Chapter 14

Dissemination

The previous chapter looked at the problems associated with the presentation of incident reports. It was argued that the format and structure of these documents must be tailored so that they support their intended recipients. It was also argued that care must be taken to ensure that the rhetoric is not used to mask potential bias in an incident report. This chapter goes on to examine the problems that are associated with the dissemination of these documents. It is of little benefit ensuring that reports meet presentation guidelines if their intended recipients cannot access the information that they provide. There are significant problems associated with such dissemination activities. For example, the FDA's Medical Bulletin that presents information about their MedWatch program is currently distributed to 1.2 million health professionals. Later sections analyse the ways in which many organisations are using electronic media to support the dissemination of incident reports. This approach offers many advantages. In particular, the development of the Internet and Web-based tools ensures that information can be rapidly transmitted to potential readers across the globe. There are, however, numerous problems. It can be difficult to ensure the security and integrity of information that is distributed in this way. It can also be difficult to help investigators search through the many thousands of incidents that are currently being collected in many of these systems. The closing sections of this chapter present a range of techniques that are intended to address these potential problems.

14.1 Problems of Dissemination

Chapters 11.5 and 12.4 have already described some of the problems that complicate the dissemination of information about adverse occurrences and near miss incidents. For example, it can be difficult to ensure that information is made available in a prompt and timely fashion so that potential recurrences are avoided. It can also be difficult to ensure that safety recommendations reach all of the many different groups that might make use of this information. The following pages build on these previous chapters to analyse these barriers to dissemination in greater detail.

14.1.1 Number and Range of Reports Published

It is important not to underestimate the scale of the task that can be involved in the dissemination of incident reports. Even relatively small, local systems can generate significant amounts of information. For instance, one of the Intensive care Units that we have studied generated a total of 111 recommendations between August 1995 and November 1998. 82 of these were 'Remind Staff' statements. The 29 other recommendations concerned the creation of new procedures or changes to existing protocols (e.g. 'produce guidelines for care of arterial lines'), or were equipment related (e.g. 'Obtain spare helium cylinder for aortic pump to be kept in ICU').

As one might expect the task of keeping staff and management informed of recent incidents and recommendations is significantly more complex in national and international systems. This is illustrated by Table 14.1, which presents the total number of different reports that were published by

	1995	1996	1997	1998	1999	2000	2001 (-Aug)
Hazard Notices	6	12	16	2	6	13	2
Device Bulletins	5	7	4	6	3	5	4
Safety Notices	33	40	20	43	41	28	24
Device Alerts	-	-	-	-	-	8	5
Advice Notices	-	-	-	-	6	1	0
Pacemaker Notes	6	6	3	4	7	4	4
Total Reports	50	65	43	55	63	59	39
Total Incidents	4,298	4,330	5,852	6,125	6,860	7,352	-

Table 14.1: Annual Frequency of Publications by the UK MDA

the UK Medical Devices Agency (MDA) over the last five years. It should be noted that the figures for 2001 are currently only available until August. As we have seen from the Maritime examples in the previous chapter, incident reporting agencies produce a range of different publications to disseminate their recommendations. In Table 14.1, hazard notices are published following death or serious injury or where death or serious injury might have occurred [543]. A medical device must also be clearly implicated and immediate action must be necessary to prevent recurrence. Device bulletins address more general management interests. They are derived from adverse incident investigations and consultation with manufacturers or users. They are also informed by device evaluations. In contrast, safety notices are triggered by less 'serious' incidents. Their are published in circumstances where the recipients' actions can improve safety or where it is necessary to repeat warnings about previous problems. Device alerts are issued if there is the potential for death or serious injury particularly through the long term use of a medical device. Finally, Pacemaker Technical Notes publish information about implantable pacemakers, defibrillators and their associated accessories. For the purpose of comparison, Table 14.1 also contains the total number of adverse incidents that were reported in the MDA's annual reports [540]. As can be seen, there has been a gradual rise in the frequency of incident reports while the total number of publications has remained relatively stable. Such an analysis must, however, be qualified. The total incident frequencies are based on the MDA's reporting year. Hence the figure cited for 1996 is, in fact, that given for 1996-1997. However, the number of reports and associated publications provides some indication of the scale of the publication tasks that confront organisations such as the MDA.

	1997	1998	1999	2000	2001 (-Aug)
Safety Alerts	55	55	54	67	38
Drug Labeling	239	519	512	505	241
Biologics Safety	-	24	10	29	14
Food and Applied Nutrition	3	2	2	2	2
Devices and Radiology	9	12	9	4	3

Table 14.2: Annual Frequency of Publications by the US FDA's MedWatch Program

The level of activity indicated in Table 14.1 is mirrored by the figures in Table 14.2. This presents publication figures for the US Food and Drug Administration's MedWatch initiative. This Safety Information and Adverse Event Reporting Programme is intended to 'serves both healthcare professionals and the medical product-using public' [272]. It covers a braid range of medical products, 'including prescription and over-the-counter drugs, biologics, dietary supplements, and medical devices'. It, therefore, has a slightly wider remit than that of the UK MDA. The primary MedWatch publication provides Safety Alerts about drugs, biologics, devices and dietary supplements. As can be seen in Table 14.2, the MedWatch programme also publishes information from several different groups within the FDA. It publishes safety-related drug labeling change summaries that have been approved by FDA Center for Drug Evaluation and Research. It also incorporates recalls, withdrawals and safety issues identified by the Center for Biologics Evaluation and Research. The program also publishes selected warnings and other safety information identified by the Center for Food Safety and Applied Nutrition. Finally, the Medwatch initiative incorporates safety alerts, public health advisories and notices from the Center for Devices and Radiological Health. We have not calculated totals for Table 14.2 as we did for Table 14.1 because of the inherent difficulty of calculating the frequency of recommendations to change drug labelling in the FDA adverse event reporting programme. Some drugs form the focus of several reports in the same year. Recommendations can be applied to a particular generic named product or to the different brands of that product. We have chosen to calculate frequencies on the basis of named drugs identified in the Center for Drug Evaluation and Research warnings.

It is also important to stress that the reports identified in Tables 14.1 and 14.2 only represent a small subset of the publications that the FDA and the MDA publish in response to adverse incidents. For instance, the MedWatch programme also disseminates articles that are intended to support the continuing education of Healthcare professionals. These include information about the post-marketing Surveillance for Adverse Events After Vaccination and techniques for assuring drug quality. The FDA also provides more consumer oriented publications to encourage contributions from the general public. It uses incident information to address specific consumer concerns, for instance in special reports on drug development and the interaction between food and drug. The wide scope of these publication activities is also illustrated by the User Facility Reporting Bulletins. This quarterly publication is specifically targeted at hospitals, nursing homes and 'device user facilities'.

Dissemination activities are not only focussed on the generation of specific incident reports or articles on more general issues. They also include the organisation of workshops, such as the 1998 meeting on 'Minimising Medical Product Errors' [263]. The UK MDA host similar events, such as their annual conference which in 2001 will address the theme 'Protecting Patients - Maintaining Standards' [546]. The MDA also holds more focussed study days. These provide staff training to address common problems with particular devices. For example, the MDA set up a recent meeting for nurses on best practice and the potential pitfalls in operating infusion systems [544].

14.1.2 Tight Deadlines and Limited Resources

The previous paragraphs illustrate the high frequency and the diversity of dissemination activities that are conduced by many incident reporting organisations. It is difficult to under-estimate the logistical challenges that such activities can impose upon finite resources. There is also increasing pressure in many sectors to increase the efficiency of many reporting bodies. For instance, one government measure estimates that the MDA managed to increase its output by 9% with a stable workforce between 2000-2001. These pressures can also be illustrated by some of the objectives being promoted by the MDA. For 2001-2002, it is intended that all Hazard Notices will be issued within 20 working days; 90% of Safety Notices will be issued within 60 days and 75% within 50 days. It is also intended to increase the number of adverse incident reports that will be published by a further 9% while at the same time making 'efficiency' savings of 2% [540].

The results of tight financial constraints can also be seen in the manner in which the FDA has altered it's publication policy in recent years [868]. Previous sections have mentioned the User Facility Reporting Bulletin, this publication is intended for hospitals, nursing homes and other end-user facilities. The initial twenty, quarterly issues of the Bulletin were printed in the conventional manner and were posted to any organisation that requested a copy. At its peak, 77,000 subscribers received copies of these documents that presented summarised reports based on recent incident reports. Budgetary restrictions forced the FDA to review this policy. In Issue 17, readers were asked to respond to a retention notice. If they did not respond then they were removed from the distribution list. It was hoped that the high initial administrative overhead associated with this initiative would yield longer term savings in distribution costs. By 1999, however, Federal funding cuts prevented any distribution in paper form. The twenty-first issue of the Bulletin was, therefore, distributed through electronic means including an automated Fax system. The FDA summarised their feelings about this situation; 'we regret the need to move to this new technology if it means that many of our current readers will no longer have access to the Bulletin' [868].

The joint pressures imposed by the need to disseminate safety information in a timely fashion and the need to meet tight financial objectives has resulted in a number of innovations in incident reporting. Many of these systems start with the premise that it is impossible to elicit and analyse voluntary incident reports across an entire industry. Even with mandatory reporting systems there will be problems of contribution bias that result in a very partial view of safety-related incidents. These problems stem partly from the cost and complexity of large scale voluntary reporting systems. They also stem from the passive nature of most mandatory systems that simply expect contributors to meet their regulatory requirements when responding to an adverse occurrence. As we have seen, even if a potential contributor wants to meet a reporting obligation they may fail to recognise that a safety-related event has occurred. Most regulatory and investigatory organisations lack the resources necessary to train personnel across an industry to distinguish accurately between reportable and non-reportable events. Similarly, there are significant financial barriers that prevent routine inspections and audits to review compliance.

Sentinel reporting systems provide an alternative solution that is intended to reduce the costs associated with incident reporting and, thereby ensure that recommendations are disseminated in a timely fashion. This approach identifies a sample of all of the facilities to be monitored. This sites within the sample are then offered specialist training in both mandatory and voluntary incident reporting. The incidents that are reported by these sentinel sites can then form a focus for more general safety initiatives across an industry. These ideas are extremely suasive to many governmental organisations. For instance, the FDA Modernisation Act (1997) required that the FDA make a report to Congress in late 1999 about progress towards such a sentinel system [264]. In September 1996, CODA Inc. was awarded a contract to conduct a study to evaluate the feasibility and effectiveness of a sentinel reporting system for adverse event reporting of medical device use. The explicit intention was to determine not whether a sentinel system could supplement passive, voluntary systems, such as MedWatch, but to replace them entirely. The trial ran for twelve months and the final report emphasised the importance of feedback and dissemination in the success of any sentinel system. The CODA trial provide several different forms of feedback. These included a newsletter, faxes of safety notices, responses to questions presented by Study Coordinators. The individual reports that were received by the project were summarised, anonymised and then published in bimonthly newsletters for Study Coordinators. These coordinators acted as an efficient means of disseminating safety related information within the sample sites.

This project not has important implications for the efficient dissemination of safety-related information. It also provides important insights about the practical problems that can arise when attempting to ensure the timely dissemination of incident recommendations and reports. Many reporting systems endeavour to ensure that operators, safety managers and regulators are provided with information about incidents according to a sliding timescale that reflects the perceived seriousness of the incident. This is the case with the MDA targets, cited above. It can, however, be extremely difficult to estimate the seriousness of an incident. Previous chapters have referred to the 'worst plausible outcome' that is often invoked to support such assessments. The practical problems of applying such heuristics can be illustrated by the CODA pilot study. The project analysts determined that only 14% of the reports received would have been clearly covered by the existing mandatory systems. 56% of the reports described less serious incidents that fell within the voluntary reporting provisions. 30% of all submission, or 96 reports, fell between these two categories; 'the determination of serious patient injury according to FDA's definition was difficult to make'. Of these 96, 60% were submitted on voluntary forms. 25% of the reports clearly documenting serious patient injury also were submitted on voluntary forms. If these results provide an accurate impression of the true severity of the incidents then they indicate that analysts cannot accept the contributors' severity assessments at face value. Two senior nurse-analysts agreed to review all reports and classify them urgency using a scale of: very urgent, urgent, routine monitoring, well-known problem or not important. Approximately one-third (113) were classified as very urgent or urgent. Of these, only 19 were clearly mandatory reports. This is a significant concern given that distribution deadlines focus on a rapid response to mandatory reports.

The results of this analysis can be presented in another way. As mentioned, 14% of all reports clearly fell within the existing mandatory systems. About half of these, according to the nurses' analysis, needed only routine monitoring. The FDA cite the example of a 'problem with a catheter in which there was medical intervention, but for which FDA already had taken action, so that additional reports would not make a very valuable contribution to the agency' [264]. However, 50 of the 175 reports that fell under voluntary reporting rules were rated as very urgent or urgent. This creates considerable problems for the prompt dissemination of safety-related information. Delays in a regulatory response do not simply stem from the time required to analyse a report and make recommendations. They also stem from the amount of time that it takes a contributor to actually gfile a report in the first place. Some of the contributors complained about the time limits that were recommended for reporting particular classes of incidents. In some cases, what contributors classed as less severe occurrences went unreported for more than ten days. The previous paragraphs have questioned the reliability of such severity assessments and so it seems likely that such delays may be a significant factor in ensuring the prompt dissemination of alerts and warnings.

14.1.3 Reaching the Intended Readership

Chapter 12.4 argued that the task of drafting incident reports is complicated by the diverse readerships that these documents can attract. This section extends this argument. The diverse readership of these documents not only complicates the drafting of an incident report but also exacerbates their dissemination. There is an immediate need to ensure that individuals within a working unit are informed of any recommendations following an incident. Chapter 4.3 argued that such actions are essential to demonstrate that contributions are being acted on in a prompt manner. In particular, reports should be sent to the individuals who initially provided notification of an adverse occurrence. These requirement apply to the dissemination of incident reports in both large and small scale systems. National and international schemes face additional distribution problems. In particular, incident reports must forwarded to other 'at risk' centres. This is a non-trivial requirement because it can often be difficult to determine precisely which centres might be affected by any potential recurrence. Within these associated working groups, it is important to identify who will assume responsibility for ensuring the reports are read by individual members of staff. This can involve close liaison between the investigators who draft a report and safety managers or other senior staff distributed throughout their organisation.

It is possible to identify a number of different dimensions that characterise the distribution of incident reports. The following list summarises these different dimensions. Particular reporting system may tailor the approach that they adopt according to the nature of the incident. They may also use hybrid combinations of these techniques. For example, a closed distribution policy might be exploited within the organisation that generate the report to ensure that information was not prematurely leaked to the media. However, a horizontal approach might also be used to ensure that key individuals in other companies are also made aware of a potential problem:

- *Closed distribution.*
  This approach restricts the dissemination of incident reports to a few named individuals within an organisation. This creates considerable problems in ensuring that those individuals and only those individuals actually receive copies of a report. It is also important to note throughout this analysis that the receipt of a report does not imply that it will be either read or acted upon.

- *Horizontal distribution.*
  This approach allows the dissemination of incident reports to other companies in the same industry. The distribution may be further targeted to those organisations that operate similar application processes.

- *Vertical distribution.*
  This approach allows the dissemination of reports to companies that occur within the same supply chain as the organisation that was notified about an incident. Reports can be passed

down the supply chain to ensure that companies, which rely on the products and services of the contributor organisation, are altered to a potential problem. Supply companies may also be informed if an incident occurs as the result of problems at previous stages in the supply chain.

- *Parallel distribution.*
  This approach ensures that reports are distributed to companies in *other* industries that operate similar processes. For example, incidents involving the handling and preparation of nuclear materials can have implications in the defence, medical and power generation industries. It is for this reason that organisations such as the US Chemical Safety and Hazard Investigation Board were set up to span several related domains.

- *Open distribution.*
  This approach allows the free distribution of incident reports. Increasingly, this approach is being adopted by regulatory organisations, including the FDA and MDA, and by independent research organisations, such as the NHS Centre for Reviews and Dissemination [192] As we shall see, these open publication initiatives increasingly rely upon Internet-based distribution techniques.

The healthcare industry provides extreme examples of the problem associated with distributing incident reports to a diverse audience. In 2000, the MDA received 7,249 reports of adverse incidents involving medical devices. These resulted in 4,466 investigations after an initial risk assessment. 49 safety warnings were published; the MDA's annual report bases this figure on the sum of the numbers of Hazard Notices, Safety Notices and Device Alerts in Table 14.1 [540]. Safety Notices are primarily distributed through the Chief Executives of Health Authorities, NHS Trusts and Primary Care Trusts as well as the directors of Social Services in England. These individuals a responsible for ensuring that they are brought to "the attention of all who need to know or be aware of it" [536]. Each Trust appoints a liaison officer who ensures that notices are distributed to the 'relevant managers'. Similarly, each local Health Authority appoints a liaison officer to ensure that notices are distributed to Chairs of Primary Care Groups, Registration Inspection Units, Independent Healthcare Sector and representatives of the Armed Services. The MDA also requires that notices are sent to the Chief Executives of Primary Care Trusts who are then responsible for onward distribution to their staff. Social Services Liaison Officers play a similar role but are specifically requested to ensure distribution to Registration Inspection Units and Residential Care Homes.

The distribution responsibilities of the individuals in the MDA hierarchy are presented in Table 14.3. The detailed responsibilities of each individual and group are, however, less important than the logistic challenges that must be addressed by the MDA when they issue a Safety Notice. For instance, any individual warning will only be sent to some portion of the total potential audience. Many Safety Notices are not relevant to the work of Social Services. In consequence, each published warning comes with a list of intended recipients. These are identified by the first level in the distribution hierarchy: Health Authorities, NHS Trusts, Primary Care Trusts and Social Services. Liaison officers are then responsible for ensuring that information is directed to those at the next level in the hierarchy. This selective distribution mechanism creates potential problems if, for example, a Social Service department fails to identify that a particular Safety Notice is relevant to their operations. The MDA, therefore, issue a quarterly checklist that is intended to help liaison officers ensure that they have received and recognised all applicable warnings.

The MDA distribution hierarchy illustrates a number of important issues that affect all reporting systems. There is a tension between the need to ensure that anyone with a potential interest in a Safety Notices receives a copy of the warning. This implies that Liaison Officers should err on the side of caution and disseminate them as widely as possible. On the other hand, this may result in a large number of potentially irrelevant documents being passed to personnel. The salience of any subsequent report might then be reduced by the need to filter these less relevant warning. These arguments, together with the expense associated with many forms of paper-based distribution, implies that Liaison Officers should target any distribution as tightly as possible. Later sections of this chapter will return to this tension when examining the generic problems of precision and recall in information retrieval systems.

