousands of other people with that name, but the particular John Doe associated with this URI (whoever created the URI defines the association). Moreover, since there is a URI to refer to John Doe, he is now a full-fledged resource and additional information can be recorded about him simply by adding additional RDF statements with John’s URI as the subject.

What we basically have now is the logical formula P(x,y)

, where the binary predicate P relates the object x to the object y – we may also think of this formula as written in the form x,P,y

. In fact, RDF offers only binary predicates (properties). If more complex relationships are to be defined, this is done through sets of multiple RDF triples. Therefore, we can describe the statement as:

<http://www.johndoesite.com> <http://purl.org/dc/elements/1.1/creator> <http://www.johndoesite.com/aboutme>

There are several conventions for writing abbreviated RDF statements, as used in the RDF specifications themselves. This shorthand employs an XML qualified name (or QName) without angle brackets as an abbreviation for a full URI reference. A QName contains a prefix that has been assigned to a namespace URI, followed by a colon, and then a local name. The full URI reference is formed from the QName by appending the local name to the namespace URI assigned to the prefix. So, for example, if the QName prefix foo is assigned to the namespace URI http://example.com/somewhere/, then the QName “foo:bar” is a shorthand for the URI http://example.com/somewhere/bar.

In our example, we can define the namespace jds for http://www.johndoesite.com and use the Dublin Core Metadata namespace dc for http://purl.org/dc/elements/1.1/.

So, the shorthand form for the example statement is simply:

jds: dc:creator jds:aboutme

Objects of RDF statements can (and very often do) form the subjects of other statements leading to a graph-like representation of knowledge. Using this notation, a statement is represented by:

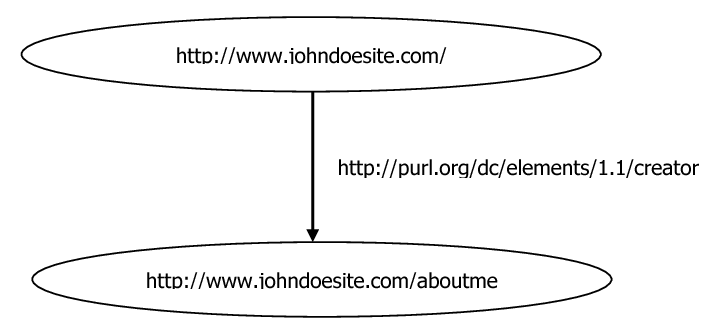
a node for the subject;

a node for the object;

an arc for the predicate, directed from the subject node to the object node.

So the RDF statement above could be represented by the following graph:

\_images/graphical\_triple.png



This kind of graph is known in the artificial intelligence community as a ‘semantic net’.

In order to represent RDF statements in a machine-processable way, RDF uses mark-up languages, namely (and almost exclusively) the Extensible Mark-up Language (XML). Because an abstract data model needs a concrete syntax in order to be represented and transmitted, RDF has been given a syntax in XML. As a result, it inherits the benefits associated with XML. However, it is important to understand that other syntactic representations of RDF, not based on XML, are also possible. XML-based syntax is not a necessary component of the RDF model. XML was designed to allow anyone to design their own document format and then write a document in that format. RDF defines a specific XML mark-up language, referred to as RDF/XML, for use in representing RDF information and for exchanging it between machines. Written in RDF/XML, our example will look as follows:

<?xml version="1.0" encoding="UTF-16"?>

<rdf:RDF

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:jds="http://www.johndoesite.com/">

<rdf:Description rdf:about="http://www.johndoesite.com/">

<dc:creator rdf:resource="jds:aboutme">

</rdf:Description>

</rdf:RDF>

Note

RDF/XML uses the namespace mechanism of XML, but in an expanded way. In XML, namespaces are only used for disambiguation purposes. In RDF/XML, external namespaces are expected to be RDF documents defining resources, which are then used in the importing RDF document. This mechanism allows the reuse of resources by other people who may decide to insert additional features into these resources. The result is the emergence of large, distributed collections of knowledge.

Also observe that the rdf:about attribute of the element rdf:Description is equivalent in meaning to that of an ID attribute, but it is often used to suggest that the object about which a statement is made has already been ‘defined’ elsewhere. Strictly speaking, a set of RDF statements together simply forms a large graph, relating things to other things through properties, and there is no such concept as ‘defining’ an object in one place and referring to it elsewhere. Nevertheless, in the serialised XML syntax, it is sometimes useful (if only for human readability) to suggest that one location in the XML serialisation is the ‘defining’ location, while other locations state ‘additional’ properties about an object that has been ‘defined’ elsewhere.

Properties

Properties are a special kind of resource: they describe relationships between resources, e.g., written by, age, title, and so on. Properties in RDF are also identified by URIs (in most cases, these are actual URLs). Therefore, properties themselves can be used as the subject in other statements, which allows for an expressive ways to describe properties, e.g., by defining property hierarchies.

Named graphs

A named graph (NG) is a set of triples named by a URI. This URI can then be used outside or within the graph to refer to it. The ability to name a graph allows separate graphs to be identified out of a large collection of statements and further allows statements to be made about graphs.

Named graphs represent an extension of the RDF data model, where quadruples <s,p,o,ng> are used to define statements in an RDF multi-graph. This mechanism allows, e.g., the handling of provenance when multiple RDF graphs are integrated into a single repository.

From the perspective of GraphDB, named graphs are important, because comprehensive support for SPARQL requires NG support.

RDF Schema (RDFS)

While being a universal model that lets users describe resources using their own vocabularies, RDF does not make assumptions about any particular application domain, nor does it define the semantics of any domain. It is up to the user to do so using an RDF Schema (RDFS) vocabulary.