Organisation	Liaison Officer forwards to	For onward distribution to
NHS Trust	Appropriate Manager	Relevant staff to include Medical Directors, Nurse Executive Directors, Directors of Anaesthetics, Directors of Midwifery, Special Care Baby Units/Pediatric Intensive Care, Maternity Wards, Operating Theatres, Ambulance NHS Trusts and Accident and Emergency Units.
Health Authority	Primary Care	Directors of Primary Care Local Representatives Committees Chief Executives of Primary Care Groups Individual GP Practices Dentists Opticians Pharmacists
	Registration Inspection Units	Care in the Community, Homes (Group Homes), Nursing homes, Managers of independent sector establishments, Private hospitals, Clinics and hospices
Social Services Department	In-house services	Residential Care Homes (elderly, learning difficulties, mental health, physical disabilities, respite care), Day Centres, Home Care Services (in-house and purchased), Occupational Therapists, Children's Services, Special Schools, Other appropriate Local Authority departments (for example Education departments for equipment held in schools).
	Registration Inspection Unit	Any of the above services provided by the independent sector.

Table 14.3: MDA's Distribution Hierarchy for Safety Notices

The success of the MDA distribution hierarchy relies on individual Liaison Officers. They exercise discretion in disseminating particular warnings to appropriate managers and directors. The significance of the Liaison Officer is also acknowledged by the MDA in a range of practical guidelines that are intended to ensure the integrity of these distribution mechanisms. For instance, healthcare organisations must identify a fax number and e-mail address for the primary receipt of Hazard Notices and Device Alerts. They must also arrange for someone to deputise in the Liaison Officer's absence. The Liaison Officer is responsible for ensuring that Hazard Notices and Device Alerts are distributed immediately after publication. Safety Notices can take a less immediate route, as described in previous paragraphs. Liaison Officers are also responsible for documenting the actions that are taken following the receipt of Hazard Notices, Device Alerts and Safety Notices. In particular, they must record the recipients of these various forms of incident report. The documentation should also record when the reports were issued and a signed assurance from the recipient that any required actions have been taken.

Liaison Officers not only pass on Safety Notices to 'appropriate' managers, they can also choose to distribute particular warnings to staff. For instance, such direct actions might be used to ensure that new employees or contract staff are brought up to date with existing warnings. These groups of workers create particular problems for the distribution of incident reports in many different industries. Not only do they create the need for special procedures in the national system operated by the MDA, they also complicate the task of communicating recommendations from local systems. Changes in working procedures in individual hospital departments create significant training overheads for temporary 'agency' staff who may be transferred between different units over a relatively short period of time. When such training is not explicitly provided then it is likely that communications problems will occur during shift hand-overs [344].

Previous sections have argued that Liaison Officers play an important role within the particular distribution mechanisms that are promoted by the UK MDA. Aspects of their role are generic; they characterise issues that must be addressed by all reporting systems. For example, the conflict between the need for wide distribution and the problems of overloading busy staff apply in all contexts. Many reporting systems must also ensure that new workers and contract staff are brought up to date. Similarly, there is a generic tension between enumerating the intended recipients of a report and allowing local discretion to determine who receives a report. This last issue can be illustrated by the way in which particular, critical reports constrain or guide the actions of Liaison Officers. For example, a recent Device Bulletin into patient injury from bed rails explicitly stated that it should be distributed to all staff involved in the procurement, use, prescription and maintenance of bed rails. Liaison officers were specifically directed to ensure that copies of the report were forwarded to 'health and safety managers; loan store managers; MDA liaison officers (for onward distribution); nurses; occupational therapists; residential and nursing home managers; risk managers.' [539] In contrast, other Device Bulletins explicitly encourage Liaison Officers to adopt a far broader dissemination policy. A report into the (ab)use of single-use medical devices enumerated the intended recipients as all Chief executives and managers of organisations where medical devices are used, all professionals who use medical devices, all providers of medical devices and all staff who reprocess medical devices [537].

Previous paragraphs have focussed on the problems of ensuring that incident reports are disseminated effectively within the organisations that participate in a reporting scheme. We have not, however, considered the additional problems that arise when any lessons must be shared between organisations that operate their own independent reporting systems. Legislation is, typically, used to provide regulators with the authority necessary to ensure that safety-related information is shared through national or industry-wide systems. Such legal requirements often fail to address the concerns that many companies might have about providing information to such reporting systems. There is a clear concern that commercially sensitive information will be distributed to competitors. The exchange of safety-related information often raise questions about confidentiality and trust:

> "(The) FDA is keenly aware of and sensitive to the impacts of these new regulatory requirements on the pace of technological advancement and economic well-being of the medical device industry. At the same time, the agency is cognizant of the usefulness of information about the clinical performance of medical devices in fulfilling its public health mandate... FDA may require the submission of certain proprietary information because it is necessary to fully evaluate the adverse event. Proprietary information will be kept confidential in accordance with Sec. 803.9, which prohibits public disclosure of trade secret or confidential commercial information.." [254]

Less critical information, for instance about near-miss occurrences, may be retained within corporate reporting systems. Other organisations can then be prevented from deriving any insights that such reports might offer. The ability to overcome these barriers often depends upon the micro-economic characteristics of the particular industry. For instance, it can be difficult to encourage the altruistic sharing of incident reports in highly competitive industries. In other markets, especially those that are characterised by oligopolistic practices, it can be far easier to ensure the cooperation and participation of potential rivals. For example, the major train operating companies combined with the infrastructure provides to establish the CIRAS reporting system on Scottish railways [198]. This scheme has a lot in common with the CNORIS regional reporting system that has recently been established across Scottish NHS hospitals [419]. Another feature of these systems is, however, that the lessons are seldom disseminated beyond the small group of companies or organisations within the oligopoly.

The increasing impact of a global economy has raised a number of difficult moral issues that were not initially considered by the early proponents of reporting systems. For example, there have been situations in which the operators of a non-punitive reporting system have identified failures by individuals who work in counties that do operate punitive, legal approaches to adverse occurrences [423]. Such situations can create particular problems when individual employees may have contributed an incident report on the understanding that they were participating in a 'no blame' system. Although these dilemmas are relatively rare, it is important to acknowledge the increasing exchange of data between different reporting systems. For instance, the 49 MDA warnings, cited in previous paragraphs, resulted in 32 notifications being issued to other European Union member states [540].

The direct distribution of reports by the MDA to other EU member states represents one of several approached to the international dissemination of safety-related information. It effectively restricts the dissemination of information, in the first instance, to the other participants in the political and economic union. Other distribution mechanisms must be established on a country-by-country basis for the wider distribution of information, for example with the US FDA. The Global Aviation Information Network initiatives represent an alternative approach to the dissemination of safety-related incident reports [310]. As the name suggests, the intention is to more beyond regional distribution to provide global access to this safety information. Similar initiatives can be seen in the work of the International Maritime Organisation (IMO) and the International Atomic Energy Authority. Such distribution mechanisms face immense practical and organisational barriers. The same issues of trust and confidentiality that complicate the exchange of information between commercial organisations also affect these wider mechanisms. There is also an additional layer of political and economic self-interest when incidents may affect the viability and reputation of national industries. These problems partly explain the halting nature of many of these initiatives. They are addressing the *distribution problem* by making information available to many national and regional organisations. However, they often fail to address the *contribution problem* because very few reports are ever received from some nations.

14.2 From Manual to Electronic Dissemination

Previous paragraphs have argued that the problems of disseminating information about adverse occurrences and near miss incidents stem from the frequency and diverse range of publications; from tight publication deadlines and resource constraints and from the difficulty of ensuring that the intended readership can access a copy of the report. These problems have been addressed in a number of ways. For example, the last section examined a number of distribution models that are intended to ease the logistics of disseminating incident reports. The hierarchical approach adopted by the MDA was used to illustrate the manner in which key individuals, such as Liaison Officers, often lie at the heart of hierarchical approaches. In contrast, this section moves on from these organisation techniques to look at the way in which different technologies can be recruited to address some of the problems that complicate the dissemination of incident reports.

14.2.1 Anecdotes, Internet Rumours and Broadcast Media

It is important not to overlook the way in which information about an incident can be disseminated by word of mouth. This can have very unfortunate consequences. For instance, the U. S. Food and Drug Administration's Center for Food Safety and Applied Nutrition describe how the company at the centre of an investigation first became aware of a potential problem through the circulation of rumours about their involvement [255]. They report that 'the first news the dairy plant received that they were being investigated in relation to this outbreak was through rumour on the street'. The plant operators then demanded to know what was going on; 'Apparently someone had heard someone else talking about the Yersinia outbreak and how it was connected to the dairy plant'. These informal accounts then had to be confirmed with a consequent loss of confidence in the investigatory procedures that had prevented disclosure of the potential incidents before the rumour began.

Informal channels are often faster and, in some senses, more effective at disseminating information than more official channels. Rumours often circulate about the potential causes well before they are published by investigatory organisations. Very often official reports into an adverse occurrence or near-miss come as little surprise to many of the individuals who work in an industry. The dissemination of safety information by word of mouth is not entirely negative. Many organisations, such as the FDA and the MDA rely upon such informal measures given limited printing budgets and the vast audiences that they envisage for some warnings. Similarly, the use of anecdotes about previous failures has provided an important training tool well before formal incident reporting systems were ever envisaged or implemented.

There is a danger, however, that the information conveyed by these informal means will provide a partial or biased account of the information that is published by more official channels. Word of mouth accounts are likely to provide an incomplete view before the official report is distributed. This can also occur after the official publication of an incident report if individuals mis-understand or forget the main findings of an investigation. They may also be unconvinced by investigators' findings. In such circumstances, there is a tendency to develop alternative accounts that resolve uncertainties about the official report. These unauthorised reports are, typically, intended to gain the listeners' attention rather than to improve the safety of application processes. It is difficult to underestimate the impact of such informal accounts. They can undermine the listeners' confidence in the investigatory agency even though they may retain significant doubts over the veracity of the alternative account [280].

In recent years, the informal dissemination of incident related information has taken on a renewed importance. The growth of electronic communication media has provided significant opportunities for investigatory agencies to distribute 'authorised' accounts. The same techniques also enable engineers, focus groups and members of the general public to rapidly exchange information about adverse occurrences and near miss incidents. The recognition that e-mail, Internet chat rooms and bulletin boards can facilitate the 'unauthorised' dissemination of such information has attracted significant attention from organisations, such as the FDA. The issues surrounding these informal communications are extremely complicated. For example, there is a concern that drugs companies and device manufacturers might exploit these communication media to actively promote their products. This resulted in a recent initiative to directly consider the position of the FDA towards 'Internet Advertising and the Promotion of Medical Products' [258]. During this meeting, a representative of one pharmaceutical company argued that they had an obligation to make sure that the information available to the public was as accurate as possible. Given the lack of Internet moderation, however, it was impossible for companies to correct every misconception that might arise; 'we do not correct every piece of graffiti that may be painted in some remote area of Australia or Alabama or Philadelphia, but we do respond where we feel this is significant and we need to clarify the issues'. A representative of another drug company addressed rumours about adverse events more directly: "there may be a rumour that a certain product is going to be withdrawn at a certain time and if no one comes in and steps in who has a authoritative information and says, 'This is not true', that kind of rumour can absolutely snowball and can become uncontrollable if it is not quashed right when it starts" [258].

Companies are not the only organisations that can have a direct interest in refuting what can be termed *Internet rumours*. The FDA recently had to launch a sustained initiative to counter rumours about the safety of tampons [265]. The FDA identified three different versions of this rumour:

1. One Internet claim is that U.S. tampon manufacturers add asbestos to their products to promote excessive menstrual bleeding in order to sell more tampons. The FDA countered this rumour by stating that 'asbestos is not, and never has been, used to make tampon fibers, according to FDA, which reviews the design and materials for all tampons sold in the United States' [276].

2. Another rumour alleged that some tampons contain dioxin. The FDA reiterated that 'although past methods of chlorine bleaching of rayon's cellulose fibers could lead to tiny amounts of dioxin (amounts that posed no health risk to consumers), today, cellulose undergoes a chlorine-free bleaching process resulting in finished tampons that have no detectable level of dioxin'.

3. A final Internet rumour argued that rayon in tampons causes toxic shock syndrome (TSS) and could make a woman more susceptible to other infections and diseases. The FDA responded that 'while there is a relationship between tampon use and toxic shock syndrome–about half of TSS cases today are associated with tampon use–there is no evidence that rayon tampons create a higher risk than cotton tampons with similar absorbency'.

In order to counter these various rumours, the FDA launched a coordinated distribution of information on the Internet and to the broadcast media. This response indicates the seriousness with which they regard the Internet as a distribution medium for alternative or 'unofficial' accounts of particular incidents, in this case involving Toxic Shock Syndrome. Such actions do not, however, come without a price. They help to ensure that the public are aware of the scientific evidence in support of the FDA claims. They also inadvertently raise the profile of those Internet resources that disseminate the rumours in the first place. The FDA's response, therefore, adds a form of reflected legitimacy to the original arguments about the link between Tampon's and TSS. It is important to emphasise that our use of the term 'rumour' is not intended to be pejorative. In many cases, the informal dissemination of information can provide a useful corrective to the partial view put forward by more 'official' agencies. Such alternative sources of information must, however, support their claims and statements with appropriate warrants. In particular, it can be argued that these informal sources of information force official agencies to focus more directly on the issues and concerns that affect the general public. The Internet rumours about the relationship between tampons and TSS may have contained numerous statements that could not subsequently be supported, however, they did persuade the FDA to clarify the existing evidence on any potential links.

The previous case study illustrates some of the complex changes that are occurring in the manner in which information about adverse occurrences is being disseminated. Internet bulletin boards and chat rooms help to publicise rumours that are then picked up by the popular media. At this stage, regulatory authorities must often intervene to correct or balance these informal accounts. It is, however, insufficient simply to publish a response via an official web site which is unlikely to attract many of the potential readers who have an interest in a particular topic. The regulatory agency is, therefore, compelled to exploit more traditional forms of the broadcast media to refute rumours that were primarily disseminated via the web and related technologies.

This reactive use of the media represents a relatively recent innovation. More typically, investigatory agencies have used the press, radio and television in a more pro active manner to disseminate the findings of incident reports. As we have seen, this use of the media requires careful planning; there is a danger that the parties involved in an investigation may learn more about their involvement from the press than from more official channels. The FDA is similar to many national agencies in that it follows detailed guidelines on the use of the media to disseminate information. For example, media relations must be explicitly considered as part of the strategy documents that are prepared before each product recall. The dissemination of information in this manner must be treated extremely carefully. It is important that the seriousness of any recall is communicated to the public. It is also important to avoid any form of panic or any adverse reaction that might unduly influence the long term commercial success of the companies that may be involved in an incident. The sensitive nature of such recall notices is recognised in the FDA provision that the warnings may be released either by the FDA or by the recalling firm depending on the circumstances surrounding the incident [259]. The political sensitivity of these issues is also illustrated by the central role that is played by the FDA's Division of Federal-State Relations during Class I recalls. This classification is used when is expected and when the 'depth' of the recall is anticipated to require action by a retailers and consumers. The Federal-State Relations division is required to use e-mail to notify state and local officials of recalls that are associated with serious health hazards or where publicity is anticipated. These officials are then issued with enforcement papers that are prepared by the FDA Press Relations Staff. This mechanism illustrates the manner in which investigatory agencies may operate several parallel dissemination activities each with very different intentions. In addition to the publication of incident reports, press releases are prepared to initiate actions by the public and by retailers. These may be distributed at press conferences, by direct contact with particular reporters and by releases to all Associated Press and United Press International wire services. Further distribution mechanisms must also ensure that individuals within relevant organisations are 'well briefed' to respond to questions from the press.

It is also important to acknowledge the central role of press and media relations staff. Not only does this department warn other members of the organisation of media interest. They also ensure that their colleagues are adequately briefed to respond to media interest. Their ability to perform these tasks is dependent upon them being notified in the early stages of any incident investigation. FDA regulations require that the Press Relations Staff are notified by any unit that 'publicity has occurred relating to the emergency condition, as well as pending requests for information from the media and/or public' [259]. The senior media relations staff then liaise directly with the officials closest to the scene to ascertain what information needs to be released and when it should be disseminated to best effect. It can, however, be difficult to ensure that such press releases will be given the prominence that is necessary in order to attract the publics' attention to a potential hazard. Some warnings have a relatively high news value. The FDA's Consumer magazine often provides journalists with a valuable starting point for these incidents. For instance, a recent warning centred on a particular type of sweet or candy that had resulted in three children choking to death. Some of these products carried warning labels, suggesting that they should not be eaten by children or the elderly. Other labels warn of a choking hazard and say to chew the sweets thoroughly. Some were sold without any warning. This story attracted immediate and focussed media interest. Another warning, which was issued on the same day as the one described above, attracted far less media attention. This concerned the potential dangers of consuming a mislabeled poisonous plant called Autumn Monkshood [268] The packages containing the plant were mistakenly labeled with the statement, 'All parts of this plant are tasty in soup'. They should have indicated that consumption of the plant can lead to aconitine poisoning and that death could occur due to ventricular arrhythmias or direct paralysis of the heart. Simply releasing information to the media about potentially fatal incidents does not imply that all incidents will be equally news worthy nor that they will receive equal prominence in press, radio or television broadcasts.

As we have seen, it can be difficult for regulatory and investigatory agencies to use the media as a means of disseminating safety information. This involves the coordination of press releases and conferences. It also involves the training of key staff, such as press liaison officers, and the use of electronic communications techniques to ensure that other members of staff are informed how to respond to media questions. Even if this infrastructure is established there is no guarantee, without legal intervention, that a particular warning will receive the prominence that is necessary to attract public attention. Such problems are most often encountered by large-scale national systems. The issues that are raised by media dissemination of incident information are, typically, quite different for smaller scale systems. There can also be a strong contrast in media relations when incident information attracts 'adverse' publicity. This is best illustrated by the phenomenon known as 'doctor bashing' which has emerged in the aftermath of a number of incidents within the UK healthcare industries. Many professionals find themselves faced by calls from the government and from the media to be increasingly open about potential incidents. For example, Alan Milnburn the UK Health Secretary has argued that the "National Health Service needs to be more open when things go wrong so that it can learn to put them right" [111]. Together with this increased openness "they would also have to be accountable for their errors and prepared to take responsibility". Some doctors have described such statements and the associated media publicity as 'hysterical'. Recent BBC reports summed up this attitude by citing a General Practitioner from the North West of England; "Shame on the media for sensationalising and exaggerating incidents...shame on you for failing to report accurately adverse clinical events" [110].