RDF Schema is a vocabulary description language for describing properties and classes of RDF resources, with a semantics for generalisation hierarchies of such properties and classes. Be aware of the fact that the RDF Schema is conceptually different from the XML Schema, even though the common term schema suggests similarity. The XML Schema constrains the structure of XML documents, whereas the RDF Schema defines the vocabulary used in RDF data models. Thus, RDFS makes semantic information machine-accessible, in accordance with the Semantic Web vision. RDF Schema is a primitive ontology language. It offers certain modelling primitives with fixed meaning.

RDF Schema does not provide a vocabulary of application-specific classes. Instead, it provides the facilities needed to describe such classes and properties, and to indicate which classes and properties are expected to be used together (for example, to say that the property JobTitle will be used in describing a class “Person”). In other words, RDF Schema provides a type system for RDF.

The RDF Schema type system is similar in some respects to the type systems of object-oriented programming languages such as Java. For example, RDFS allows resources to be defined as instances of one or more classes. In addition, it allows classes to be organised in a hierarchical fashion. For example, a class Dog might be defined as a subclass of Mammal, which itself is a subclass of Animal, meaning that any resource that is in class Dog is also implicitly in class Animal as well.

RDF classes and properties, however, are in some respects very different from programming language types. RDF class and property descriptions do not create a straight-jacket into which information must be forced, but instead provide additional information about the RDF resources they describe.

The RDFS facilities are themselves provided in the form of an RDF vocabulary, i.e., as a specialised set of predefined RDF resources with their own special meanings. The resources in the RDFS vocabulary have URIs with the prefix http://www.w3.org/2000/01/rdf-schema# (conventionally associated with the namespace prefix rdfs). Vocabulary descriptions (schemas) written in the RDFS language are legal RDF graphs. Hence, systems processing RDF information that do not understand the additional RDFS vocabulary can still interpret a schema as a legal RDF graph consisting of various resources and properties. However, such a system will be oblivious to the additional built-in meaning of the RDFS terms. To understand these additional meanings, the software that processes RDF information has to be extended to include these language features and to interpret their meanings in the defined way.

Describing classes

A class can be thought of as a set of elements. Individual objects that belong to a class are referred to as instances of that class. A class in RDFS corresponds to the generic concept of a type or category similar to the notion of a class in object-oriented programming languages such as Java. RDF classes can be used to represent any category of objects such as web pages, people, document types, databases or abstract concepts. Classes are described using the RDF Schema resources rdfs:Class and rdfs:Resource, and the properties rdf:type and rdfs:subClassOf. The relationship between instances and classes in RDF is defined using rdf:type.

An important use of classes is to impose restrictions on what can be stated in an RDF document using the schema. In programming languages, typing is used to prevent incorrect use of objects (resources) and the same is true in RDF imposing a restriction on the objects to which the property can be applied. In logical terms, this is a restriction on the domain of the property.

Describing properties

In addition to describing the specific classes of things they want to describe, user communities also need to be able to describe specific properties that characterise these classes of things (such as numberOfBedrooms to describe an apartment). In RDFS, properties are described using the RDF class rdf:Property, and the RDFS properties rdfs:domain, rdfs:range and rdfs:subPropertyOf.

All properties in RDF are described as instances of class rdf:Property. So, a new property, such as exterms:weightInKg, is defined with the following RDF statement:

exterms:weightInKg rdf:type rdf:Property .

RDFS also provides vocabulary for describing how properties and classes are intended to be used together. The most important information of this kind is supplied by using the RDFS properties rdfs:range and rdfs:domain to further describe application-specific properties.

The rdfs:range property is used to indicate that the values of a particular property are members of a designated class. For example, to indicate that the property ex:author has values that are instances of class ex:Person, the following RDF statements are used:

ex:Person rdf:type rdfs:Class .

ex:author rdf:type rdf:Property .

ex:author rdfs:range ex:Person .

These statements indicate that ex:Person is a class, ex:author is a property, and that RDF statements using the ex:author property have instances of ex:Person as objects.

The rdfs:domain property is used to indicate that a particular property is used to describe a specific class of objects. For example, to indicate that the property ex:author applies to instances of class ex:Book, the following RDF statements are used:

ex:Book rdf:type rdfs:Class .

ex:author rdf:type rdf:Property .

ex:author rdfs:domain ex:Book .

These statements indicate that ex:Book is a class, ex:author is a property, and that RDF statements using the ex:author property have instances of ex:Book as subjects.

Sharing vocabularies

RDFS provides the means to create custom vocabularies. However, it is generally easier and better practice to use an existing vocabulary created by someone else who has already been describing a similar conceptual domain. Such publicly available vocabularies, called ‘shared vocabularies’, are not only cost-efficient to use, but they also promote the shared understanding of the described domains.

Considering the earlier example, in the statement:

jds: dc:creator jds:aboutme .

the predicate dc:creator, when fully expanded into a URI, is an unambiguous reference to the creator attribute in the Dublin Core metadata attribute set, a widely used set of attributes (properties) for describing information of this kind. So this triple is effectively saying that the relationship between the website (identified by http://www.johndoesite.com/) and the creator of the site (a distinct person, identified by http://www.johndoesite.com/aboutme) is exactly the property identified by http://purl.org/dc/elements/1.1/creator. This way, anyone familiar with the Dublin Core vocabulary or those who find out what dc:creator means (say, by looking up its definition on the Web) will know what is meant by this relationship. In addition, this shared understanding based upon using unique URIs for identifying concepts is exactly the requirement for creating computer systems that can automatically process structured information.