Public and government pressure to increase the dissemination of information about medical incidents must overcome many doctor's fear of adverse or 'sensational' press coverage. At present, many NHS trusts have still to face up to the consequences of this apparent conflict. They are reluctant to disclose information about previous incidents even to their own staff for fear that details might 'leak' to the press. In this domain at least, we are a very long way from the culture of openness that the proponents of incident reporting systems envisage as a prerequisite for the effective implementation of their techniques. It is important not to simply view these tensions as simply the result of media interest in disseminating sensational accounts of adverse incidents. They reflect deeper trends in society. The chairman of the British Medical Association's Junior

Doctors' Committee saw this when he argued that "we have a more consumerist society... people are complaining more about everything... there is a lot of doctor-bashing in the press" [108]. Such quotations illustrate the way in which the media not only inform society, as in the case of FDA warnings, but they also reflect the concerns of society.

This section has focussed on the 'informal' dissemination of information about adverse incidents. In particular, it has focussed on the way in which electronic and Internet-based communications have provided new means of distributing alternative accounts of near-misses and adverse occurrences. We have also described how regulatory organisations have used the same means to rebutt these alternative reports. The conventional media is routinely used to support these initiatives. It can also be used to publicise more general safety warnings and can initiate investigations where other forms of reporting have failed to detect safety-related incidents. This more positive role must be balanced with the problems of media distortion that dissuade managers from disseminating the findings of incident reporting systems. There is a stark contrast between the use of the media to publicise necessary safety information and the fear of publicity in the aftermath of an adverse event.

14.2.2 Paper documents

The previous section has done little more that summarise the informal communication media that support the distribution of safety related information. Similarly, we have only touched upon the complex issues that stem from the role of the media in incident reporting. These related topics deserve books in their own right, however, brevity prevents a more sustained analysis in this volume. In contrast, the remainder of this chapter focuses on more 'official' means of disseminating incident reports. In particular, the following section analyses the strengths and weaknesses of conventional paper-based publications to disseminate safety-related information.

One of the most suasive reasons for supporting the paper-based dissemination of incident reports is to meet regulatory obligations. The importance of this media is clearly revealed in the various regulations that govern the relationship between the FDA, manufacturing companies and the end-users of healthcare products. The primary focus of these regulations is on the exchange of written or printed documentation. This emphasis is not the result of historical factors. It is not simply a default option that has been held over from previous versions of the regulations that were drafted in an age before electronic dissemination techniques became a practical alternative. As we have seen, the recommendations in some incident reports can have a legal force. Companies may be required to demonstrate that they have taken steps to meet particular requirements. This creates problems for the use of electronic media where it can be very difficult to determine the authenticity of particular documents. It would be relatively easy to alter many of the reports that are currently hosted on regulatory and governmental web-sites. Later sections will describe a range of techniques, such as the use of electronic watermarks, that can increase a reader's confidence about the authenticity of the documents that are obtained over the Internet. Unfortunately, none of the existing incident reporting sites have adopted this technology. In consequence, paper versions continue to exist as the 'gold standard' against which compliance is usually assessed. Copies obtained by other distribution mechanisms are, therefore, seens as in some way additional to this more traditional form of publication.

A further benefit of conventional, paper-based dissemination techniques is that regulatory agencies can exploit existing postal distribution services. A host of external companies can also be used to assist with the formatting, printing and mailing of these documents. The technology that is required to perform these tasks is well understood and is also liable to be readily available within most organisations. These are important considerations. Simplicity and familiarity help to reduce the likelihood of failures occurring in the distribution process, although as we have seen they are not absolute guarantees! Minimal staff training is needed before information can be disseminated in this way. It is for this reason that most small scale reporting systems initially exploit this approach. Typically, newsletters are duplicated using a photocopying machine and are then made available either in staff common areas or in a position that is close to a supply of reporting forms.

Paper-based dissemination techniques simplify the task of distributing incident reports because they can exploit existing mechanisms, including staff distribution lists as well as both internal and

Very Well	Well	Not Well	Not at All
17,862,477	7,310,301	4,826,958	1,845,243

Table 14.4: 1990 US Census Data for Self-Reported Ability in English

state postal services. There are further advantages. No additional technology, such as a PC with an Internet connection or CD-ROM, is required before people can access safety-related information. This is a critical requirement for the dissemination of some incident reports. One participant at a recent FDA technical meeting was extremely irritated by the continual reference to web sites as a primary communication medium. He asked the others present whether they knew how many American could access the Internet or could understand English [262]. Such comments act as an important reminder that paper-based publications continue to have an important role in spite of the proliferation of alternative dissemination techniques. For the record, Table 14.4 provides the latest available figures from the 1990 US Census describing self-reported English ability. The total US population was reported as 230,445,777 of which there were some 198,600,798 individuals who reported that they could only speak English. There were 31,844,979 who described themselves as being primarily non-English speakers. The self-reported figures for the standard of English amongst this community are shown in Table 14.4. The proportion of the population who express problems in understanding English appears to be relatively small. However, there may be a significant proportion of the population who did not return a census form and there is a concern that the proportion of non-English speakers might be relatively high in this community. There is also a natural tendency to over-estimate linguistic ability in such official instruments. Such factors motivate the provision of alternate language versions of safety-related information [823]. The 2000 census provided further insights into the growth of the Internet amongst the US population [824]. The census asked 'Is there a personal computer or laptop in this household?'. The returns indicated that 54,000,000, or 51%, of households had one or more computers in August 2000. This was an increase of 42% from December 1998 45,000,000, or 42%, of households had at least one member who used the Internet at hone, This had risen from only 26% in 1998 and 18% in 1997. Such statistics reinforce the point that significant proportions of the population in what is arguably the world's most technologically advanced nation still do not have Internet access. This is liable to be less significant for incident reports that are targeted at commercial organisations, for which one might expect a higher percentage of Internet connectivity. The census statistics are, however, a salient reminder for more general reports and warning such as those issued by the FDA that are deliberately intended for the general public.

Paper-based dissemination techniques are also resilient to hardware failures. It is a relatively simple matter to find alternative printing facilities and postal services. It can be far more complex to introduce alternative web-servers or automatic fax routing services. The reliability of the distribution service is only one aspect to this issue. There can also be considerable problems in ensuring that the intended recipients of incident reports can successfully retrieve alternative formats. Postal services are seldom swamped by the volume of mail. The same cannot be said by web servers or even by the use of fax-based distribution techniques. If the intended recipient's fax machine is busy at the time when an automated distribution service attempts to distribute an incident report, critical information can be delayed by hours and even days. At peak times of the day, many requests can either fail entirely or be significantly delayed as users request incident reports from the FDA or MDA web-sites. One particular problem here is that many government web sites only make limited use of more advanced techniques, such as predictive cacheing or mirror sites [418]. Similarly, the servers that provide access to incident reports may also be used to provide access to other documents that attract a large volume of users throughout the day. There is a certain irony in the manner in which some incident reporting web sites also elicit user-feedback about the failure of those sites that are intended to provide access to other forms of incident reports. Even if readers can download a computer-based report, there is no guarantee that they possess the application software that may be required to view it. Chapter 12.4 described how most incident reporting sites exploit either HTML and PDF. The former supports the dissemination of web-based documents because no additional support is required beyond a browser. Unfortunately, there is no guarantee that a document, which is formatted in HTML will be faithfully reconstructed when printed. This is significant because the psychological literature points to numerous cognitive and perceptual problems associated with the on-screen reading of long and complex documents [876]. In consequence, many organisations exploit Adobe's proprietary PDF format. PDF readers can be downloaded for most platforms without any charge. Problems arise, however, when incident reports that have been prepared for viewing under one version of the reader cannot then be viewed using other versions. For instance, a recent MDA report into Blood Pressure Measurement Devices contained the following warning: "Adobe Acrobat v.4 is required to view on screen the content of the tables at p.9 + 16...Adobe Acrobat v.3 can view remainder of document and can print in full". Paper-based dissemination techniques avoid such problems, which present a considerable barrier for many users who might otherwise want to access these documents.

There are further benefits to more traditional dissemination techniques. For instance, the physical nature of paper-based publications enables regulators to combine documents in a single mailshot. This is important because potential readers can skim these related items to see whether or not they are relevant to their particular tasks. This can be far more difficult to achieve from the hypertext labels that are, typically, used to encourage readers to access related items over the web [758]. The flexible nature of printed media can be illustrated by the way in which Incident Report Investigation Scheme news and safety alerts were directly inserted into printed copies of the Australian Therapeutic Goods Administration newsletter [45]. Similar techniques have been adopted by many different investigation schemes. Safety-related information is included into publications that are perceived to have a wider appeal. This is intended to ensure that more people will consider reading this information than if they had simply been sent a safety-related publication.

There are also situations in which investigatory and regulatory organisations have no alternative but to use printed warnings. For example, the FDA took steps to ensure that printed warnings were distributed about the danger of infection from vibrio vulnificus as a result of eating raw oysters [256]. The signs and symptoms of previous cases were described and the resulting warnings were posted at locations where the public might choose to buy or consume these products. The use of the Internet or of broadcast media provides less assurance that individuals who are about to consume raw Oysters are aware of the potential risks. This incident also illustrates some of the limitations of paper-based dissemination techniques. Many of the cases of infection were identified in and around Los Angeles. The FDA soon discovered that, as noted above, a significant proportion of this community could not speak or read English at the level which was required to understand the signs that had been posted. The States of California, Florida, and Louisiana only required Oyster vendors to post signs in English. In consequence, the FDA supplemented these printed warnings with a 24-hour consumer 'Seafood Hotline' that provided information in English and Spanish.

There are a number of problems that limit the utility of paper-based dissemination techniques as a means of distributing the documents that are generated by incident reporting systems. The most obvious of these issues is the cost associated with both the printing and shipping of what can often be large amounts of paper. These costs can be assessed in purely financial terms. They are also increasingly being measured in terms of their wider ecological impact, especially for large scale reporting systems that can document many thousands of contributions each year. Many organisations attempt to defray the expenses that are associated with the generation and distribution of incident reports by charging readers who want to obtain copies of these documents. This raises a number of complex, ethical issues. For example, the cost of obtaining a copy of an incident report can act as a powerful disincentive to the dissemination of safety-related information. This should not be underestimated for state healthcare services where any funds that are used to obtain such publications cannot then be spent on more direct forms of patient care. Some regulatory bodies, therefore, operate a tiered pricing policy. For example, the MDA do not make a charge for any of the Device Bulletins requested by members of the National Health Service. In contrast, Table 14.5 summarises the prices that must be paid to obtain copies of a number of recent MDA documents by those outside the national health system [541].

The costs illustrated in Table 14.5 do not simply reflect the overheads associated with the printing and shipping of these documents. They also, in part, reflect the costs of maintaining a catalogue of previous publications. This can prove to be particularly difficult with paper-based reports given the

Device Bulletins - 2001

Number	Title	Issue Date	Price
DB 2001(04)	Advice on the Safe Use of Bed Rails	July 2001	£15
DB 2001(03)	Guidance on the Safe Transportation of Wheelchairs	June 2001	£25
DB 2001(02)	MDA warning notices issued in 1995	May 2001	Free
DB 2001(01)	Adverse Incident Reports 2000	March 2001	Free

Device Bulletins - 2000

Number	Title	Issue Date	Price
DB 2000(05)	Guidance on the Purchase, Operation and Maintenance of Benchtop Steam Sterilisers	October 2000	£25
DB 2000(04)	Single-Use Medical Devices: Implications and Consequences of Reuse Replaces DB9501	August 2000	£15
DB 2000(03)	Blood Pressure Measurement Devices - Mercury and Non-Mercury	July 2000	£15
DB 2000(02)	Medical Devices and Equipment Management: Repair and Maintenance Provision	June 2000	£25
DB 2000(01)	Adverse Incident Reports 1999 Reviews adverse incidents reported during 1999 and describes MDA actions in response.	March 2000	Free

Table 14.5: Pricing Policy for Recent MDA Device Bulletins

storage that is required to hold the large numbers of publications that were described in the opening pages of this chapter. The MDA has published well over 300 different reports in the last five years. The logistics of supporting the paper-based distribution of such a catalogue has led many similar organisations to abandon such archival services. The Australian Institute of Health and Welfare now only provide the detailed back-up data and tables for many of their publications in electronic format [41].

A number of further limitations affect the use of paper-based dissemination techniques. The previous paragraphs have argued that such approaches do not suffer from the problems of server saturation and network loading that can affect electronic distribution mechanisms. Unfortunately, more convention dissemination mechanisms can suffer from other forms of delay that can be far worse than those experienced with Internet retrieval tools. Even with relatively efficient administration procedures there can be a significant delay between the printing of a report and the time of its arrival with the intended readership. These delays are exacerbated when safety managers or members of the general public require access to archived information about previous incidents. For instance, the MDA promise to dispatch requested reports by the next working day if they are in stock [542]. If they are not currently in print then they will contact the person or organisation making the request within forty-eight hours. These delays can be exacerbated by the use of the UK's second-class postal service to dispatch the requested copies of the report. This reduces postage costs, however, it also introduces additional delays. The second class service aims to deliver by the third working day from when it was posted. In the period from April to June 2001, 92.5% of second class 'impressions' satisfied this target. This is an important statistic because it implies that even if there is a relatively long delay before any requested report can be delivered, the duration of this delay is relatively predictable. In the same period, the UK postal servise achieved close to 100% reliability in terms of the number of items that were lost. The high volume of postal traffic does, however, mask the fact that Consignia received 223,495 complaints about lost items, 40,529 complaints about service delays and 37,256 complaints about mis-deliveries by the Royal Mail service from April to June 2001.

The delays introduced by a reliance on the postal service or similar distribution mechanisms also creates problems in updating incident reports. In consequence, organisations may be in the process of implementing initial recommendations at a time when these interim measures have already been revised in the final report. Updating problems affect a wide range of the publications that are produced from incident reporting systems. For instance, the FDA explicitly intended that their Talk Papers, which are prepared by the Press Office to help personnel respond to questions from the public, are subject to change 'as more information becomes available' [269]. Even when revisions are made over a longer time period, it is important not to underestimate the administrative burdens and the costs of ensuring that all interested parties receive new publications about adverse incidents. This point can be illustrated by the problems surrounding Temporomandibular Joints (TMJs). These implants have been used in several dental procedures. They were initially introduced onto the market before a 1976 amendment that required manufacturers to demonstrate that such products were both safe and effective. TMJs were, therefore, exempt from the terms of the amendment. From 1984 to June 1998, the FDA received 434 adverse event reports relating to these devices. 58% of these incidents resulted in injury to the patient. In 1993, the Dental Products Advisory Panel reclassified TMJs into their highest risk category (III). All manufacturers of TMJ devices were then required to submit a Premarket Approval Application, demonstrating safety and effectiveness, when called for by the FDA. In December 1998, the FDA called for PMAs from all manufacturers of TMJ implants. This was followed up by the publication in 1999 of a consumer handbook entitled, 'TMJ Implants - A Consumer Informational Update'. In April 2001 this was updated to present further information about the changing pattern of incidents involving these devices. As can be seen, adverse occurrences led to the publication of reclassification information in 1993. This had to be disseminated to all device manufacturers. This was revised in 1998 when the Premarket Approval Applications were called for. This change has considerable implications; the FDA have to ensure that they contact all of the commercial organisations that might be affected by such a change. TMJ's are relatively specialist devices and so only a hand-full of companies are involved in manufacturing them in the United States. It is important to recognise, however, that the Class III categorisation also applied to the sale of foreign imports. One solution to the potential problems that might arise in such circumstances is to use legal powers to require that all device manufacturers take measures to ensure that they are aware of any changes to the regulatory status of the devices that they produce. Such an approach is, however, infeasible for members of the general public and even for clinicians. It would clearly not be a productive use of FDA resources if their administrative staff had to answer repeated requests from concerned individuals who were simply wanting to check whether or not they had received the most recent information about particular devices.

The web offers considerable benefits for the dissemination of updated information about adverse occurrences and near-miss incidents. A single web-site can act as a clearing house for informations about particular products, such as TMJs, users can then access this page in order to see whether or not the information there had been updated. This approach raises interesting questions about the relationship between the reader and the regulatory or investigatory organisation that disseminates the information. In a conventional paper-based approach, a *push* model of distribution was used. The incident reporting organisation actively sent concerned individuals updated copies of information that they had registered an interest in. This enabled regulators to have a good idea about who read their reports. The overheads associated with this approach persuaded some organisations to adopt a *pull* model in which interested readers had to explicitly request particular documents. The dissemination of reports could then be targeted on those who actually wanted them rather than simply sending everyone a copy of every report. The administrative costs associated with such a scheme have persuaded many organisations to adopt the electronic variant of this approach in which individuals are expected to *pull* updated reports from a web page of information. This removes many of the costs associated with the production and distribution of paper-based reports. It also prevents regulators and investigators from having any clear idea of who has read the incident reports and associated publications that they have produced. Web server logs can prove to be misleading, given the prevalence of cacheing and other mechanisms for storing local copies of frequently accessed information [418].

14.2.3 Fax and Telephone Notification

Telephone and fax-based systems provide a compromise between the push-based approach of paper dissemination and the pull-based techniques of more recent, Internet approaches. In their simplest form, a pre-recorded message can be used to list all of the most recent updates and changes to paper-based documentation. This can help potential readers to identify the report that they want without consuming the regulator's finite administrative resources. It also enable frequent and rapid updates to be made to the information that is pre-recorded. Unfortunately, the linear nature of recorded speech can make this approach impractical for agencies that publish many different reports. A potential reader would have to listen to the recording for many minutes before hearing about a potential item of interest.

The use of pre-recorded messages to provide an index of updates to incident reports still does not address many of the administrative and resources problems that can arise from the paper-based distribution of these documents. At some point, copies of the report have to be printed and shipped to the prospective readers. One solution to these problems is to use fax machines to distribute incident reports. This approach has numerous benefits. For instance, fax-servers can be pre-programmed with large sets of telephone numbers. They will then automatically ensure that a faxed document is sent to every number of the list. More advanced systems will suspect a call if the fax machine is busy and will then re-try the number later in the run. The use of fax machines can also help regulatory authorities to keep track of the recipients of particular documents. For example, the UK MDA's institutional Business Plan for 2001-2002 includes the objective to monitor 'first time' fax failures when urgent safety warnings are issued to liaison officers. Of course, it is not possible to ensure that named individuals will have received and read a document that is sent in this manner. There can, however, be a reasonable degree of assurance that the fax has been received by the organisation associated with a particular fax number.

The FDA has pioneered the development and use of a more refined version of the systems described in previous paragraphs. They have developed a fully automated 'Facts on Demand' system. The user dials up the service and they then hear a series of instructions. If, for example, they press '2' on their keypad then they can hear more detailed instructions on how to use the system. If they press '1' then they can choose to order a document. If they dial 'INDX' or 4639 on the keypad then they can order an index of all documents on the system. If they choose to order an index, the system will call them back to fax a catalogue of publications. This currently runs to more than 50 page, however, it avoids the problems associated with listening to a pre-recorded listing for several hours! Callers can then use this faxed index to identify the identifier of the document that they want to retrieve. They must then call the system again, select the required option and then enter the document identifier. The system will then automatically fax them back with the required publication. The only technical requirement for the user of such a system is that they have access both to a fax machine and to a touch-tone telephone [266].