However, the use of URIs does not solve all identification problems, because different URIs can be created for referring to the same thing. For this reason, it is a good idea to have a preference towards using terms from existing vocabularies (such as the Dublin Core) where possible, rather than making up new terms that might overlap with those of some other vocabulary. Appropriate vocabularies for use in specific application areas are being developed all the time, but even so, the sharing of these vocabularies in a common ‘Web space’ provides the opportunity to identify and deal with any equivalent terminology.

Dublin Core Metadata Initiative

An example of a shared vocabulary that is readily available for reuse is The Dublin Core, which is a set of elements (properties) for describing documents (and hence, for recording metadata). The element set was originally developed at the March 1995 Metadata Workshop in Dublin, Ohio, USA. Dublin Core has subsequently been modified on the basis of later Dublin Core Metadata workshops and is currently maintained by the Dublin Core Metadata Initiative.

The goal of Dublin Core is to provide a minimal set of descriptive elements that facilitate the description and the automated indexing of document-like networked objects, in a manner similar to a library card catalogue. The Dublin Core metadata set is suitable for use by resource discovery tools on the Internet, such as Web crawlers employed by search engines. In addition, Dublin Core is meant to be sufficiently simple to be understood and used by the wide range of authors and casual publishers of information to the Internet.

Dublin Core elements have become widely used in documenting Internet resources (the Dublin Core creator element was used in the earlier examples). The current elements of Dublin Core contain definitions for properties such as title (a name given to a resource), creator (an entity primarily responsible for creating the content of the resource), date (a date associated with an event in the life-cycle of the resource) and type (the nature or genre of the content of the resource).

Information using Dublin Core elements may be represented in any suitable language (e.g., in HTML meta elements). However, RDF is an ideal representation for Dublin Core information. The following example uses Dublin Core by itself to describe an audio recording of a guide to growing rose bushes:

<rdf:RDF

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:dc="http://purl.org/dc/elements/1.1/">

<rdf:Description rdf:about="http://media.example.com/audio/guide.ra">

<dc:creator>Mr. Dan D. Lion</dc:creator>

<dc:title>A Guide to Growing Roses</dc:title>

<dc:description>Describes planting and nurturing rose bushes.

</dc:description>

<dc:date>2001-01-20</dc:date>

</rdf:Description>

</rdf:RDF>

The same RDF statements in Notation-3:

@prefix dc: <[http://purl.org/dc/elements/1.1/]> .

@prefix rdf: <[http://www.w3.org/1999/02/22-rdf-syntax-ns#]> .

<http://media.example.com/audio/guide.ra> dc:creator "Mr. Dan D. Lion" ;

dc:title "A Guide to Growing Roses" ;

dc:description "Describes planting and nurturing rose bushes." ;

dc:date "2001-01-20" .

Ontologies and knowledge bases

In general, an ontology formally describes a (usually finite) domain of related concepts (classes of objects) and their relationships. For example, in a company setting, staff members, managers, company products, offices, and departments might be some important concepts. The relationships typically include hierarchies of classes. A hierarchy specifies a class C to be a subclass of another class C' if every object in C is also included in C'. For example, all managers are staff members.

Apart from subclass relationships, ontologies may include information such as:

properties (X is subordinated Y);

value restrictions (only managers may head departments);

disjointness statements (managers and general employees are disjoint);

specifications of logical relationships between objects (every department must have at least three staff members).

Ontologies are important because semantic repositories use ontologies as semantic schemata. This makes automated reasoning about the data possible (and easy to implement) since the most essential relationships between the concepts are built into the ontology.

Formal knowledge representation (KR) is about building models. The typical modelling paradigm is mathematical logic, but there are also other approaches, rooted in the information and library science. KR is a very broad term; here we only refer to the mainstream meaning of the world (of a particular state of affairs, situation, domain or problem), which allow for automated reasoning and interpretation. Such models consist of ontologies defined in a formal language. Ontologies can be used to provide formal semantics (i.e., machine-interpretable meaning) to any sort of information: databases, catalogues, documents, Web pages, etc. Ontologies can be used as semantic frameworks: the association of information with ontologies makes such information much more amenable to machine processing and interpretation. This is because ontologies are described using logical formalisms, such as OWL, which allow automatic inferencing over these ontologies and datasets that use them, i.e., as a vocabulary. An important role of ontologies is to serve as schemata or ‘intelligent’ views over information resources. This is also the role of ontologies in the Semantic Web. Thus, they can be used for indexing, querying, and reference purposes over non-ontological datasets and systems such as databases, document and catalogue management systems. Because ontological languages have formal semantics, ontologies allow a wider interpretation of data, i.e., inference of facts, which are not explicitly stated. In this way, they can improve the interoperability and the efficiency of using arbitrary datasets.

An ontology O can be defined as comprising the 4-tuple.

O = <C,R,I,A>

where

C is a set of classes representing concepts from the domain we wish to describe (e.g., invoices, payments, products, prices, etc);

R is a set of relations (also referred to as properties or predicates) holding between (instances of) these classes (e.g., Product hasPrice Price);

I is a set of instances, where each instance can be a member of one or more classes and can be linked to other instances or to literal values (strings, numbers and other data-types) by relations (e.g., product23 compatibleWith product348 or product23 hasPrice €170);

A is a set of axioms (e.g., if a product has a price greater than €200, then shipping is free).

Classification of ontologies

Ontologies can be classified as light-weight or heavy-weight according to the complexity of the KR language and the extent to which it is used. Light-weight ontologies allow for more efficient and scalable reasoning, but do not possess the highly predictive (or restrictive) power of more powerful KR languages. Ontologies can be further differentiated according to the sort of conceptualisation that they formalise: upper-level ontologies model general knowledge, while domain and application ontologies represent knowledge about a specific domain (e.g., medicine or sport) or a type of application, e.g., knowledge management systems.