Most incident reporting systems continue to use paper-based dissemination techniques. Technological approachs, such as that described above, provide additional facilities that build on these more traditional approaches. This situation is gradually changing under increasing financial and administrative pressures. These influences can be seen behind the decision to move to the electronic publication of the FDA's User Facility Reporting Bulletin. In 1997, it was decided this it was no longer possible to print and mail this document out to anyone who requested it:

> "Time, technology, and budget restrictions have come together in the Food and Drug Administration. Ten years ago, our computer capability allowed us to communicate only within FDA. Now, with advanced computer technology we can globally communicate through the Internet and through Fax machines. As you would expect, Congressional budget cuts have affected all parts of government. FDA did not escape these cuts. In the search for ways to reduce our expenses, printing and mailing costs for distribution of publications in traditional paper form have come to be viewed as an extravagant expenditure... Now, budget restrictions prevent future distribution in paper form. We regret the need to move to this new technology if it means that many of our current readers will no longer have access to the Bulletin. We would like to remind you that you

> can also obtain copies through our Facts-on- Demand System or the World Wide Web." [868]

The concerns voiced in this quotation are understandable given the relatively low penetration of Internet connections into many areas of the US economy in 1997. Fax systems, such as 'Facts on Demand', provided an alternative dissemination technique. It can, however, be argued that they are likely to be replaced as more and more companies invest in Internet technology. Computers-based dissemination will then become the primary means of distributing incident reports. Before this can happen, however, we will need to address security concerns and the legal status of electronic reports. We will also need to consider the consequences of providing electronic access to large collections of safety-critical incident reports.

14.3 Computer-Based Dissemination

There are many diverse reasons that motivate the increasing use of information technology to support incident reporting systems. These approaches offer the potential for almost instantaneous updates to be disseminated across large distances. As we shall see, computer-based systems also offer security and access control facilities that cannot easily be provided using paper-based dissemination techniques. The same technological infrastructure that supports the rapid dissemination of individual incident reports also offers mass access to historical data about previous incidents and accidents. There are further motivations for providing this form of access to incident databases:

- *supporting risk assessment.* An important benefit of providing wider access to incident databases is that safety managers can review previous incidents to inform the introduction of new technologies or working practices. This information must be interpreted with care; contribution and reporting biases must be taken into account. Even so, incident databases have been widely used to support subjective estimates about the potential likelihood of future failures [423].

- *identifying trends.* Databases can be placed on-line so that investigators and safety managers can find out whether or not a particular incident forms part of a more complex pattern of failure. This does not simply rely upon identifying similar causes of adverse occurrences and near misses. Patterns may also be seen in the mitigating factors that prevent an incident developing into a more serious failure. This is important if, for example, safety managers and regulators were to take action to strengthen the defences against future accidents.

- *monitoring the system.* Regulators and safety managers can monitor incident data to determine whether particular targets are being achieved. Incident databases have been monitored to demonstrate reductions in particular types of incidents. They have also been used to support arguments about overall safety improvements. The following chapter will address the problems that affect this use of reporting systems. For instance, any fall in the number of contributions to a reporting system can reflect a lack of participation rather than an increased 'level' of safety.

- *encouraging participation.* If potential contributors can monitor previous contributions, they can be encouraged to participate in a reporting system. Information about previous incidents helps to indicate the types of events and near misses that fall within the scope of the system. Evidence of previous participation can also help to address concerns about retribution or of accusations about 'whistle blowing'.

- *transparency and the validation of safety initiatives.* Wider access to incident data helps to validate any actions that are taken in the aftermath of an adverse occurrences. For example, several reporting systems have used their incident data to draft a 'hit' list of the most serious safety problems [36]. By providing access to the underlying data that supports such initiatives, manufacturers and operators can see the justifications for subsequent regulatory intervention.

- *information sharing.* The development of on-line incident databases enables safety managers and regulators to see whether similar incidents have occurred in the past. This information technology provides further benefits. For the first time, it is becoming possible to extend the search to include the on-line databases of incidents in other countries and even in different industries. It is important not to underestimate the opportunities that this creates. For instance, it is possible to directly view incident data relating to the failure of medical devices in the USA prior to their approval for use in the UK. Conversely, it is becoming possible for the authorities in the USA to view elements of the submission for approval for particular forms of drugs that are submitted to the UK authorities. The electronic indexing of all of this data can help investigators, regulators and safety-managers to search through a mass of information that would otherwise overwhelm their finite resources.

The FDA recently summarised the advantages of electronic over paper based systems for incident reporting in the healthcare domain [274]. Firstly, they argued that automated databases enable readers to perform more advanced searches of information. This is important because it is likely that individuals may miss relevant information if they are expected to perform manual inspections of large paper-based data sets. Secondly, FDA also argued that computer-based retrieval systems can be used to view a single collection of information from a number of different perspectives. For instance, it is possible to present summaries of all incidents that relate to particular issues. This might be done by issuing a request to show every incident that involves a software bug or the failure of particular infusion devices. Similarly, other requests might be issued across the same collection of reports to identify incidents that occurred in particular geographical locations or over a specified time period. Such different views can, in principle, be derived from paper-based documentation of adverse occurrences and near-miss incidents. The costs of obtaining such information are, however, likely to be prohibitive. The third justification that the FDA identified for the use of electronic information systems was that they support the analysis of trends and patterns. Many incident reporting systems are investigation in new generations of 'data mining' applications and 'search engine' that can identify subtle correlations within a data-set over time. Finally, the FDA argued that electronic information systems avoid 'initial and subsequent document misfiling that may result from human error' [274]. As we shall see, this particular benefit can be more than offset by the problems that many users experience when they attempt to use computer-based systems to retrieve particular incident reports. Many applications require the use of arcane command languages or pre-programmed queries that are often poorly understood by the people that must use the information that is returned to them.

14.3.1 Infrastructure Issues

It is important not to automatically assume that all incident reporting systems are following a uniform path in the application of information technology. There is an enormous diversity of techniques. Some systems, such as the CIRAS rail application [198], deliberately avoid the use of computer networks. Security concerns have persuaded them to use stand-alone machines. This has profound concerns for the collation of distributed data. Paper forms are used throughout the Scots rail network and extensive use is made of telephone interviewing, even though it can often be difficult to arrange times when contributors can be contacted in this manner. Other organisations, such as the Swedish Air Traffic Control Organisation, have deliberately created computer-based reporting systems that exploit the benefits of networked applications. Individuals can log-onto the system from many diverse locations both to submit an incident report and to monitor the progress of any subsequent investigation. The following pages, therefore, review the strengths and weaknesses of the technological infrastructures that have supported incident reporting systems. The term 'technological infra-structure' is used here to refer to the means of distributing computer-based records of adverse occurrences and near-miss incidents. Subsequent sections examine issues that relate more to the retrieval of reports once they have been disseminated. In other words, this section looks at the techniques that investigatory bodies can use to push information out to end-users. Later sections look at the systems that end-users can exploit to search through that data and pull out information about particular incidents.

Stand-Alone Machines

Many incident reporting systems initially make very limited use of computer-based tools. Typically, they recruit mass-market desktop applications such as spreadsheets and text editors to help with managerial and logistical tasks. There are also more mature systems that have deliberately adopted the policy not to use more advanced computational tools. This decision can be justified in a number of ways:

- *security concerns.* The most pressing reason not to exploit more advanced technology in general, and network connectivity in particular, is that many safety managers have concerns about their ability to maintain the security of commercially sensitive incident data. Even if those operating the system have satisfied themselves that precautions can be taken against potential threats, senior levels of management may intervene to prevent the use of local or wide area networks. Security concerns are often most significant when an independent reporting agency holds data on behalf of operating companies. In such circumstances, the integrity and privacy of that information is often a prerequisite for running the system in the first place;

- *cost issues.* The declining costs associated with computer hardware have not been matched by similar reductions in connectivity charges within particular areas of the globe. In consequence, many small scale reporting systems may not be able to justify the additional expense associated with anything more advanced than a stand-alone machine. This is particularly important when the potential contributors to a reporting system are geographically distributed and may be involved in occupations that do not directly involve the use of information technology. For example, high wiring costs and legacy buildings have prevented many NHS trusts from providing direct network access to all of the wards in every hospital.

- *lack of relevant expertise.* Costs not only stem from the physical infrastructure. They also relate to the additional technical expertise that is required to connect stand-alone machines to local and wide area networks. It is relatively simple to register a single machine with an Internet service provider. Most incident reporting systems, however, depend on gather information from and disseminating information to large networks of contributors. Previous sections have also stressed the importance of maintaining connectivity within these networks to ensure that safety information is disseminated in a timely fashion. These factors combine to make it likely that any move beyond a stand-alone architecture will incur additional costs in terms of technical support to maintain the computer-based infrastructure.

- *'not invented here' syndrome.* The previous justifications for the continued use of stand-alone machines are well considered and appropriate. There is, however, a further reason for the longevity of this simple architecture that does little credit to the systems that embody it. As mentioned, many incident reporting systems begin by using commercial, off-the shelf packages to collate statistical data about previous incidents. As we shall see, many of the individuals who maintain the automated support are acutely aware of the problems and limitations that this software imposes on its users. There can, however, be a reluctance to move beyond these initial steps to elicit professional support from software engineers. This stems form a justifiable fear that the use of more advanced systems may imply a loss of control. In consequence, we see safety-critical data being held on mass-market systems whose licenses explicitly prohibit their use in such critical applications.

Stand-alone computers, without network connectivity, continue to play a significant role in many incident reporting systems. It is possible to identify a number of different modes of operation. They can be used simply to disseminate forms that can then be edited locally. Contributors can then print out the details of a particular incident and post the completed form back to a central agency. This is the approach advocated by the US Joint Commission on Accreditation of Healthcare Organisation's Sentinel Event system [431]. They request that all organisations transmit 'root cause analysis, action plan, and other sentinel event-related information to the Joint Commission through the mail' rather than by electronic means. The UK MDA also offer a range of electronic forms in PDF and in Microsoft Word format that can be printed and posted back when they have been completed [538].

The dissemination of incident reporting forms that are then intended to be printed and posted back to central agencies creates something of a paradox. Organisations that use stand-alone machines to coordinate their reporting policies must first obtain copies of these documents before they can complete their submission in a secure manner. The most common means of doing this is to use another networked machine to download the form then copy this across to the isolated machine on a disk. This is clearly a protracted mechanism. It may also fail to achieve the level of security that many people believe it ought to. For example, stand-alone machines are equally vulnerable to the free distribution of passwords and the problems associated with unlocked offices [1].

A slightly more complex use of information technology is to distribute a suite of programs that not only helps with drafting incident reports but also helps with the investigation and analysis of near misses and adverse events. Many of these systems are not primarily intended to support the exchange of information between institutions but to ensure that clear and coherent procedures are adopted within an institution. Rather than supporting the dissemination of information about incidents, regulators disseminate software support for local reporting systems. A recent collaboration between Harvard School of Public Health, the MEDSTAT Group, Mikalix & Company and the Center for Health Policy Studies illustrates this approach [3]. They devised the Computerised Needs-Oriented Quality Measurement Evaluation System (CONQUEST) system. This is intended to support general information about quality assurance in healthcare. It does, however, share much in common with many incident reporting systems. For instance, it was designed to help managers and clinicians derive answers to the following questions:

1. Did the clinician do the right thing at the right time?

2. Was effective care provided to each patient?

3. Was care provided safely and in an appropriate time frame for each patient?

4. Was the outcome as good as could be expected, given each patient's condition, personal characteristics, preferences, and the current state of medical science? [3]

The project began by devising a classification scheme for the information that it was to maintain. It then "became obvious that a computer database was the logical way to store and retrieve the data". A mass-market, single-user database application was then chosen as the implementation platform for this system. This decision reflects considerable expediency. It is possible to use these applications to quickly craft a working application that can be used to store information about several thousand incidents. This is sufficient for individual hospitals, however, it will not provide indefinite support as the number of incidents increases over time. Nor do these systems provide adequate support for the storage and retrieval of incident information on a national scale.

Further problems complicate the use of stand-alone architectures to support local incident reporting systems. Many of the distributed software systems that run on these applications offer means of tailoring the format of the data to meet local requirements. In general, this is an excellent approach because it provides safety managers with a means of monitoring incident-related information that might have particular importance within their working context but which has been ignored at a national level. Unfortunately, one consequence of this flexibility is that local systems often develop electronic data formats and classification schemes that are entirely inconsistent with those used by other organisations. These problems even occur when organisations exploit the same version of incident reporting software. Local changes can, over time, partition incident data so that it is difficult to join the individual approaches into a coherent overview of incidents across an industry. Technically, it is possible to match variant fields in different systems providing that it is possible to identify relationships between this information. This will not, however, resolve the problems of missing or partial data.

Electronic Mail

If incident information is only ever to be held and used locally then many of the previous caveats are of limited importance. In most cases, however, there is a need to support the exchange of incident information about near-misses and adverse occurrences. Arguably, the most common means of supporting such transmission is through the use of electronic mail. This generic term is usually applied to a range of software applications that exploit the Internet-based Simple Mail Transfer Protocol (SMTP). This enables users to send arbitrary messages between different accounts. The 'Seafood Network' provides a good example of the effective use of this simplest form of electronic communication [688]. This ensures that any message sent to central account, or list server, is automatically distributed to everyone who is registered with the service. The primary purpose of this "Internet based seafood network is to facilitate information exchange about the Hazard Analysis and Critical Control Point (HACCP) system of food safety control in the seafood industry". The HACCP system can be thought of as a form of incident reporting scheme that also disseminates more general safety-related information. There are, however, certain pitfalls that can arise from this *relatively* simple use of electronic dissemination techniques:

> "It has become apparent that some network users may not realise that the e-mail address, seafood@ucdavis.edu, automatically distributes a message to over 400+ subscribers worldwide. By all means, if you want everyone on the seafood network to read your message, address it to seafood@ucdavis.edu To communicate privately as a follow-up, please respond to the individual's e-mail address that is listed on the message.

As mentioned, SMTP supports the exchange of simple text messages. Some email applications also support Multipurpose Internet Mail Extensions (MIME). These enable senders to attach files of a particular type to their mail messages. This is useful if, for example, the sender of a mail message wanted to ensure that the recipient used CONQUEST or a similar system to open the file that they had attached to their email. MIME not only sends the file but also send information, that is usually hidden from the user, about those programs that can be used to open the attachment. It is clear that the recipients of any message must be able to interpret its contents. In the case of standard SMTP mail, the human reader must be able to understand the contents of any message. In the case of a MIME attachment, the associated program must be able to interpret the data that is associated with the mail message. Some industries are more advanced than others in specifying the format that such transmissions should take. In particular, there are many reasons why this issue should be of particular interest to healthcare professionals. There must be clear standards for the transmission of information if doctors are to interpret patient-related information sent from their colleagues. Similarly, the increasing integration of testing equipment into some hospital networks has created situations in which results can be automatically mailed to a clinician. In consequence, many professional bodies have begun to collaborate on standards for the transmission of clinical information. For example, Health Level 7 is an initiative to develop a Standard Generalised Markup Language similar to that used on the web, for healthcare documents [185]. These standards are not primarily intended to support incident reporting. They can, however, provide convenient templates for the transmission and dissemination of incident reports that are consistent with emerging practices in other areas of healthcare. This approach is also entirely consistent with attempts to develop causal taxonomies for incident reporting. The leaf nodes in techniques, such as PRISMA described in Chapter 10.4, might be introduced within such languages to record the results of particular investigations.

At present, most reporting systems do not exploit such general standards in the electronic transmission of incident information. Instead, they rely upon a range of formats that are tailored to particular reporting systems and its associated software. One of the most advanced examples of this approach is provided by the Australian Incident Monitoring System [36]. Initially, like many reporting systems, this relied upon the paper-based submission of incident forms from hospitals and other healthcare organisations. As the system grew, it established a network of representatives within 'healthcare units'. These representatives now collate the paper-based reports and enter them into a database that exploits the Structured Query Language (SQL). SQL can be viewed as a standard that describes the language that is used when forming requests for information. The software embodies the Generic Occurrence Classification (GOC). This supports the categorisation and subsequent analysis of the incident records that are held within the system. As mentioned, this initial data entry is performed within the 'health units'. At this stage, the system contains confidential information on those involved in the incident. It is, therefore, protected from legal discovery under Australian Commonwealth Quality Assurance legislation. For monitoring purposes, the Australian Patient Safety Foundation (APSF) then collates information from the individual units. Before this is done, all identifying information is removed from the individual reports. Current versions of the AIMS software enable individual units to email this information to the APSF system using the MIME techniques, described above. An important aspect of this transmission is that the individual records are encrypted prior to transmission. Later sections will describe how SMTP mail services are insecure and that such techniques are a necessary precaution against the unauthorised access to such information.

As mentioned, the AIMS approach is both innovative and well-engineered. There are, however, a number of potential problems with the techniques that it exploits. It adopts a model that is very similar to the stand-alone architecture, which was described in previous paragraphs. The collated database of anonymised incident information is held by the APSF as a central resources. This not only secures the data it also, potentially, acts as a bottleneck for other healthcare professional who might have a legitimate interest in analysing the data. In particular, safety managers in individual health units cannot directly pose queries to determine whether an incident forms part of a wider pattern. They must go through the mediation of the APSF. Increasing incident reporting systems are adopting a more egalitarian model in which anonymised incident data is also be distributed by electronic means. The move has been inspired by work in the aviation community and, in particular, the metaphor of an information warehouse that has been promoted by the GAIN initiative [310]. Those who contribute data should also have direct access to the data that is contributed by peer organisations. This egalitarian approach poses considerable logistical problems, including the difficult of ensuring the security of information transfers. As we shall see, a range of encryption techniques can be used during transfer. Password protection can also be used to restrict access. Neither of these techniques addresses the problems of ensuring the consistency of incident databases that may be replicated in each of the peer organisations. In other words, if the AIMS database were to be distributed more widely there would be a danger that one hospital might be using a collection that was updated in 2000 while others used more recent versions of the database. This problem arises because information about new incidents must not only be sent to the central clearing house operated by the APSF, it must also be sent to all peer organisations. Email can be used for this but there is no guarantee that every message will be acted upon and incorporated into the existing database. It is also likely that the size of many reporting systems would prevent any attempt to email out the entire collection at regular intervals. It is for this reason that organisations such as the National Transportation Safety Board (NTSB) exploit a mixture of on-line updates and CD-ROM digests of their incident databases. Organisations can then either download each new set of reports as they become available or simply order a CD-ROM of the entire updated collection.