Finally, ontologies can be distinguished according to the sort of semantics being modelled and their intended usage. The major categories from this perspective are:

Schema-ontologies: ontologies that are close in purpose and nature to database and object-oriented schemata. They define classes of objects, their properties and relationships to objects of other classes. A typical use of such an ontology involves using it as a vocabulary for defining large sets of instances. In basic terms, a class in a schema ontology corresponds to a table in a relational database; a relation – to a column; an instance – to a row in the table for the corresponding class;

Topic-ontologies: taxonomies that define hierarchies of topics, subjects, categories, or designators. These have a wide range of applications related to classification of different things (entities, information resources, files, Web-pages, etc). The most popular examples are library classification systems and taxonomies, which are widely used in the knowledge management field. Yahoo and DMoz are popular large scale incarnations of this approach. A number of the most popular taxonomies are listed as encoding schemata in Dublin Core;

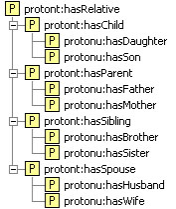
Lexical ontologies: lexicons with formal semantics that define lexical concepts. We use ‘lexical concept’ here as some kind of a formal representation of the meaning of a word or a phrase. In Wordnet, for example, lexical concepts are modelled as synsets (synonym sets), while word-sense is the relation between a word and a synset, word-senses and terms. These can be considered as semantic thesauri or dictionaries. The concepts defined in such ontologies are not instantiated, rather they are directly used for reference, e.g., for annotation of the corresponding terms in text. WordNet is the most popular general purpose (i.e., upper-level) lexical ontology.

Knowledge bases

Knowledge base (KB) is a broader term than ontology. Similar to an ontology, a KB is represented in a KR formalism, which allows automatic inference. It could include multiple axioms, definitions, rules, facts, statements, and any other primitives. In contrast to ontologies, however, KBs are not intended to represent a shared or consensual conceptualisation. Thus, ontologies are a specific sort of a KB. Many KBs can be split into ontology and instance data parts, in a way analogous to the splitting of schemata and concrete data in databases.

Proton

PROTON is a light-weight upper-level schema-ontology developed in the scope of the SEKT project, which we will use for ontology-related examples in this section. PROTON is encoded in OWL Lite and defines about 542 entity classes and 183 properties, providing good coverage of named entity types and concrete domains, i.e., modelling of concepts such as people, organisations, locations, numbers, dates, addresses, etc. A snapshot of the PROTON class hierarchy is shown below.



Logic and inference

The topics that follow take a closer look at the logic that underlies the retrieval and manipulation of semantic data and the kind of programming that supports it.

Logic programming

Logic programming involves the use of logic for computer programming, where the programmer uses a declarative language to assert statements and a reasoner or theorem-prover is used to solve problems. A reasoner can interpret sentences, such as IF A THEN B, as a means to prove B from A. In other words, given a collection of logical sentences, a reasoner will explore the solution space in order to find a path to justify the requested theory. For example, to determine the truth value of C given the following logical sentences

IF A AND B THEN C

B

IF D THEN A

D

a reasoner will interpret the IF..THEN statements as rules and determine that C is indeed inferred from the KB. This use of rules in logic programming has led to ‘rule-based reasoning’ and ‘logic programming’ becoming synonymous, although this is not strictly the case.

In LP, there are rules of logical inference that allow new (implicit) statements to be inferred from other (explicit) statements, with the guarantee that if the explicit statements are true, so are the implicit statements.

Because these rules of inference can be expressed in purely symbolic terms, applying them is the kind of symbol manipulation that can be carried out by a computer. This is what happens when a computer executes a logical program: it uses the rules of inference to derive new statements from the ones given in the program, until it finds one that expresses the solution to the problem that has been formulated. If the statements in the program are true, then so are the statements that the machine derives from them, and the answers it gives will be correct.

The program can give correct answers only if the following two conditions are met:

The program must contain only true statements;

The program must contain enough statements to allow solutions to be derived for all the problems that are of interest.

There must also be a reasonable time frame for the entire inference process. To this end, much research has been carried out to determine the complexity classes of various logical formalisms and reasoning strategies. Generally speaking, to reason with Web-scale quantities of data requires a low-complexity approach. A tractable solution is one whose algorithm requires finite time and space to complete.

Predicate logic

From a more abstract viewpoint, the subject of the previous topic is related to the foundation upon which logical programming resides, which is logic, particularly in the form of predicate logic (also known as ‘first order logic’). Some of the specific features of predicate logic render it very suitable for making inferences over the Semantic Web, namely:

It provides a high-level language in which knowledge can be expressed in a transparent way and with high expressive power;

It has a well-understood formal semantics, which assigns unambiguous meaning to logical statements;

There are proof systems that can automatically derive statements syntactically from a set of premises. These proof systems are both sound (meaning that all derived statements follow semantically from the premises) and complete (all logical consequences of the premises can be derived in the proof system);

It is possible to trace the proof that leads to a logical consequence. (This is because the proof system is sound and complete.) In this sense, the logic can provide explanations for answers.

The languages of RDF and OWL (Lite and DL) can be viewed as specialisations of predicate logic. One reason for such specialised languages to exist is that they provide a syntax that fits well with the intended use (in our case, Web languages based on tags). The other major reason is that they define reasonable subsets of logic. This is important because there is a trade-off between the expressive power and the computational complexity of certain logic: the more expressive the language, the less efficient (in the worst case) the corresponding proof systems. As previously stated, OWL Lite and OWL DL correspond roughly to description logic, a subset of predicate logic for which efficient proof systems exist.

Another subset of predicate logic with efficient proof systems comprises the so-called rule systems (also known as Horn logic or definite logic programs).

A rule has the form:

A1, ... , An → B

where Ai and B are atomic formulas. In fact, there are two intuitive ways of reading such a rule:

If A1, ... , An are known to be true, then B is also true. Rules with this interpretation are referred to as ‘deductive rules’.

If the conditions A1, ... , An are true, then carry out the action B. Rules with this interpretation are referred to as ‘reactive rules’.

Both approaches have important applications. The deductive approach, however, is more relevant for the purpose of retrieving and managing structured data. This is because it relates better to the possible queries that one can ask, as well as to the appropriate answers and their proofs.

Description logic

Description Logic (DL) has historically evolved from a combination of frame-based systems and predicate logic. Its main purpose is to overcome some of the problems with frame-based systems and to provide a clean and efficient formalism to represent knowledge. The main idea of DL is to describe the world in terms of ‘properties’ or ‘constraints’ that specific ‘individuals’ must satisfy. DL is based on the following basic entities:

Objects - Correspond to single ‘objects’ of the real world such as a specific person, a table or a telephone. The main properties of an object are that it can be distinguished from other objects and that it can be referred to by a name. DL objects correspond to the individual constants in predicate logic;

Concepts - Can be seen as ‘classes of objects’. Concepts have two functions: on one hand, they describe a set of objects and on the other, they determine properties of objects. For example, the class “table” is supposed to describe the set of all table objects in the universe. On the other hand, it also determines some properties of a table such as having legs and a flat horizontal surface that one can lay something on. DL concepts correspond to unary predicates in first order logic and to classes in frame-based systems;

Roles - Represent relationships between objects. For example, the role ‘lays on’ might define the relationship between a book and a table, where the book lays upon the table. Roles can also be applied to concepts. However, they do not describe the relationship between the classes (concepts), rather they describe the properties of the objects that are members of that classes;

Rules - In DL, rules take the form of “if condition x (left side), then property y (right side)” and form statements that read as “if an object satisfies the condition on the left side, then it has the properties of the right side”. So, for example, a rule can state something like ‘all objects that are male and have at least one child are fathers’.

The family of DL system consists of many members that differ mainly with respect to the constructs they provide. Not all of the constructs can be found in a single DL system.

The Web Ontology Language (OWL) and its dialects

In order to achieve the goal of a broad range of shared ontologies using vocabularies with expressiveness appropriate for each domain, the Semantic Web requires a scalable high-performance storage and reasoning infrastructure. The major challenge towards building such an infrastructure is the expressivity of the underlying standards: RDF, RDFS, OWL and OWL 2. Even though RDFS can be considered a simple KR language, it is already a challenging task to implement a repository for it, which provides performance and scalability comparable to those of relational database management systems (RDBMS). Even the simplest dialect of OWL (OWL Lite) is a description logic (DL) that does not scale due to reasoning complexity. Furthermore, the semantics of OWL Lite are incompatible with that of RDF(S).

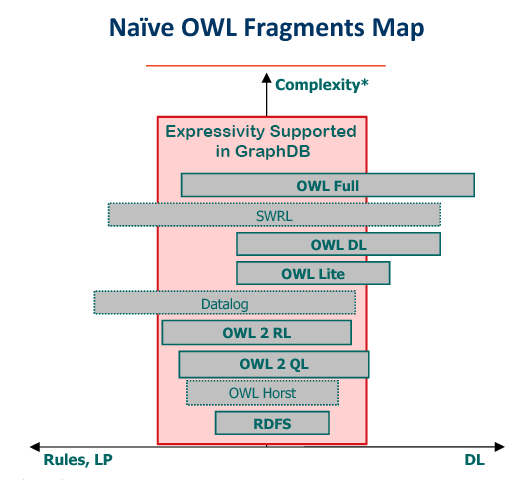


Figure 1 - OWL Layering Map

OWL DLP

OWL DLP is a non-standard dialect, offering a promising compromise between expressive power, efficient reasoning, and compatibility. It is defined as the intersection of the expressivity of OWL DL and logic programming. In fact, OWL DLP is defined as the most expressive sublanguage of OWL DL, which can be mapped to Datalog. OWL DLP is simpler than OWL Lite. The alignment of its semantics to RDFS is easier, as compared to OWL Lite and OWL DL dialects. Still, this can only be achieved through the enforcement of some additional modelling constraints and transformations.

Horn logic and description logic are orthogonal (in the sense that neither of them is a subset of the other). OWL DLP is the ‘intersection’ of Horn logic and OWL; it is the Horn-definable part of OWL, or stated another way, the OWL-definable part of Horn logic.

DLP has certain advantages:

From a modeller’s perspective, there is freedom to use either OWL or rules (and associated tools and methodologies) for modelling purposes, depending on the modeller’s experience and preferences.

From an implementation perspective, either description logic reasoners or deductive rule systems can be used. This feature provides extra flexibility and ensures interoperability with a variety of tools.

Experience with using OWL has shown that existing ontologies frequently use very few constructs outside the DLP language.

OWL Horst

In “Combining RDF and Part of OWL with Rules: Semantics, Decidability, Complexity” ter Horst defines RDFS extensions towards rule support and describes a fragment of OWL, more expressive than DLP. He introduces the notion of R-entailment of one (target) RDF graph from another (source) RDF graph on the basis of a set of entailment rules R. R-entailment is more general than the D-entailment used by Hayes in defining the standard RDFS semantics. Each rule has a set of premises, which conjunctively define the body of the rule. The premises are ‘extended’ RDF statements, where variables can take any of the three positions.