CD-ROMS

Compact Disk-Read Only Memory provides a relatively cheap means of disseminating incident information without requiring that the recipient exposes their machine to the security risks associated with a network connection. This technology also avoids the level of technical support that can be associated with network administration. The popularity of is format is based on storage capacity of this media. Most CD-ROMs provide nearly 0.7 gigabytes of data; this is equivalent to almost 500 high-density floppy disks. The emerging successors to this format, such as Digital Versatile Disc (DVD-ROMs), expand this capacity to 8.5 gigabytes. CD-ROMs are also highly portable and light weight making the postal distribution of large amounts of data far cheaper using this format than printed documentation. It is also possible to encrypt the information on a CD-ROM; this provides added protection against the problems that can arise if critical documents go missing in the postal system. All of these technical attributes make this format particularly well suited as a communication medium for the dissemination of incident databases.

The CD-ROM format has further advantages. In particular, they provide a maximum data rate of between 2.8 and 6 megabytes per second on a 40x drive. This might seem a relatively trivial, technical statistic. However, such speed enable regulators and investigatory agencies to include multi-media resources, such as short audio and video clips, in addition to textual information and static images. Previous sections have described how CD-ROMs can be used to ensure consistency through periodic updates to the many different users of a reporting database. These databases do not, typically, make use of multimedia resources. In contrast, the additional facilities provided by the CD-ROM format tend to be exploited by some of the other publications that are generated by incident reporting systems. Hence, the sheer storage capacity of this medium provides means of disseminating textual incident databases. The access speeds supported by CD-ROM enable regulators to disseminate multimedia training presentations that are intended to guide safety managers and operators who must follow particular recommendations in the aftermath of near miss incidents and adverse occurrences.

The use of CD-ROMs to distribute information about adverse occurrences raises a number of further problems. In large, distributed organisations it may not be possible to provide all members of staff with access to personal computers that can play these disks. A number of innovative solutions have been devised to address this problem. One of these is illustrated by a staff training scheme that was developed by the Royal Adelaide Hospital. This scheme was closely tied to the Australian Incident Monitoring System, mentioned above. This hospital developed 'Mobi-ed' units that resemble the information booths that are found in public areas such as shopping malls and airports; 'the cabinet has solid wheels with brakes, a handle at the back for moving it around, and locks the computer behind two separate doors' [40]. They were based on standard desktop PCs. The booths also had the advantage that they could be left in common areas where a number of members of staff might have access to them during many different shift patterns. The units were deliberately moved between locations in the hospital; "this appearance and disappearance of the units encourages staff to check them out whenever they appear in their ward or work area" [40]. Multimedia training material was obtained to address specific training needs identified frm incident reports. The Mobi-Units also provided staff with access to an on-line tutorial about how and when to complete a submission to the AIMS reporting system.

It is important to recognise that CD-ROMs are simply a storage technology that supports the distribution of incident databases and training material. A number of deeper questions can, however, be raised about the effectiveness of the material that is distributed using this technology. Chapter 14.5 will analyse the effectiveness of incident databases. The following paragraphs provide a brief appraisal of multimedia training packages that are increasingly being developed by reporting agencies for dissemination on CD-ROM. We initially became interested in the effectiveness of this approach when developing materials for Strathclyde Regional Fire Brigade. Their training requirements are very similar to those of the healthcare professionals in the Royal Adelaide Hospital, mentioned above. Staff operate shift patterns and geographically distributed across a number of sites. There are also differences, for instance the training activities of a particular watch will be disrupted if they are called out in response to a call from a member of the public. Such circumstances increase the appeal of CD-ROM based training. The intention was to develop a series of multimedia courses that supported key tasks, which had caused particular concern during previous incidents. Fire-fighters could work through a course at their own pace. They could also suspend an activity and return to it after a call-out. There was also a perception that the use of computer-based technology might address the subjective concerns of staff who found conventional lectures to be ineffective. Figure 14.1 presents the results of a questionnaire that was issued to 27 staff within the Brigade, At the start of the project, they did not have access to any computer-based training. They received information about safety-related issues through paper publications, lectures and videos as well as drill-based instruction. However, many fire-fighters viewed 'real' incidents as an important means of acquiring new knowledge and reinforcing key skills. This finding has important health and safety implications. Incidents should reinforce training gained by other means. They are clearly not a satisfactory delivery mechanism for basic instruction.

As mentioned, a series of CD-ROM based training packages were developed to address the perceptions identified in Figure 14.1. These applications were developed in collaboration with the Brigade

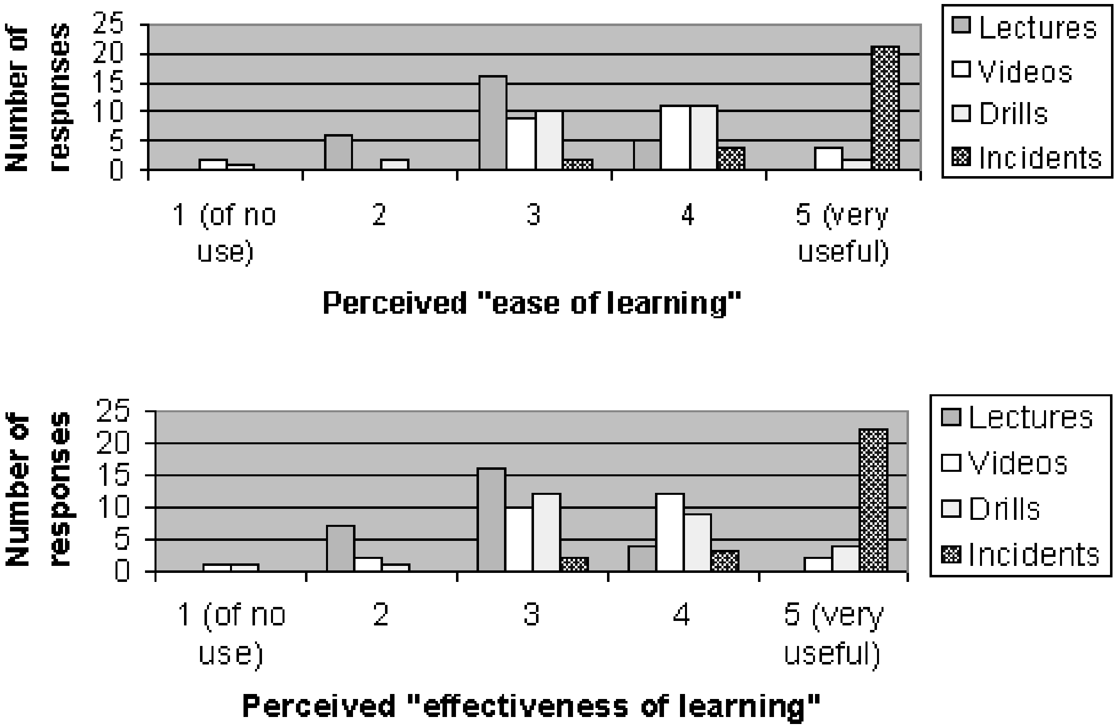


Figure 14.1: Perceived 'Ease of Learning' in a Regional Fire Brigade

training officer, Bill West, and two multimedia developers who were employed by the Brigade, Brian Mathers and Alan Thompson. Figure 14.2 illustrates one of these tools. This exploited the desktop virtual reality techniques, introduced in Chapter 7.3. Fire-fighters could use a mouse and keyboard to 'walk' into a Heavy Rescue Vehicle. They could then look inside equipment lockers and obtain a brief tutorial on the design and use of particular rescue devices. Previous incidents had illustrated the difficulty of providing officers with enough time to train on this particular vehicle given an operational requirement to keep it 'on call' as much as possible.

As mentioned, multimedia applications can be devised to address particular concerns that emerge in the aftermath of near miss incidents and adverse occurrences. The performance characteristics, in particular the access speeds of CD-ROMS, make this the favoured distribution media for such materials given network retrieval delays. It is important, however, to both understand and assess the various motivations that can persuade organisations to invest in tools such as that illustrated in Figure 14.2. In particular, we have argued that the strong motivational appeal of computer-based systems can support staff who find it difficult to be motivated by more traditional forms of training. We were, however, concerned that the introduction of computer-based techniques should not compromise particular learning objectives. We, therefore, conducted an evaluation that contrasted a computer-based system with more traditional techniques.

A matched subject design was adopted; each fire-fighter was paired with another officer of equivalent rank and each member of the pair was then randomly assigned to one of two groups. Both groups were given access to the same computer based training package on techniques to support the effective application of foam to combat particular types of fire. The technology used to produce the interactive application was similar to that used in Figure 14.2. One group was then given a CD-ROM based Comprehension Tool. This guided the officers through a series of questions about the training material and provided immediate feedback if any problems were diagnosed [425]. The

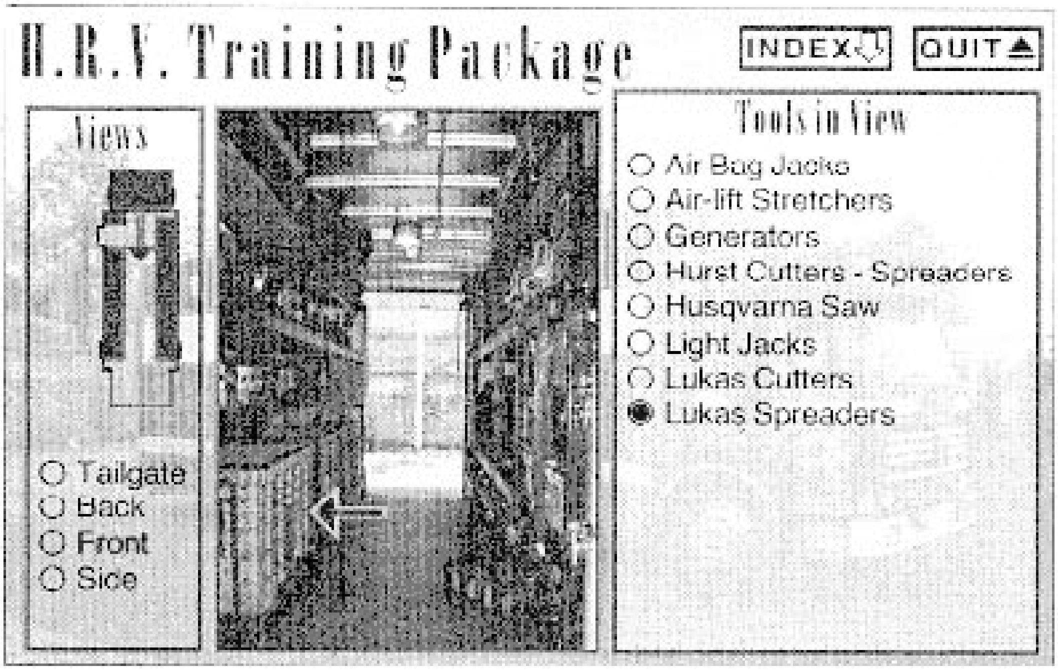


Figure 14.2: The Heavy Rescue Vehicle Training Package

second group was given a pencil and paper test without any feedback about the accuracy of their responses. One week later both groups were re-tested using the Comprehension Tool. It was hypothesised that the group that had previous access to the CD-ROM based self-assessment tool would achieve significantly higher scores than the group that had performed the pencil and paper test. A weakness in this experimental design is that learning effects might improve the results of the group that already had some experience with the Comprehension Tool. These effects were minimised by ensuring that both groups were entirely confident in the use of the tool before the second test began. Tables 14.6 and 14.7 present the results obtained for the two groups involved in this evaluation.

Rank	Number	Comprehension Tool or Paper test?	Score 1	Score 2
Sub officer	1	Paper	72	76
Leading fire-fighter	2	Comprehension tool	60	56
Leading fire-fighter	3	Comprehension tool	80	68
Fire-fighter	4	Paper	36	44
Fire-fighter	5	Paper	88	84
Fire-fighter	6	Comprehension tool	64	52
Fire-fighter	7	Paper	40	40
Fire-fighter	8	Comprehension tool	52	32

Table 14.6: Results for the first group of Fire Fighters

This evaluation forms part of a far wider attempt to validate the use of computer-based learning techniques. It does, however, provide a case study in the problems that can arise during these validation exercises. For instance, subtle differences in position 1 of the ranking schemes in Figures 14.6 and 14.7 complicate the task of making accurate comparisons. The station officer's performance is better than that of the sub-officer. There differences reflect the operating characteristics and compo-

Rank	Number	Comprehension Tool or Paper test?	Score 1	Score 2
Station officer	1	Comprehension tool	96	88*
Leading fire-fighter	2	Paper	80	60*
Leading fire-fighter	3	Paper	92	80
Fire-fighter	4	Comprehension tool	68	48
Fire-fighter	5	Comprehension tool	64	68
Fire-fighter	6	Paper	52	40*
Fire-fighter	7	Comprehension tool	84	64
Fire-fighter	8	Paper	60	60

Table 14.7: Results for the second group of Fire Fighters (* post fire)

sition of this organisation; they could not simply be changed for experimental expediency. Further problems arose because the second group of fire-fighters was called out while we administered the retest. It would have been unethical to prevent them from responding until after they had completed the evaluation! This study also provided some direct insights into the use of computer-based training techniques. Statistical T-tests failed to show any significant differences in the re-test scores between those who had access to the CD-ROM based tool and those who sat the pencil and paper test. We were unable to establish that the computer-based tool was any better than the more traditional techniques. Or put another way, it was no worse than existing methods for the dissemination of safety-related information.

The previous paragraphs are intended to correct the euphoria that often promotes the use of CD-ROM based technology for the publication of safety-related training materials. Often managerial and political pressures encourage the use of 'leading edge' technology without any careful analysis of whether this technology will support key learning objectives. Cost constraints can also act to limit many organisation's ability to disseminate the insights gained from previous incidents and accidents in any other format. Some regulatory and investigatory bodies have, however, continued to resist these pressures. For example, the State Training Teams of the FDA's Office of Regulatory Affairs support a 'lending library' of courses that must be presented by "trained state/federal facilitators that have already completed the original satellite course" [267]. These trained mentors are support by course videos, books, exams and answer key forms. It is interesting to note that these courses cover topics that are perceived to have a relatively high degree of importance in the FDA's regulatory role. For example, they include two courses on the investigation and reporting of adverse occurrences. This might imply that many of the issues addressed in previous chapters of this book cannot easily be taught using computer-based techniques.

Local and Wide Area Networks

The previous section has identified some of the strengths and weaknesses of CD-ROM technology for disseminating the information that can be derived from incident reporting schemes. They can be used to distribute the multimedia training resources that are intended to address previous failures. It can, however, be difficult to demonstrate the effectiveness of these resources. In contrast, CD-ROMs offer numerous benefits for the distribution of incident databases. They are relatively cheap. They offer relatively high storage capacity together with a relatively compact, lightweight format that is rugged enough to survive most postal services. Data can also be encrypted to provide additional security should a CD-ROM be lost or intercepted. There are, however, a number of limitations with this use of CD-ROM technology. In particular, it can be difficult to use this approach to issue more immediate updates to safety-related information. We have already describe the delays that can be introduced through the use of postal services to distribute physical media, such as CD-ROM. In contrast, many organisations are increasingly using computer networks to support the more rapid dissemination of information about adverse occurrences and near miss incidents. The MDA provide an example of this in their Business Plan for 2001-2002. They express the intention to develop closer links with the National Health Service and the Commission for Health Improvement with the objective of 'improving the dissemination' of information about adverse events. This will be done by 'electronic dissemination through our website and other Internet systems, so that healthcare professionals will increasingly have important safety information at their fingertips' [545].

It is convenient to identify two different sorts of incident data that can be accessed over computer networks. Firstly, incident databases help to collate information about large numbers of adverse occurrences and near misses. Secondly, incident libraries provide access to small numbers of analytical reports that may summarise the findings from many different incidents. In either case, these electronic documents must be stored in a particular format if they are to be disseminated across computer networks. Chapter 13.5 described the strengths and weaknesses of two of these formats. Hypertext Markup Language (HTML) documents can be viewed using standard browsers and are easily indexed by search engines but cannot easily be printed. Adobe's Portable Display Format (PDF) avoids this problem but most search engines have to be adapted to search this proprietary format for the keywords that are then used when users issue search requests. Incident data can also be stored in the file format that are supported by commercial mass-market databases and spreadsheets. This approach tends to be associated with incident databases. They are used to provide access to summary data about large numbers of individual incidents. PDF and HTML are more commonly used to support the dissemination of analytical surveys and the longer reports that are contained in on-line 'reading rooms'. The distinction between on-line libraries and incident databases is significant not simply because it influences file formats and retrieval techniques but also because it reflects important distinctions in the policies that determine what is, and what is not, made available over computer networks.

Private Databases and Public Libraries. Some organisations maintain private electronic databases that are not mounted on machines that are accessible to a wider audience. These same organisations may, however, provide wider access to the libraries of reports and recommendations that are derived from these private databases. This approach is currently being exploited by the MDA; 'we plan to introduce web reporting facilities that will feed directly into the Adverse Incident Tracking System (AITS) software' [545]. AITS is intended to help the Agency keep all its main records in electronic form for 'action and archiving'. It will also provide MDA staff with 'flexible data analysis tools to identify trends and clusters of incidents and that will enable us to adopt a more pro-active approach to reducing adverse incidents'. It is not intended that other organisations should have access to this database. In contrast, electronic access will be provided to what the previous paragraphs have characterised as 'libraries' of analytical overview documents. In passing, it should be stressed that the technical details of the AIMS software have not been released, not is there a detailed account of the full system development plan. It may very well be that the objectives outlined in the 2001-2002 Business Plan will be revised as AIMS is implemented.

Public Databases and Public Libraries. Other organisations provide access both to 'reading rooms' of analysis and to the databases of incidents that are used to inform these more analytical accounts. For example, the FDA's Manufacturer and User Facility Device Experience Database is freely available over the Internet [270]. It is comparable to AITS because it records voluntary reports of adverse events involving medical devices. An on-line search is available which allows you to search the Centre for Devices and Radiological Health's database of incident records. It is also possible to download the data in this collection over the Internet. These files are updated every three months. They are in a text format that enables safety managers and other potential readers to import them into a commercial database or word processor for further analysis. At the same time, the information in the MAUDE system is also used to inform more detailed incident investigations and surveys of common features across several adverse occurrences. The resulting reports are also available on-line in PDF format via an electronic 'reading room' [273]. This open dissemination policy enables readers to examine the warnings that are contained in a particular safety issue or alert publication. They can then trace additional details about particular incidents, and related occurrences, using the MAUDE database. It is important to stress, however, that the provision of public databases and reading rooms need not imply that sponsor organisations do not also maintain more private systems that are not made available in the manner described above.