The head of the rule comprises one or more consequences, each of which is, again, an extended RDF statement. The consequences may not contain free variables, i.e., which are not used in the body of the rule. The consequences may contain blank nodes.

The extension of R-entailment (as compared to D-entailment) is that it ‘operates’ on top of so-called generalised RDF graphs, where blank nodes can appear as predicates. R-entailment rules without premises are used to declare axiomatic statements. Rules without consequences are used to detect inconsistencies.

In this document, we refer to this extension of RDFS as “OWL Horst”. This language has a number of important characteristics:

It is a proper (backward-compatible) extension of RDFS. In contrast to OWL DLP, it puts no constraints on the RDFS semantics. The widely discussed meta-classes (classes as instances of other classes) are not disallowed in OWL Horst. It also does not enforce the unique name assumption;

Unlike DL-based rule languages such as SWRL, R-entailment provides a formalism for rule extensions without DL-related constraints;

Its complexity is lower than SWRL and other approaches combining DL ontologies with rules.

In Figure 1, the pink box represents the range of expressivity of GraphDB, i.e., including OWL DLP, OWL Horst, OWL 2 RL, most of OWL Lite. However, none of the rulesets include support for the entailment of typed literals (D-entailment).

OWL Horst is close to what SWAD-Europe has intuitively described as OWL Tiny. The major difference is that OWL Tiny (like the fragment supported by GraphDB) does not support entailment over data types.

OWL2 RL

OWL 2 is a re-work of the OWL language family by the OWL working group. This work includes identifying fragments of the OWL 2 language that have desirable behavior for specific applications/environments.

The OWL 2 RL profile is aimed at applications that require scalable reasoning without sacrificing too much expressive power. It is designed to accommodate both OWL 2 applications that can trade the full expressivity of the language for efficiency, and RDF(S) applications that need some added expressivity from OWL 2. This is achieved by defining a syntactic subset of OWL 2, which is amenable to implementation using rule-based technologies, and presenting a partial axiomatisation of the OWL 2 RDF-Based Semantics in the form of first-order implications that can be used as the basis for such an implementation. The design of OWL 2 RL was inspired by Description Logic Programs and pD.

OWL Lite

The original OWL specification, now known as OWL 1, provides two specific subsets of OWL Full designed to be of use to implementers and language users. The OWL Lite subset was designed for easy implementation and to offer users a functional subset that provides an easy way to start using OWL.

OWL Lite is a sublanguage of OWL DL that supports only a subset of the OWL language constructs. OWL Lite is particularly targeted at tool builders, who want to support OWL, but who want to start with a relatively simple basic set of language features. OWL Lite abides by the same semantic restrictions as OWL DL, allowing reasoning engines to guarantee certain desirable properties.

OWL DL

The OWL DL (where DL stands for Description Logic) subset was designed to support the existing Description Logic business segment and to provide a language subset that has desirable computational properties for reasoning systems.

OWL Full and OWL DL support the same set of OWL language constructs. Their difference lies in the restrictions on the use of some of these features and on the use of RDF features. OWL Full allows free mixing of OWL with RDF Schema and, like RDF Schema, does not enforce a strict separation of classes, properties, individuals and data values. OWL DL puts constraints on mixing with RDF and requires disjointness of classes, properties, individuals and data values. The main reason for having the OWL DL sublanguage is that tool builders have developed powerful reasoning systems that support ontologies constrained by the restrictions required for OWL DL.

Query languages

In this section, we introduce some query languages for RDF. This may beg the question why we need RDF-specific query languages at all instead of using an XML query language. The answer is that XML is located at a lower level of abstraction than RDF. This fact would lead to complications if we were querying RDF documents with an XML-based language. The RDF query languages explicitly capture the RDF semantics in the language itself.

All the query languages discussed below have an SQL-like syntax, but there are also a few non-SQL-like languages like Versa and Adenine.

The query languages supported by RDF4J (which is the Java framework within which GraphDB operates) and therefore by GraphDB, are SPARQL and SeRQL.

RQL, RDQL

RQL (RDF Query Language) was initially developed by the Institute of Computer Science at Heraklion, Greece, in the context of the European IST project MESMUSES.3. RQL adopts the syntax of OQL (a query language standard for object-oriented databases), and, like OQL, is defined by means of a set of core queries, a set of basic filters, and a way to build new queries through functional composition and iterators.

The core queries are the basic building blocks of RQL, which give access to the RDFS-specific contents of an RDF triplestore. RQL allows queries such as Class (retrieving all classes), Property (retrieving all properties) or Employee (returning all instances of the class with name Employee). This last query, of course, also returns all instances of subclasses of Employee, as these are also instances of the class Employee by virtue of the semantics of RDFS.

RDQL (RDF Data Query Language) is a query language for RDF first developed for Jena models. RDQL is an implementation of the SquishQL RDF query language, which itself is derived from rdfDB. This class of query languages regards RDF as triple data, without schema or ontology information unless explicitly included in the RDF source.

Apart from RDF4J, the following systems currently provide RDQL (all these implementations are known to derive from the original grammar): Jena, RDFStore, PHP XML Classes, 3Store and RAP (RDF API for PHP).

SPARQL

SPARQL (pronounced “sparkle”) is currently the most popular RDF query language; its name is a recursive acronym that stands for “SPARQL Protocol and RDF Query Language”. It was standardised by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is now considered a key Semantic Web technology. On 15 January 2008, SPARQL became an official W3C Recommendation.

SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns. Several SPARQL implementations for multiple programming languages exist at present.

SeRQL

SeRQL (Sesame RDF Query Language, pronounced “circle”) is an RDF/RDFS query language developed by Sesame’s developer - Aduna - as part of Sesame (now RDF4J). It selectively combines the best features (considered by its creators) of other query languages (RQL, RDQL, N-Triples, N3) and adds some features of its own. As of this writing, SeRQL provides advanced features not yet available in SPARQL. Some of SeRQL’s most important features are:

Graph transformation;

RDF Schema support;

XML Schema data-type support;

Expressive path expression syntax;

Optional path matching.

Reasoning strategies

There are two principle strategies for rule-based inference: Forward-chaining and Backward-chaining:

Forward-chaining

to start from the known facts (the explicit statements) and to perform inference in a deductive fashion. Forward-chaining involves applying the inference rules to the known facts (explicit statements) to generate new facts. The rules can then be re-applied to the combination of original facts and inferred facts to produce more new facts. The process is iterative and continues until no new facts can be generated. The goals of such reasoning can have diverse objectives, e.g., to compute the inferred closure, to answer a particular query, to infer a particular sort of knowledge (e.g., the class taxonomy), etc.

Advantages: When all inferences have been computed query answering can proceed extremely quickly.

Disadvantages: Initialisation costs (inference computed at load time) and space/memory usage (especially when the number of inferred facts is very large).

Backward-chaining

involves starting with a fact to be proved or a query to be answered. Typically, the reasoner examines the knowledge base to see if the fact to be proved is present and if not it examines the ruleset to see which rules could be used to prove it. For the latter case, a check is made to see what other ‘supporting’ facts would need to be present to ‘fire’ these rules. The reasoner searches for proofs of each of these ‘supporting’ facts in the same way and iteratively maps out a search tree. The process terminates when either all of the leaves of the tree have proofs or no new candidate solutions can be found. Query processing is similar, but only stops when all search paths have been explored. The purpose in query answering is to find not just one but all possible substitutions in the query expression.

Advantages: There are no inferencing costs at start-up and minimal space requirements.

Disadvantages: Inference must be done each and every time a query is answered and for complex search graphs this can be computationally expensive and slow.

As both strategies have advantages and disadvantages, attempts to overcome their weak points have led to the development of various hybrid strategies (involving partial forward- and backward-chaining), which have proven efficient in many contexts.

Total materialisation

Imagine a repository that performs total forward-chaining, i.e., it tries to make sure that after each update to the KB, the inferred closure is computed and made available for query evaluation or retrieval. This strategy is generally known as materialisation. In order to avoid ambiguity with various partial materialisation approaches, let us call such an inference strategy, taken together with the monotonic entailment. When new explicit facts (statements) are added to a KB (repository), new implicit facts will likely be inferred. Under a monotonic logic, adding new explicit statements will never cause previously inferred statements to be retracted. In other words, the addition of new facts can only monotonically extend the inferred closure. Assumption, total materialisation.

Advantages and disadvantages of the total materialisation:

Upload/store/addition of new facts is relatively slow, because the repository is extending the inferred closure after each transaction. In fact, all the reasoning is performed during the upload;

Deletion of facts is also slow, because the repository should remove from the inferred closure all the facts that can no longer be proved;

The maintenance of the inferred closure usually requires considerable additional space (RAM, disk, or both, depending on the implementation);

Query and retrieval are fast, because no deduction, satisfiability checking, or other sorts of reasoning are required. The evaluation of queries becomes computationally comparable to the same task for relation database management systems (RDBMS).

Probably the most important advantage of the inductive systems, based on total materialisation, is that they can easily benefit from RDBMS-like query optimisation techniques, as long as all the data is available at query time. The latter makes it possible for the query evaluation engine to use statistics and other means in order to make ‘educated’ guesses about the ‘cost’ and the ‘selectivity’ of a particular constraint. These optimisations are much more complex in the case of deductive query evaluation.

Total materialisation is adopted as the reasoning strategy in a number of popular Semantic Web repositories, including some of the standard configurations of RDF4J and Jena. Based on publicly available evaluation data, it is also the only strategy that allows scalable reasoning in the range of a billion of triples; such results are published by BBN (for DAML DB) and ORACLE (for RDF support in ORACLE 11g).

Semantic repositories

Over the last decade, the Semantic Web has emerged as an area where semantic repositories became as important as HTTP servers are today. This perspective boosted the development, under W3C driven community processes, of a number of robust metadata and ontology standards. These standards play the role, which SQL had for the development and spread of the relational DBMS. Although designed for the Semantic Web, these standards face increasing acceptance in areas such as Enterprise Application Integration and Life Sciences.

In this document, the term ‘semantic repository’ is used to refer to a system for storage, querying, and management of structured data with respect to ontologies. At present, there is no single well-established term for such engines. Weak synonyms are: reasoner, ontology server, metastore, semantic/triple/RDF store, database, repository, knowledge base. The different wording usually reflects a somewhat different approach to implementation, performance, intended application, etc. Introducing the term ‘semantic repository’ is an attempt to convey the core functionality offered by most of these tools. Semantic repositories can be used as a replacement for database management systems (DBMS), offering easier integration of diverse data and more analytical power. In a nutshell, a semantic repository can dynamically interpret metadata schemata and ontologies, which define the structure and the semantics related to the data and the queries. Compared to the approach taken in a relational DBMS, this allows for easier changing and combining of data schemata and automated interpretation of the data

# References

Argyris, C. and Schön, D. (1974) Theory in Practice: increasing professional effectiveness, Jossey-Bass

Argyris, C. (1999) On Organisational Learning, Blackwell Publishers

Bawden, D., Holtham, C. and Courtney, N. (1999) ‘Perspectives on information overload’, Aslib Proceedings, Vol. 51, No. 8, 249–255

Bonaventura, M. (1997) ‘The benefits of a knowledge culture’, Aslib Proceedings, Vol. 49, No. 4, 82–89