Private Database and Private Summaries. Some incident reporting systems restrict access to both their database information and the reports that are derived from them. This policy is reflected in the way in which the AIMS system restricts access to its central database and only provides feedback on comparative performance to the individual units that participate in the scheme. These private summaries can be distributed over networks, either using the e-mail systems that have been described in previous paragraphs or using more explicit forms of file transfer [187]. This approach is entirely understandable given the sensitive nature of incident reporting within individual hospitals. There are other circumstances in which computer networks have been developed to support incident investigation and analysis within an organisation. The intention has never been that the data should be made public but that it should support specific tasks and objectives within the particular teams that must act upon incident data. This is illustrated by the U.S. Department of Health and Human Services' PulseNet system [261]. This system was established to distribute information generated from a molecular technique, pulsed-field gel electrophoresis (PFGE), that can be used to identify similarities between different samples of E. coli O157:H7. The PFGE technique was first used during a food-borne illness in 1993. This enabled laboratories in different locations to determine that they were fighting a common strain of bacteria. The lack of efficient computer networks to distribute the information from the independent PFGE tests prevented analysts from identifying these common features for the first week of the outbreak. Seven hundred people became ill and four children died in the outbreak. PulseNet is intended to reduce the interval taken to detect future incidents down to approximately 48 hours. This system is not intended to support the public dissemination of information about such events. It is, however, intended to support the analytical and decision making tasks that are necessary in order to detect common features between apparently isolated incidents.

It is important to emphasise that the increasing use of computer networks in incident reporting is only a small part of a wider move to intergrate diverse information sources about potential hazards. This integration is intended to support decision making. In other words, the dissemination of incident-related information is not an end in itself. This is illustrated by the manner in which epidemiologists can use PulseNet to trace common features in E. coli outbreaks. It is also illustrated by recent attempts to integrate diverse Federal databases to support the FDA field officers that have to determine whether or not to admit medical devices into the United States of America. The intention behind this initiative is to provide officers with rapid access to the range of data that they require in order to reach a decision. This data includes information about any previous adverse events involving particular products. However, this is only one part of a more complex set of requirements. For example, officers will also have to access the FDA's Operational and Administrative System for Import Support (OASIS) and Customs' Automated Commercial System (ACS). These data sources can be used to identify whether the product violates particular regulations by virtue of its point of origin. They can also be used to determine whether or not previous samples conformed with regulations when explicitly tested by the FDA. There is not intention that this integrated system should be widely accessible over public computer networks. It is, however, possible to access some of these information sources. For instance, it is possible to view the safety alerts that apply to particular products over the World Wide Web. This provides an example of the flexibility that such computer networks can offer for the provision of safety-related information. Users can choose whether to view warning that are sorted by particular industries, by country of origin or by FDA reference number. Users can also conduct free-text searches over the database of import alerts.

The FDA's import system contradicts the binary distinction between public and private distribution that was made in previous paragraphs. In practise, computer networks enable their users to make fine grained decisions about who can and who cannot access incident information. In the case of the import system, full functionality is reserved for field officers. Individual elements of the entire system are, however, made available for the public to access over the Internet. The same techniques can also be used more generally to restrict access to particular information about previous incidents. These approaches implement access control policies. The most common approach is to erect a 'firewall' that attempts to prevent access from anyone who is not within the local network that hosts the system. The following section discusses some of the consequences that such measures have for the implementation and maintenance of incident reporting systems. In particular, it is argued that compromises must often be made between restricting access and simplifying the procedures that users must follow in order to obtain access to incident data.

14.3.2 Access Control

Security deals with the unauthorised use and access to the hardware and software resources of a computer system. For example, *unauthorized disclosure* occurs when a individual or group can read information that they should not have access to. They, in turn, can then pass on information to other unauthorized parties. For instance, an unauthorised party might pass on information about an adverse occurrence to the press or broadcast media before that incident has been fully investigated. *Unauthorised modification* occurs when an unauthorised individual or group can alter information. They might have permission to 'read' data items but this does not automatically imply that they should also be able to modify data. It is, therefore, important to distinguish between different levels of permission. For example, an individual hospital contributing incident reports to a central database may have permission to access and modify their own reports. They might, in contrast, only be able to read reports from other hospitals without being able to modify them. Finally, *unauthorised denial of service* occurs when an individual or group can shut-down a system without authority for taking such an action. Unauthorised denial of service is a general problem in computer security. For example, the propagation of viruses can deny other applications of the computational resources that they require. I am unaware of any specific instances in which this form of attack has been a particular problem for incident reporting. It is important to stress, however, that unauthorised denial of service could have potentially profound consequences as incident reporting system become more tightly integrated into complex decision support systems, such as the FDA's Import application.

The issue of security affects incident reporting systems in a number of ways. For example, the Central Cardiac Audit Database (CCAD) project identifies two main threats [185]. The first centres on the security of data during transmission. When data is transmitted across open networks, such as the Internet, it can be intercepted unless it is encrypted. The second set of security concerns centres on controlling access to incident information after it has been collated. This is important because it is often necessary to ensure that different individuals and groups have different degrees of access to sensitive information. Some may be denied access to particular records. Other groups may be entitles to read data without being able to modify or 'write' it.

	Unit 1's Reports	Unit 2's Reports	Unit N's Reports
Administrator	read, write	read, write	read, write
Regulator	read	read	read
Unit 1	read, write	read	read
Unit 2	read	read, write	read
Unit N	read	read	read, write

Table 14.8: General Form for an Access Control Matrix

The distinction between 'read' and 'write' permissions has led to the development of access control policies. In their simplest form, these techniques implement a matrix that associates particular privileges with the users of a system and the objects that are held by that system. This is illustrated by Table 14.8. As can be seen, system administrators must be able to access and modify the reports that are submitted from all of the units that contribute to a reporting system. This requirement is, typically, enforced so that they can implement any revisions or updates that may subsequently prove to be necessary for the maintenance of the system. For example, they can automatically insert additional fields into the record of an incident. External regulators, in this instance, are provided with read-only access to all reports. Each of the contributing units can also read the reports from other contributors. They can also modify their own information. It is important to stress that the exact form of an access control matrix depends upon the nature of the reporting system. For example, some applications only provide read access to its contributors. They cannot modify their own data and all updates must be performed through an administrator who is entirely responsible for any 'write' actions. This reflects elements of the AIMS approach. In this case, the access control matrix would only contain write entries in the Administrator row. It is also important to emphasise that access is only granted if it is explicitly indicated in the matrix. By default, all other permissions are denied. In Table 14.8, the general public would not have any right to obtain or modify incident data.

Access control matrices are explicitly embodied within many of the more sophisticated software applications that have been developed to support incident reporting schemes. When a user makes a request to access a particular item of information, the system identifies the row associated with that user in the matrix. It then looks along the columns until it finds an entry associated with the object of the request. If the user does not have the relevant permissions then the request is denied. Unfortunately, this approach is not a feature of single-user systems. In general, access control makes little sense when there is only one row in the matrix. This has important consequences for many reporting systems that continue to use mass-market, desktop applications to support the dissemination of incident information. Single-user spreadsheets and databases, typically, have no means of making the fine grained distinctions implied by Table 14.8. In consequence, if a user is granted access to the system then they have complete permission to access all data. It is, typically, possible to apply locking techniques to the information that is held by these systems. This prevents unauthorised modification. However, this 'all or nothing' approach is usually too restrictive for large-scale systems [187].

Access control matrices provide numerous benefits to incident reporting systems. They explicitly represent the security policy that is to be enforced during the distribution of potential sensitive information. They are not, however, a panacea. As we have seen, it is entirely possible for individuals or groups to abuse their access permissions. For example, Unit 1 might pass on information about Unit 2 to a third party that is not entitled to this access, according to Table 14.8. In such circumstance, it is possible for system administrators to identify the potential sources of any 'leak' by inspecting the entries in the column that is associated with any disclosed information. Table 14.8 makes a number of simplifying assumptions. For instance, we have not considered 'grant privileges'. These enable particular users or groups to provide access permissions on certain objects. This is most often necessary when new Units join the system. The administrator would have to ensure that they were granted permission to read the contributions from all of the other Units. The entries associated with the system administrators would, therefore, be revised to *read, write, grant.* Paradoxically, the ability yo grant access also implies the ability to remove or deny access permissions. For instance, if incident information were being leaked to a third party then administrators might take the decision to remove all read permissions except those that apply to the Unit that contributed particular reports.

14.3.3 Security and Encryption

Access control matrices define the policy that is to be followed in the distribution and modification of incident information. In order to implement such a policy, most software systems rely upon encryption algorithms. As might be expected, these techniques take the original document, or plain text, and produce a cipher. Ideally, it should not be possible for an unauthorised person or group to derive the plain text from the cipher. One means of helping to prevent this is to create an encryption algorithm that relies not only on the plain text but also an additional input parameter known as a key. This can be illustrates by Caesar's algorithm. Caesar's algorithm replaces each letter in the plain text with the next letter in the alphabet. The letter 'a' would be replaced by 'b' in the cipher, 'c' would be replaced by 'd' and so on. This is a relatively simple algorithm to guess and so we might require that the user also supplies a key. The key could be the number of places that each letter is offset. For example, in order to decipher the following phrase we must know that each letter has been shifted by 14 places in the alphabet:

Cipher:	VAPVQRAG ERCBEGVAT
Plain text:	INCIDENT REPORTING

This simple algorithm is vulnerable to many different forms of attack. For instance, we can examine letter frequencies in the cipher to make guesses about the identity of particular letters. It does, however, illustrate the key features of *secret or private key* encryption. In order for this approach to work, it is important that the key is never disclosed to unauthorised individuals. The previous example also illustrates further aspects of secret key encryption. For instance, this approach can be strengthened by choosing a suitably complex algorithm. Alternatively, it is often more convenient to choose a relatively simple algorithm but a very complex key. Without the key, even if the algorithm is well known, it can be extremely difficult for unauthorised individuals to decipher a message. It is for this reason that encryption software often emphasizes the size of the keys that they support, 64 or 128 bits for example.

Secret key encryption is a feature of many national reporting systems. Regulatory and investigatory agencies provide each unit in the system with the encryption software and a password. This is then used to protect incident reports when they are transmitted across computer networks. This approach was adopted by the APSF's AIMS system when they moved from paper to electronic submissions [35]. One problem with secret key encryption is that it can be difficult to secure the dissemination of keys. They cannot be transmitted over the computer network because this would defeat attempts to secure transmission over that network. Conversely, it is impossible to secure the network until the secret key has been agreed upon.

Public key encryption provides an alternative to secret key encryption. This relies upon algorithms that require different keys to encode and decode the plain text. Typically, users distribute a public key to anyone who might want to send them secure information. They can do this because they know that this key can only be used to encode data. Only they have access to the second private key that is required in order to read any message. This is the approach adopted by the 'Pretty Good Privacy' or PGP mechanisms that are widely available over the Internet. PGP is one of several systems that are recommended for transmission of data to the Central Cardiac Audit Database, mentioned earlier [185]. The PGP package provides a variety of utilities for the generation, management and use of encryption keys. It has the advantage of being low or no cost and is widely available. Secure/Multipurpose Internet Mail Extensions (S/MIME) is an application of public key encryption to the MIME technology that is widely used for the dissemination of incident data. This approach is supported by recent versions of Netscape Communicator and Internet Explorer. S/MIME applies encryption to individual files that are mailed from one machine to another. At is also possible to create secure links that encrypt information passing between two or more machines. This approach is deliberately designed to support more interactive forms of communication, such as web browsing, where information can be passing in both directions over the connection. The best known application of this approach is known as the Secure Socket Layer (SSL) protocol. This applies public key encryption over an entire session rather an individual item of mail.

PGP, S/MIME and SSL have all been used to secure the data that is transmitted between the contributors of incident reports and regulatory or investigatory agencies [274, 185]. It is important to emphasise that these technologies do not provide any absolute guarantees about the security of any electronic communication. It is theoretically possible to break most implementations. The technical expertise and computation resources do make it extremely unlikely that this will occur, at least in the short term. These observations re-iterate an important concept; it is seldom possible to achieve absolute security. Safety managers must adopt an informed approach to risk assessment. The degree of technological sophistication applied to secure incident data must be proportional to the sensitivity of that data. It is important, however, that these issues are explicitly considered as more and more reporting systems use computer-based networks as a cheap and effective means of disseminating information about near misses and adverse occurrences [380].

One way in which cryptography has been used by incident reporting systems is to support digital signatures. A digital signature is a means of encoding a message in such a manner that it authenticates the sender's identity. This is important if regulators and investigation agencies are to ensure that reports of an incident have not been sent for malicious reasons. Both private and secret key techniques can be used to implement digital signatures. The fact that the recipient can decode a message that was encoded using the secret key agreed with the sender might, at first sight, seem to be sufficient for a secret key implementation. No other person should know the secret key. Unfortunately, this is vulnerable for a number of reasons. This approach is vulnerable to a replay attack in which a previous message is saved and later resent by some unauthorised agent. Similarly, some portion of a previous message may be cut and pasted to form a new message. This is feasible because it is possible to make inferences about the contents of a message even even if it is impossible to completely decipher all of its contents. For these reasons, secret key implementations of digital signatures usually also encode characteristics of the entire message, such as the date when it was sent and the number of characterise in the plain text. When the message has been decoded the recipient can check this additional information to ensure the integrity of the content. This technique can also be used when the message is not, itself, encoded. A signature block can be encrypted at the end of the message. Again, the recipient can decode the signature block and use the techniques described above to establish its authenticity. This approach has given rise to a range of more elaborate techniques that support the concept of *electronic watermarks* .

Public key implementations of a digital signature can be slightly more complicated. For instance, a contributor might encrypt a message using it's secret key. It will then encrypt the results of this encryption using the public key of the regulator. The message is then sent over a computer-based network. The regulator first decodes the message using their secret key. Ideally, no other users can complete this first step assuming that the regulators secret key is not compromised. Next, the regulator can apply the contributor's public key to extract the plain text. The regulator knows that the contributor sent the message because only the contributor has access to their secret key.

It might seem that such details are a long way removed from the practical issues that must be considered in the development and operation of incident reporting systems. The key point about digital signatures is that they enable organisations to transmit information in a secure manner that can be granted the same legal status as conventional, paper-based documents. For instance, in 1997 the FDA issued regulations that identified the criteria that would have to be met for the use 'of electronic records, electronic signatures and handwritten signatures executed to electronic records as equivalent to paper records and handwritten signatures executed on paper' [274]. These regulations applied to all FDA program areas and were intended to support the widest possible use of electronic technology 'compatible with FDA's responsibility to promote and protect public health'. The FDA requirements illustrate the importance of understanding some of the concepts that have been introduced in previous paragraphs. For instance, Section 11.70 requires that 'electronic signatures and handwritten signatures executed to electronic records must be linked to their respective records so that signatures cannot be excised, copied, or otherwise transferred to falsify an electronic record by ordinary means'. Such concerns have motivated the development of message-specific signature blocks mentioned above.

There are many reasons why incident reporting systems are forced to introduce some of these more advanced security measures. The FDA regulations reflect a concern to ensure that electronic reports of adverse occurrences have the same legal status as their paper-based counterparts. These more advanced security techniques can also be implemented in response to the concerns expressed by potential contributors. There is often fear that an individual's identity will be revealed. Similarly, safety managers in the healthcare industry must also respect patient confidentiality. There is, however, usually a trade-off to be made between the security of a system and the ease with which it can be used by its operators. For example, the use of cryptographic techniques implies a considerable managerial overhead both on the part of systems administrators and the users who must remember the passwords that protect and sign their information. Recent studies have suggested that for every user of a system there is a request for support staff to reset a forgotten password every three-four months [1].

It is difficult to under-emphasise the human element in any secure system. For example, recent attempts to introduce public and private key cryptography into one national reporting system produced a series of responses from potential contributors. They commented that these human issues would compromise the most advanced technology. It is possible for an owner to 'lend' a file, as a collaborative fraudulent gesture, or to unwittingly assist a fraudulent colleague in an 'emergency'. The FDA has acknowledged that 'such fraudulent activity is possible and that people determined to falsify records may find a means to do so despite whatever technology or preventive measures are in place' [274]. There are also more 'mundane' threats to the security of incident reporting systems. Previous research has suggested that people are often very lax in their selection of passwords [881]. The use of recognisable words, of names of friends, of addresses makes the entire system vulnerable to dictionary attacks. These take the form of repeated requests to access a system where a different word or phrase is supplied from an electronic dictionary in response to each password request. Eventually this form of attack will succeed unless care has been taken in selecting a password. Further problems arise when passwords are distributed to friends and colleagues. More commonly, however, systems are compromised by simply leaving a machine connected to the network. Users issue the requested password and then leave the machine unattended. I recently saw an example of this on a hospital visit. A screen-saver was used with the warning 'unauthorised access on this machine is prohibited'. Anyone who ignored this message could have directly accessed and altered the patient records for that ward while the nurse was away from her station.

Mnny of the usability problems that affect secure systems can be reduced through the use of biometric authentication. These techniques avoid the need for users to remember arbitrary passwords. They include the use of fingerprints, retinal patterns, signatures and voice recognition [441] This technology has not yet been widely used to secure the transmission of incident data. It is, however, likely that it will be used within future systems. Part of the reason for this is apparent in recent observations made by the Central Cardiac Audit Database Project. They argue that such security techniques appear to be 'an extremely expensive solution to a non-problem, but public fears about Internet security, mostly unfounded but encouraged by the popular media, would need to be allayed before widespread medical data transmission via the Internet would be acceptable' [185].

14.3.4 Accessibility

'Accessibility' can be thought of as the converse of access control. Just as it is important to ensure that unauthorised people are denied access to an incident report, it is equally important that authorised individuals can obtain necessary information. This implies that any computer-based resource should be evaluated to ensure that the human-computer interface does not embody inappropriate assumptions about the potential users of such as system. This is particularly important because the computer-based dissemination of incident reports can offer particular advantages to certain groups of users providing that their requirements are considered during the early stages of systems development. People with visual disabilities can use a range of computer-based systems to access incident databases in a manner that cannot easily be supported using paper-based techniques. Unfortunately, the use of icons and complex menu structures can prevent many users from exploiting screen readers and similar devices. It is for this reason that many organisations publish minimum standards in this area, such as Federal Regulations Section 508 on the accessibility of electronic information. In order to satisfy these requirements it is important that the developers of incident reporting systems provide some means for users to communicate any difficulties they might have experiences in access their data. This can, however, create a recursive problem in which the users of the information resource cannot access the information resource in order to learn of alternative format or other forms of help:

> "The U.S. Department of Health and Human Services is committed to making its web sites accessible to all users. If you use assistive technology (such as a Braille reader, a screen reader, TTY, etc.) and the format of any material on our web sites interfere with your ability to access the information, please use the following points of contact for assistance. To enable us to respond in a manner most helpful to you, please indicate the nature of your accessibility problem, the preferred format in which to receive the material, the web address of the requested material, and your contact information..." [31]

Although Section 508 of the US Accessibility Act focuses on users with special needs, there is also a more general requirement to ensure that people can access incident information. In consequence, observational studies and laboratory-based evaluations may be conducted to ensure that users can operate computer-based information systems. This implies that designers must consider the previous expertise of their users and of their ability to exploit particular human-computer interaction techniques. Brevity prevents a more detailed introduction to the design and evaluation of interactive computer systems in general. Preece et al provide a survey of techniques in this area [687]. In contrast, the following pages focus more narrowly on techniques that can be used to identify patterns of failure in large-scale collections of incident reports.

14.4 Computer-Based Search and Retrieval

Previous sections have considered the dissemination of information about individual incidents. In contrast, this section focuses more narrowly on the problems that arise when providing access to databases of previous incidents over computer networks. Recent technological innovations, often associated with mass-market applications of the World Wide Web, are creating new opportunities for rapidly searching large number of incident reports that can be held in many different countries across the globe. Before looking in more detail at these 'leading edge' systems, it is first necessary t understand why organisations are exploiting computer-based dissemination techniques for their incident databases.

The sheer scale of many reporting systems motivates the use of electronic dissemination techniques for incident databases. The number of incident reports that are submitted to a system can accumulate rapidly over a relatively short period of time, even in local or highly specialised systems. For instance, the Hyperbaric Incident Monitoring Study was started in 1992 to collects reports of incidents and near misses in hyperbaric medicine. Hyperbaric Oxygen Therapy is the most common form of this treatment. The patient enters a chamber that is filled with compressed air until a required pressure is reached. The patient breaths 'pure' oxygen through a mask or a transparent hood. In addition to diving recompression, this techniques has been used in the treatment carbon monoxide poisoning, wound healing and post radiation problems. This system is currently operated through the APSF, the same organisation that maintains AIMS. The Hyperbaric Incident Monitoring Study was launched internationally in 1996 and the associated forms have been translated into 4 different languages. By early 2001, there were some 900 reported incidents in the database [38]. This partly reflects the success of this system. It also creates considerable practical problems. The costs associated with maintaining even a relatively simple paper-based indexing system would be prohibitive. It would also be difficult for other organisations to access this data without replicating each paper record or posing a succession of questions to the staff who are responsible for maintaining the paper indexing system.

It is feasible but unlikely that the Hyperbaric Incident Monitoring Study database could be implemented using paper-based techniques. In contrast, other national reporting databases could not be maintained without electronic support. Firstly, the sheer volume of reports makes it essential that some form of database be used to collate and search that data. Secondly, the large number of individuals and groups who might legitimately want to retrieve incident information increase the motivation to provide access to these databases over computer networks. For instance, the FDA's Adverse Event Reporting System (AERS) for adverse events involving drugs and therapeutic biological products contains more than 2 million reports. Table 14.9 provides a break-down of the number of incidents that are entered into the FDA's databases within a single year. The Center for Biologics Evaluation and Research's Error and Accidents Reporting System records incidents that occur in the manufacture of biological products. An error or accident is a deviation from the 'good manufacturing practice' set down by FDA regulations. The Drug Quality Reporting System receives reports of similar incidents that affect the manufacturing or shipping of prescription and over-the-counter drug products. These incidents can result in problems for the formulation, packaging, or labeling of these products. Post-marketing surveillance for vaccines is handled by the Vaccine Adverse Event Reporting System. Approximately 15 percent of the reports describe a serious event, defined as either fatal, life-threatening, or resulting in hospitalization or permanent disability. The Manufacturer and User Device Experience (MAUDE) Database receives between 80,000 and 85,000 reports per year. The 1984 Medical Devices Reporting regulation required manufacturers to report device-related adverse events to the FDA. In 1990, the Safe Medical Devices Act extended this regulatory structure to include user facilities such as hospitals and nursing homes. Serious injuries that are device-related must be reported to their manufacturers. Fatalities must be reported both to the manufacturer and directly to the FDA. The MAUDE database was established in 1995 to support the Safe Medical Device Act and now contains more than 300,000 reports. Another 500,000 reports are collected in a pre-1995 database. Finally, table 14.9 records that the FDA's risk-based summary reporting system receives some 30,000 reports per annum. Products that are approved for this summary reporting process are 'well known' and have a 'well-documented' adverse event history [278]. This approach involves the periodic submission of adverse event statistics in a tabular form that yields 'economies for both the devices industry and FDA'.

Reporting System	No. of reports per annum
Adverse Event Reporting System	230,000
(CBER) Biologics Error and Accidents Reporting System	13,000
Drug Quality Reporting System	2,500
Vaccine Adverse Event Reporting System	12,000
Manufacturer and User Device Experience	80,000
Risk-Based Summary reports	30,000

Table 14.9: Annual Number of Incidents Included in FDA Databases

The volume of data that can be gathered by successful national systems justifies the use of information technology. However, the use of this technology does not provide a panacea for the management of large-scale databases. For instance, the cost implications and the requirement to use specialist hardware and software often convinces many public or Federal agencies to involve contractors to run these systems. The FDA's Drug Quality Reporting System database, mentioned above, is run in this manner. FDA staff interact with the system via an on-line interface that is intended to help them pose particular queries or questions that are then relayed to the database, which is administered by the contract organisation. The management of the Vaccine Adverse Event Reporting System database is even more complex. This is jointly administered by the FDA's Center for Biologics' Division of Biostatistics and Epidemiology and the Centers for Disease Control and Prevention, Vaccine Safety Activity, National Immunization Program. Representatives of both agencies oversee data processing and database management that is again performed by a contractor. Such complex relationships occur in other industries. For instance, the ASRS is largely funded by the FAA. NASA manages the system, which is in turn operated under contract by the Battelle Memorial Institute. In most cases, these relationships have proven to be highly successful. There are, however, considerable problems in ensuring that technological requirements are accurately communicated between all of the stake-holders in such complex, interactive applications. In consequence, previous studies have revealed considerable frustration from users who feel that many incident databases no longer support all of the retrieval tasks that they must perform [472, 416].

Further problems affect the management of large-scale incident databases. For example, it is often convenient for national regulators to establish a number of different schemes that focus upon particular incidents. For instance, Table 14.9 summarises the different schemes that are operated by the FDA, This can create problems because the same incident can fall within the scope of more than one database. This problem can be exacerbated if different agencies also run apparently complementary reporting systems. For instance, the FDA has to monitor medication error reports that are forwarded by clinical staff to the United States Pharmacopeia and to the Institute for Safe Medication Practices. Some of these incidents are then incorporated into the FDA's own databases. Similarly, the FDA must also review the medical device reports that are submitted to the MEDWATCH programme in case they have any relevance for possible medication errors. In addition to all of the systems mentioned in Table 14.9, the FDA also maintains a central database for all reports involving a medication error or potential medication error. This contains some 7,000 reports. In total contrast to this amalgamated database, the FDA's Vaccine Adverse Event Reporting database is entirely 'independent of other FDA spontaneous reporting systems' [278]. This diversity creates a flexible approach that is tailored to the various industries which are served by the FDA. It also creates considerable managerial and technical problems for those individuals who must support the exchange of data both within and between the various systems. For instance, some of these systems provide public access to incident data. There is a web-based search engine that can be used to find the detailed records held in the MAUDE system. Other applications are strictly confidential and no public access is provided. This can create problems if incident data is transferred from one system to the other. Confidential reports can be made public if they are transferred to the open system. Alternatively, if potentially relevant reports are not disclosed then valuable device-related safety information will be withheld and the MAUDE data will be incomplete. This would create doubts about the value of the system as a means of tracing more general patterns from information about previous incidents.

Further legal and ethical issues complicate the use of on-line systems to help search through incident databases. For example, a number of countries now have powerful disclosure laws that enable people to access databases in the aftermath of near miss incidents and adverse occurrences. This provides an opportunity to search through the database and identify previous failures with similar causes. In subsequent litigation, it might then be argued that responsible organisations had not shown due care because they had failed to learn from those previous incidents. Several safety managers have described their concerns over this scenario during the preparation for this book. One even commented that their company was considering deleting all records about previous incidents. These concerns are partly motivated by the difficulties that many organisations have in storing and retrieving information about previous incidents. Incidents often recur not because individuals and organisations are unwilling to learn from previous failures but because they lack the necessary technological support to identify patterns amongst thousands of previous incident reports. For instance, many users of incident databases cannot accurately interpret the information that is provided by database systems. It is also difficult for safety managers to form the commands that are necessary to retrieve information about particular types of incident. The following sections describe these problems in greater detail and a number of technological solutions are proposed.

14.4.1 Relational Data Bases

There are two central tasks that users wish to perform with large-scale incident databases. These two tasks are almost contradictory in terms of the software requirements that they impose. On the one hand, there is a managerial and regulatory need to produce statistics that provide an overview of how certain types of failures are reduced in response to their actions. On the other hand, there is a more general requirement to identify common features amongst incident reports that should be addressed by those actions in the first place. The extraction of statistical information typically relies upon highly-typed data so that each incident can be classified as unambiguously belonging to particular categories. In contrast, the more analytical uses of incident reporting systems involve people being able to explore alternative hypotheses about the underlying causes of many failures. This, in turn, depends upon less directed forms of search. It is difficult to envisage how investigatory bodies could construct a classification scheme to reflect all of the possible causal and mitigating factors that might arise during the lifetime of a reporting system. Unfortunately, most schemes focus on the development of incident taxonomies to support statistical analysis. Relatively, few support the more open analytical activities, described above. This is reflected in the way in which most reporting systems currently rely upon relational databases.

As the name suggests, relational database techniques build on the concept of a relation. This can be thought of as a table that holds rows of similar values. For example, one table might be used to record values associated with the contributor of an incident report. The cells in the table might be used to store their name, their contact information, the date when they submitted a report and so on. Each row of the table would represent a different contributor. Of course, each contributor might make several incident reports and so another table would be used to hold this relation. The columns in this table might include the name or identifier of the person making the report, the date of the report, the plausible worst case estimate of the severity of the incident and so on. Each row in this incident table would store infromation about a different report.

Relational database offer a number of important benefits for the engineering of incident databases. One of the most important of these is the relative simplicity of the underlying concept of a relation.

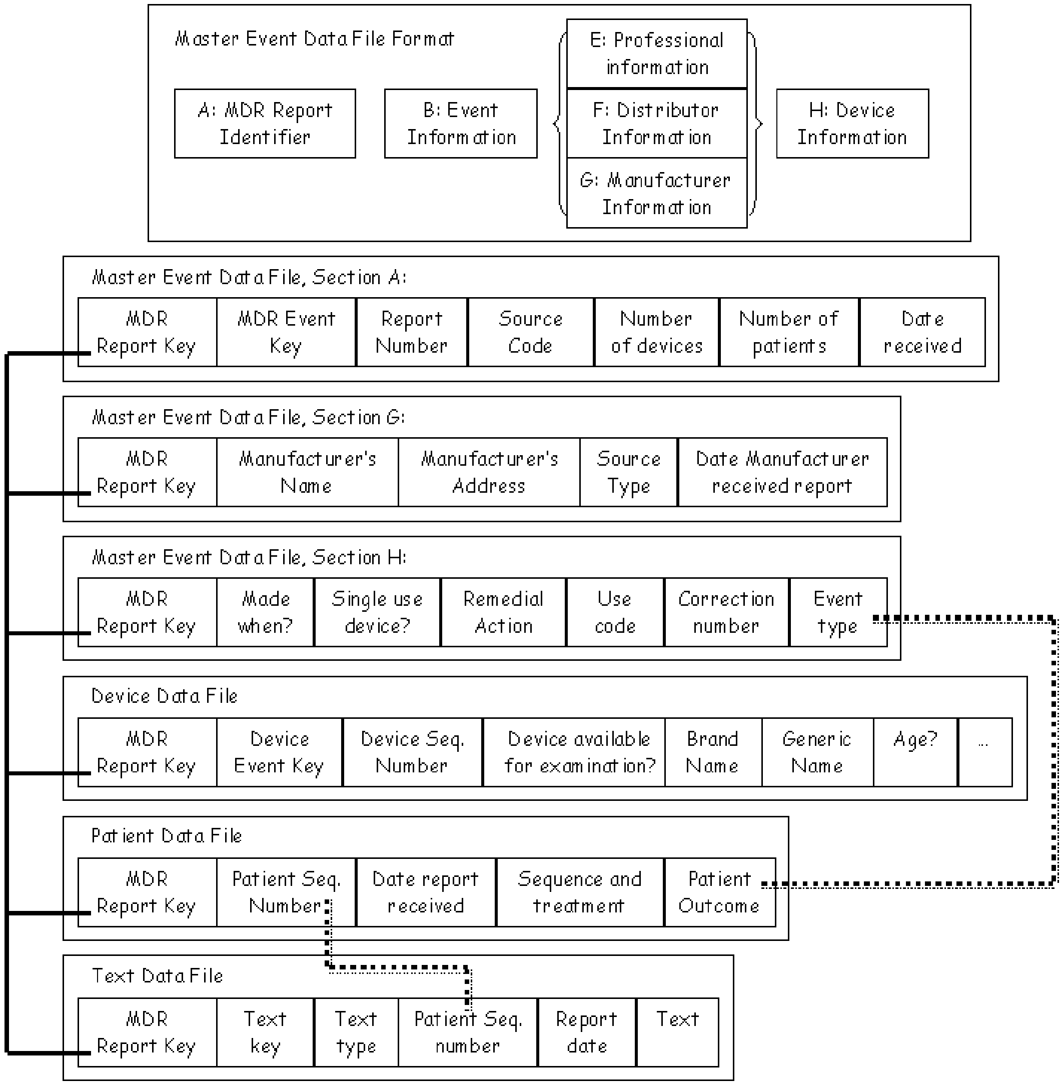


Figure 14.3: Overview of the MAUDE Relations

Unfortunately, the operational demands of many reporting systems have created the need for complex relational schemas. A schema can be thought of as a high-level model of the relationships between the different information fields that are held in the system . They are often structured in terms of objects that would be recognisable to the users of the system. Hence, as we shall see, the MAUDE schema groups information about devices, patients, manufacturers. Figure 14.3 provides a slightly simplified overview of the schema that is used to structure the FDA's MAUDE data. This overview has been reverse-engineered from the technical information that was released to enable interested parties to make use of data that is released under the US Freedom of Information provisions [271]. It is also based on a study of the way in which information is retrieved by the FDA's on-line databases. As can be seen from Figure 14.3, the data is structured around four different relations: a master event relation; a device relation; a patient relation and a text relation. For convenience, these are stored and can be retrieved over the Internet in four different files.

The Master Event Data holds information about the person or group that reports an event. This is the most complex of the relations. It distinguishes between a total of seventy-two different items of information. It also illustrates previous comments about the way in which relatively simple ideas can quickly be compromised by the operation demands of an incident reporting system. The high level structure of this file is illustrated by the top component of Figure 14.3. It is based around the idea of a nested relational schema because any entry in the Master Event Data relation is, itself, composed of more complex relations. These relations hold summary data about the nature of the event, about who has reported the event and about the devices that were involved. Section A of the meta-relation holds identification information relevant to the report. This is compulsory for all reports because, as we shall, see this acts as an index that can be used to cross-reference between the information that is held in other relations within the MAUDE system. Section B hold information about the particular event that is described in the report and again should be included for all reports. Section E is only used if the report was submitted by a healthcare professional. Similarly, Section F only applies to reports filed by device distributors. Section G only applies if the report was completed by a device manufacturer. Finally, Section H is based around a relation that is used to structure information about the devices that were involved in an incident. This should be included for all incidents. As mentioned, the precise information that is held about any particular report depends on the nature of the person or group who submitted the form. For instance, if a form was submitted by a device distributor then the master record will hold information both about the distributor and about the manufacturer that provided them with the device. In this instance, the Master Event relation would contain sections A, B, F and H. If the report is filed by a healthcare professional then they might not be in a position to enter this information into the reporting form and hence it will be omitted from the database. The relation would, therefore, be composed from sections A, B, E and H. Figure 14.3 is simplified by only showing the relations that would record a report from a manufacturer. The FDA provide summary information about the other formats [271].

It is reasonable to ask why anyone should devote this level of attention to the manner in which data is stored within an incident database. It might be argued that such details relate solely to the implementation of particular systems and are of little interest to a more general audience. Such arguments neglect the consequences that such techniques can have upon the end-users of incident databases. For instance, many local incident reporting systems adopt a more 'naive' approach and simply 'flatten out' the nested relations that we have described in the previous paragraph. This would result in every entry in the system being provided with a cell for a Health Care Professional's contact address even though the report was submitted by a distributor or manufacturer. The significance of this should be apparent if we recall that MAUDE receives between 80,000 and 85,000 submissions per year. Relatively minor changes to the relational schema can have a huge impact upon both the time that is taken to search through or download an incident database.

As mentioned, the Master-Event component of the MAUDE database is composed from nested relations. One element of this more complex structure structures the information that is necessary to unambiguously identify each incident. Figure 14.3 illustrates this by the thick line that links the MDR report key across all of the other relations in the MAUDE system. The importance of this 'key' information can be illustrated by the following example. Supposing that we wanted to find out how many patients had been injured by all of the devices produced by a particular manufacturer, we

MDR Report Key	MDR Event Key	Report Number	Source Code	No. of devices	No. of Patients	Date received
Generated	Generated	Generated	Voluntary/ User facility/ Distributor/ Manufacturer	0..Max	0..Max	Date
2339271	319405	2919016-2001-00002	Manufacturer	1	1	06/22/2001
2339103	319248	2124823-2001-00010	Manufacturer	1	1	05/20/2001

Table 14.10: The MAUDE Master-Event Relation (Section A)

could begin by using the Master Event Section G to list all of the MDR Report Keys associated with that Manufacturer's name. After having retrieved the list of MDR Report Keys we could then use Section A of the Master Event File to find out the number of patients that had been reported to be affected in initial reports to the system. It is important to note, however, that the MDR report Key is insufficient to unambiguously identify all information within the system. For instance, an incident might involve more than one device. In this case, Figure 14.3 shows how each entry in the MAUDE Device Data relation must be identified both by the MDR report key and by the Device Event Key. Again, it is important to emphasise that these implementation details have a profound impact upon the users of an incident reporting database. If they are not taken into account during the early stages of development then it can be difficult to extract critical information about previous incidents. In the example outlines above, it might be difficult to extract information about the individual devices that are involved in a single incident. If this data is not clearly distinguished then it can, in turn, become either difficult or impossible to trace the pervious performance of the manufacturers of those devices. Too often, investigators and regulators have sub-contracted the implementation of incident reporting databases with the assumption that such relational structures are both obvious and easy to implement. Equally sub-contractors have often failed to communicate the impact of these technical decisions on those who must operate incident databases. It is, therefore, hardly surprising that so many people express disappointment and frustration with the systems that they are then expected to use.