Bradley, P. (1999) ‘Efficiency in Numbers’, Information World Review,

May

British Standards Institution (1995) Code of Practice for Information Security Management (BS7799)

Buckley, P. and Clark, D. (2004) A Rough Guide to the Internet, Rough guides

Clifton, H. D. (1990) Business data systems, Prentice Hall

Cobham, D. and Curtis, G. (2004) Business Information Systems:

Analysis, Design and Practice, FT Prentice Hall

Cooke, A. (1999) ‘Neal-Schuman authoritative guide to evaluating information on the Internet’, Searcher: the Magazine for Database Professionals, June, pp.22–31

Data Protection Act 1998, The Stationery Office Ltd, www.hmso.gov.uk/acts/acts1998/19980029.htm

Davenport, T. H. and Prusak, L. (1998) Working knowledge: how organisations manage what they know, Harvard Business School

Dawson, A. (1997) 2nd edition, The Internet for Library and

Information Service Professionals, Aslib

Derr, K. quoted in The Knowledge Management Resource Center site at www.kmresource.com/exp\_cases.htm

Edmunds, A. and Morris, A. (2000) ‘The problem of information overload in business organisations: a review of the literature’, International Journal of Information Management, 20, 17–28

European Forum for Electronic Business, www.eema.org

Fishenden, J. (1997) ‘Managing intranets to improve business process’, Aslib Proceedings, Vol. 49, 4, April, pp.90–96

Galliers, R. D., Leidner, D. E. and Baker, B. S. H. (1999) 2nd edition, Strategic information management: challenges and strategies in managing information systems, Butterworth-Heinemann

Gorelick, C., April, K. and Milton, N. (2003) Performance Through

Learning: Knowledge Management in Practice, Butterworth-

Heinemann

Hawkins, S., Yen, D. C. and Chou, D. C. (2000) ‘Awareness and challenges of Internet security’, Information Management and Computer Security, Vol. 8, No. 3, 131–143

Horton, F. W. and Lewis, D. (1991) Great information disasters, Aslib

Information Management Report (2000) Bowker, May, 4

Informix (1999) ‘Executive Summary’, www.informix.com

Jay, R. and Jay, A. (2000) Effective presentation, Prentice Hall

Kolb, D. A. (1985) Learning styles inventory, McBer & Co

KPMG (2001) Global e-fr@ud Survey, www.kpmg.co.uk

Leigh, A. (1997) Persuasive reports and proposals, Chartered Institute of Personnel and Development

Lytle, R. (1991) ‘The PPS information systems development disaster in the early 1980s’, in Horton, F. and Lewis, D. (eds.) Great information disasters, Aslib

Mason, D. and Willcocks, L. (1994) Systems analysis, systems design,

Alfred Waller Ltd

McEwan, I. (1997) Enduring Love, Vintage

MelissaVirus.com, www.melissavirus.com

Miller, P. (1998) Mobilising the power of what you know: a practical guide to successful knowledge management, Century Ltd

Mitchell, R. C., Marcella, R. and Baxter, G (1999) ‘Corporate

Information Security Management’, New Library World, Vol. 100,

No. 1150, 213–227

Moore, C. (2001) ‘Tapping Knowledge’, 12 October, www.infoworld.com

Morling, J. (2000) ‘Share or else’, Consultants’ Advisory, December,

30–32

Newing, R. (2000) ‘Knowledge economy’, Consultant’s Adviser, Iss. 3, 2–4

Nickerson, R. C. (2001) 2nd edition, Business and information systems, Prentice Hall

Nonaka, I. and Takeuchi, H. (1995) The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford

University Press

Reuters (1996) Dying for information, Reuters Business Information

Shim, J. K. (2000) Information systems and technology for the noninformation systems executive, St Lucie Press

Simpson, C. W. and Prusak, L. (1995) ‘Troubles with information overload’, International Journal of Information Management, Vol. 15,

No. 6, 413–425

Skyrme, D. (1998) ‘Developing a knowledge strategy’, Strategy,

January, 18–19

Skyrme, D. (2000) ‘Knowledge Management: making it work’, www.skyrme.com

Skyrme, D. ‘Knowledge Management: making sense of an oxymoron’, www.skyrme.com

Smith, M. K. (2001) ‘Chris Argyris: Theories of action, double-loop learning and organizational learning’, The encyclopedia of informal learning, www.infed.org/thinkers/argyris.htm,

Strunk, W. and White, E. (1999) The Elements of Style Allyn & Bacon

Swan, W., Langford, N., Watson, I. and Varey, R. J. (2000) ‘Viewing the corporate community as a knowledge network’, Corporate Communications, Vol. 5, Pt. 2, 97–106

Tissen, R., Andriessen, D. and Deprez, F. L. (1998) Value-based knowledge management, Addison Wesley Longman

Turban, E., Lee, J., King, D and Chung, H. M. (2000) Electronic commerce: a managerial perspective, Prentice Hall

University of Cambridge Computing Service (1998) Technical User

Services, 23 October, www.tus.csx.cam.ac.uk/videoconf/vctips.html

Von Krogh, G., Ichijo, K. and Nonaka, I. (2000) Enabling knowledge creation: how to unlock the mystery of tacit knowledge and release the power of innovation, Oxford University Press

Walters, L. (2002) Secrets of successful speakers, McGraw-Hill

Wenger, E., McDermott, R. and Snyder, W. (2002), Cultivating

Communities of Practice, Harvard Business School Press

Willard, N. (1999) ‘Knowledge Management: foundations for a secure structure’, Managing Information, June, 45–49

Wilson, D. (2002) Managing Information: IT for Business Process,

Butterworth-Heinemann