Table 14.10 provides more details about the report identification information that is held in the Master Event relation. As can be seen, each new entry is automatically assigned three reference keys: the MDR report key, the event key and a report number. The precise meaning and purpose of each of these values is difficult to infer from the FDA documentation that is provided with the MAUDE database. However, it is apparent that the MDR report key and the event key are used to index into other sources of information in the manner described above. The source code information helps to distinguish between the various groups and individuals who might submit a report to this system. Any contribution is either voluntary or is provided to meet the regulatory requirements on end-user facilities, device distributors or manufacturers. The second row of the table summarises the type of information that can be entered in each column. The third and fourth rows of table 14.10 present the values that were entered into the MAUDE database to describe two recent software related failures.

Table 14.11 characterises the nested relation inside the Master Event record that holds information about device manufacturers. Recall that this is only used if a report is submitted by a manufacturer. As can be seen, the name and address of the contributor is stored with the record. Table 14.11 is a slight simplification because the address component of this relation is itself a nested relation containing fields to store manufacturer's street name, their city and state, their telephone number and so on. It might seem like an obvious and trivial requirement to provide such a structure to record the contact information of a contributor. However, the MAUDE structure shows considerable sophistication in the manner in which this data is handled. For instance, two fields

MDR Report Key	Manufacturer's Name	Manufacturer's Address	Source Type	Date Manufact. Received
Generated	Text	Address	Other/ Foreign/ Study/ Literature/ Consumer/ Professional/ User facility/ Company rep./ Distributor/ Unknown/ Invalid data	Date
2339271	A. Maker	Somewhere	Professional, Other	05/30/2001
2339103	Another Maker	Somewhere Else	Professional, User Facility	05/18/2001

Table 14.11: The MAUDE Master-Manufacturer Relation (Section G)

are associated with street address information. Many databases simplify this into a single field and then subsequently have problems encoding information about appartments and offices that have ‘unconventional’ addresses. These issues are not simply important for the technical operation of the incident database. They can also have significant consequences for the running of the system. It is clearly not desirable to have investigators search for a contributor to a confidential or anonymous system in order to conduct follow-up interviews.

Table 14.11 also includes an enumerated type, or list of values, that can be used to describe the source that first alerted the device manufacturer to the incident. This information is critical and is often omitted from incident databases. Chapters 2.3 and Chapters 5.4 have argued that it is important not simply to identify the causes of a near-miss or adverse occurrence. It is also important to identify those barriers that prevented such incidents from developing into more critical failures. Any mechanism that alerts a contributor to a potential failure is an important component in the defences that protect the safety of future applications. Often this information is embedded within free-text descriptions of adverse events and it can prove to be extremely difficult to collate data about such defence. The MAUDE system avoids this problem by firstly prompting the contributor to provide this information by ticking an element in a list of the incident form and then by encoding their response within the ‘source type’ field of the relation illustrated in Table 14.11. The elements of this type are instructive in their diversity. Incidents may be detected from a healthcare consumer or a healthcare professional, they can also be identified by literature reviews or other forms of field study. The final rows of Table 14.11 illustrate sample values for this relation.

Table 14.12 presents the final nested component of the Master Event relation. This is completed for all submissions and provides initial details about the devices that were involved in a near miss or adverse occurrence. Again an examination of the components of this nested relation can be used to illustrate some of the points that have been made in the previous chapters of this book. For example another enumerated type is used to categorise the immediate remedial actions that a manufacturer has taken to address any incident that has been brought to their attention. This is significant because many incident and accident analysis techniques focus directly on causal events rather than examining the critical actions that are taken in the aftermath of an adverse occurrence. In this instance, the manufacturers’ actions are important because they may pre-empt any further regulatory actions by the FDA, for instance, if a device recall has already been issued.

The care that has been taken in devising the FDA’s relational scheme is also illustrates by the use code in Table 14.12. Figure 3.3 in Chapter 2.3 illustrated the higher probability of failure that

MDR Report Key	Made When?	Single use?	Remedical Action	Use Code	Correction No.	Event Type
Generated	Date	Yes/ No	Recall/ Repair/ Replace/ Relabelling/ Other/ Notification/ Inspection/ Monitor Patient / Modification/ Adjustment/ Invalid data	Initial use / Reuse/ Unknown	Previous FDA Notification Reference	Death/ Injury/ Malfunction/ Other
2339271	-	No	-	Initial use	No	Other
2339103	05/18/2001	No	Notification	Initial use	No	Other

Table 14.12: The MAUDE Master-Device Relation (Section H)

is associated during the initial period of operation for hardware systems. This occurs because of component variations but also from the problems associated with setting up devices and of learning to operate them under particular working conditions. The use code in the Master Event Relation, therefore, distinguishes between initial use and reuse of any particular device. This field also illustrates a generic problem with incident reporting databases. Ideally, we would like to provide a meaningful value for every field in every relation. This would enable use to satisfy requests of the following form 'how many incident reports related to the initial use of a device?' or 'how many reports were immediately resolved by the manufacturer issuing a recall?'. Unfortunately, lack of data can prevent investigators from entering all of the requested data into an incident database. For example, if a healthcare professional informs a manufacturer of an incident they may neglect to pass on the information that is necessary to complete the use code. In such a circumstance, the manufacturer would tick the 'unknown' category and return the form to the FDA to be entered into the MAUDE database. This would create problems because if we attempted to answer the question "how many incident reports related to the initial use of a device?' then it would be unclear how to treat these 'unknown' values. If they were excluded then this might result in a significant underestimate of the initial device set-up problems. If they were excluded then the converse problem would occur. Similar concerns can be raised about the 'Remedial Action' field in Table 14.12. In this case, the FDA analysts can enter an 'invalid data' category rather than 'unknown'. This is worrying because manufacturers might, in fact, be exploiting a range of potentially valid remedial actions that are lost to the database simply because they do not fit easily within the categories of remedial action that are encoded within this component of the Master Event relation. These concerns lead to two key heuristics for the application of relational databases to incident reporting:

- *unknown data.* If an 'unknown' value is entered for a field then a caveat must be associated with any statistics that are derived from the data in that field. Ideally, this warning should provide information about the proportion of unknown values compared to those that are known for that field.

- *invalid data.* If 'invalid data' is entered for a field then analysts should also record a reason why this option was selected. System managers should then conduct periodic reviews to ensure that important information is not being omitted through poor form design or an incomplete relational schema.

It is important to emphasise that to entirely exclude either of these categories would place severe constraints on data entry for incident reporting systems. In contrast, these heuristics are intended to ensure that relational schemas continue to offer the flexibility that is necessary when encoding incomplete accounts of adverse occurrences and near misses. They are also intended to ensure that this flexibility does not compromise the integrity of the information that is derived from incident databases [225].

MDR Report Key	Device Event Key	Device Seq. No.	Device Available?	Age	Brand Name	Generic Name	Baseline id	...
Generated	Generated	1.. Max	Yes/ No/ Returned/ No answer	0..Max	Text	Text	Generated	...
2339271	328578	1	No	2	The Item	HNID Panel	K833027	...
2339103	328407	1	No	3	Product	Central Station	K954629	...

Table 14.13: The MAUDE Device Records

Table 14.13 illustrates how MAUDE holds further device information in a separate relation. This separation can be explained by the observation that the Master Event relation holds information that is derived from the incident report. The device relation, in contrast, can hold information that need not be available from the initial report. For instance, Table 14.13 includes a field that is intended to hold information about any baseline report that is associated with a device. Chapter 5.4 has described how baseline reports must be submitted in response to the first reportable incident involving a particular device. It provides basic device identification information including: brand name, device family designation, model number, catalogue number and any other device identification number. This information helps ensure clear, unambiguous device identification. From this it follows that if the incident described in the Master Event relation is not the first occurrence to affect a device than the baseline report summarised in the device relation will be based on previous information. It is again important to emphasise that the relation shown in Table 14.13 simplifies the data that is actually held by the MAUDE system. The device relation ic composed of 43 individual fields. These include information not only about the particular device that was involved in the incident but also about the product range or family that the device belongs to. Such details again emphasise the importance of considering each element of a relational scheme in order to ensure that it captures all of the information that may subsequently help to identify patterns of failure in similar devices.

Table 14.13 illustrates a number of further, generic issues that affect relational schemas in many different incident databases. For instance, the device sequence number helps to distinguish between the different items of equipment that can be involves in any single incident. In order to refer to any particular device record, therefore, it may be necessary to supply both the MDR report key and the sequence number. Alternatively, a device event key can also be supplied to unambiguously identify a device record. Table 14.13 also captures some forensic information, including whether or not a particular device is available for examination. As with the use code in Table 14.12, this field also permits the entry of an unknown value. In this case it is termed 'no answer'. This illustrates a potential problem for the coders who enter the data into the system. They must be trained to distinguish between, or conversely to ignore, the subtle differences in terminology that are used to represent null values in these two different contexts.

A further relation holds information about the patents that were affected by a near miss or adverse occurrence. This is completed even if there were no long term consequences for the individuals who were involved in an incident. Just as there can be several devices that are involved in an incident, there can also be more than one patient. In consequence, any individual patient record must be identified both by the MDR report key and also by the patient sequence number. The sequence numbers that are associated with any incident can be inferred from Section A of the Master Event relation because this records the total number of patients that were affected by an incident. The

MDR Report Key	Patient Sequence No.	Date Received	Sequence Treatment	Patient Outcome
Generated	0..Max	Date	Sequence-Treatment pair	Life threatening/ Hospitalization/ Disability/ Congenital Abnormality/ Requireed Intervention/ Other/ Unknown/ No information / Not applicable / Death / Invalid data
2339271	1	06/22/2001	-	Other
2339103	1	05/20/2001	-	Other

Table 14.14: The MAUDE Patient Records

integrity of the database therefore depends upon a number of assumptions:

1. the Master Event record must accurately record the total number of patients that were affected in an incident.
2. a different Patient Data record must be stored for each patient involved in an incident.
3. each Patient Data record must include a unique Patient Sequence Number and these must follow consecutively from 1 to the total number of patients stored in the Master Event record with the same MDR report key.

If any of these integrity constraints are violated then there is no guarantee that it will be possible for an implementation of the database to return the patient records of all individuals who may have been affected by a near miss or adverse occurrence. For instance, if a patient record was allocated a sequence number greater than the maximum number of patients noted in the Master Event record then doubts would be raised about the reliability of that data. It is also likely that the algorithms for assembling information about an incident might miss the additional patient record if they assumed that the Master Event record was correct.

There are a number of similar constraints that must be observed by those who maintain the MAUDE system. For example, the device sequence number might be related to the total number of devices in the same manner that the patient sequence number is related to the total number of patients. There are further examples. For instance, the Master Event relation, Section H, contains information about the nature of the adverse event. Analysts must enter whether the incident resulted in a death, injury, a malfunction or some other outcome. Similarly, the individual patient records include a 'patient outcome' field that distinguishes between the following categories: life threatening; hospitalization; disability; congenital abnormality; requireed intervention; other; unknown; no information; not applicable, death and invalid data. Clearly the integrity of the database would be compromised if a patient record indicated that a fatality was associated with a particular MDR report key while the Master Event relation showed that the outcome was a malfunction. Fortunately, many database management systems provide explicit support for automating these consistency constraints. They will alert users to potential problems if they arise during data entry. It is, however, less easy for these systems to help users distinguish between the overlapping categories that the FDA have introduced for some fields. These include the 'other', 'unknown', 'no information', 'not applicable' and 'invalid data' options, mentioned above. Later sections will describe the problems that coders have experienced in choosing between these different values when they complete report forms and enter them into incident databases.

Table 14.14 also includes information about the treatment that the patient received following an incident. Any individual patient may receive a number of different treatments. In consequence, the MAUDE relation includes a sequence-treatment pair. This simply associates a number with each of the treatments that was used on that particular individual. It would be possible to construct a further relation that holds more detailed information about each treatment. This could be indexed by the MDR report key, the patient sequence number and the sequence number of the treatment. The data that is released by the FDA from the MAUDE system does not do this. Instead it adopts this compromise approach that resembles a compound attribute [225]. The more general point here, however, is that the database records remedial actions that relate to individual devices, such as product recalls, and the treatments that are taken to counter any adverse consequences for an incident to the patients that are affected. In other words, MAUDE illustrates the broad approach that must be taken when considering what information to capture about the response to any incident.

Table 14.15 provides an overview of the final relation that is used to structure incident data in the FDA's MAUDE system. This relation os central to the success of the system and it represents a solution to a generic problem that affects all incident databases. The previous relations have provided a means of grouping or structuring related information. The patient relation holds information about an individual patient, the device record holds information about an item of equipment that is implicated in an incident and so on. Each element of information that might be placed within a field in one of these relations has an associated type. Most of these types are constrained. For instance, the event type in Section H of the Master Event record can only take the values: death; injury; malfunction or other. This helps to reduce coding problems. Analysts must only differentiate between a few values rather than the subtle differences that might exist between a larger range of potential values. They also help to provide numerical results for statistical analysis. It is relatively easy to sum the total number of incidents which were classified as resulting in a death. This would be far harder if analysts were able to enter any free text value that they liked in the event type field. Analysis might use the terms 'fatal' or 'fatality', 'dead' or 'death' and so on. An automated system would then have to predict all of these potential values and recognise that they were equivalent in calculating any summary statistics. These problems are avoided by have a small range of admissible values that are associated with the various fields in a relation schema. Similarly, the fields in the schema also define a minimum data-set that should be obtained about each incident. The previous paragraphs have described how this minimum data-set can depend upon the nature of the incident report. The information that is available for voluntary reports by a healthcare professional might be very different than that which is available following a mandatory report from a manufacturer. Similarly, we have also described how lack of evidence in the aftermath of an incident can prevent investigators from satisfying the minimum requirement implied by a relational schema. The key point is, however, that by providing 'invalid data' or 'unavailable' options in the database, investigators can be sure that this analysts were prompted for this information when they entered incident data into the system. Any omission, in principle, should be due to the constraints that characterise the aftermath of the incident rather than neglect on the part of the analyst.

Unfortunately, the strengths that the relational model derives from the explicit grouping of related fields of typed information can also be a significant weakness for many incident reporting systems. As we have seen, many near misses and adverse occurrences cannot easily be characterised into the relatively small number of fields that have been introduced in the previous pages. For example, the MAUDE relational schema offers almost no opportunity for analysts to enter the contextual information about workplace factors, such as time pressure or staffing issues, that have been stressed in previous chapters. If any database only recorded the typed information mentioned above then subsequent investigators would derive a very biased view of the causes of previous incidents. In consequence. most relational systems also provide for storage and retrieval of large textual accounts. For instance, Table 14.15 shows how there are two different types of text that can be associated with each MDR report key. Event description summarise any additional information about the immediate course of a near miss or adverse occurrence that cannot be provided in the previous fields. The manufacturer narrative, in contrast, provides an opportunity for the producers of a device to respond to any incident reports. As can be seen in Table 14.15, this response can include information about subsequent studies into the cause of an incident. Such studies must

MDR Report Key	Text Key	Text Type	Patient Seq. No.	Report Date	Text
Generated	Generated	Event description/ Manufacturer narrative	0..Max	Date	Text
2339271	1173556	Event description	1	06/22/2001	User reported a clinical isolate was identified on Microscan HNID panel read by the walk-away instrument system as Neisseria Gonorrhoeae with 99% probability. The specimen was from a blood source from PT. due to the unusual source for this organism the specimen was sent to the State Health Dept Reference Lab for confirmation. The State Reference Lab Identified the organism as Neisseria Meningitis.
2339271	1173558	Manufacturer narrative	1	06/22/2001	H.6 EVAL METHOD: obtained clinical isolate from customer and tested on products involved i.e. HNID panels and Microscan Walk-away Instrument system. Reviewed complaint history, performance evals, labeling and literature regarding reported issue. H.6. RESULTS: Biotype reported by user was duplicated by Microscan Technical Services lab. Atypical results suggest and footnotes indicate N. Gonorrhoeae identification required additional tests to confirm. Results from add'l tests should lead to a presumptive identification on N. Meningitis. Results of complaint history review revealed a very low complaint volume for this issue...

Table 14.15: The MAUDE Text Records

describe the methods used and the results that were obtained. It might, therefore, be argued that a nested relation could be used to distinguish these approaches. This would enable analysts to pose queries about which manufacturers had used a particular evaluation method in response to an incident failure. This is not, however, possible using MAUDE because a general text field is used rather than the more strongly typed approaches that are embodied in previous relations.

The examples in Table 14.15 illustrate the way in which several textual accounts can be associated with a single incident. As mentioned, there is both an event description and a manufacturer narrative for event report 339271. It is also important to realise that more than one individual may be affected by an incident. In such circumstances, an event description can be associated with each person who was, or might have been, injured. The previous table, therefore, includes the patient sequence number associated with each text report. For this it follows that in order to uniquely identify any particular report, analysts will have to supply the MDR report key, the text type and the patient sequence number. In some cases, manufacturers may make more than one response to an incident. If such multiple responses were admitted then analysts would also have to specify the date of the message that they were interested in retrieving. Such requirements appear to introduce unnecessary complexity into an incident database. It is important to remember, however, that there could be profound implications if it appeared that a subsequent response to an incident had in fact been made in the immediate aftermath of an adverse report. Unless the database supports such version control, investigators would have no means of knowing when a narrative was introduced into the system.

This section has provided a relatively detailed analysis of the relational model that is used by the FDA to structure the data contained in their Manufacturer and User Facility Device Experience Database. The level of detail in this analysis is justified by the observation that many reporting databases have failed to provide their expected benefits precisely because those who have commissioned these systems have failed to pay sufficient interest to these details. Conversely, the sub-contractors who are typically enlisted to implement these systems often fail to explain the consequences of particular relational schema both on the queries that can be posed of the system and on the performance that can be obtained as the size of the system grows. Our analysis has also helped to identify a range of benefits that can be derived through an appropriate use of the relational model that is embodied in most incident databases:

- *Analytical help in developing the relational schema.* It can be argued that the process of developing the relational schemas that underly many databases can help to indentify key information requirements. This process is supported by a range of well-documented methods, including entity-relationship modelling and the analysis of normal forms [225]. In particular, these approaches can help to expose the integrity constraints that must be satisfied by any implementation. Although these techniques have their limitations and none are specifically intended to support the development of incident databases, they do have the strong advantage that they are 'industry standard' and hence widely understood. This introduces an important paradox because the reverse enginering of many incident databases has revealed important structural weaknesses, which suggest that many of these systems have been built without the benefit of these relatively simple engineering techniques [417].

- *Analytical help in guiding incident classification.* The development of a relational schema is intended to enable investigators, regulators and safety managers to classify incident reports so that they can be analysed and retrieved at a later date. The process of constructing a relational schema, therefore, forces people to consider the forms of analysis that any system must support. This leads to an important decision. Either the person submitting a form must indicate appropriate values from an incident classification or the analysts must codify a less structured account into the fields that are included in a relational schema or a hybrid model can be adopted where the contributor performs a 'first pass' classification that is then refined by the investigator. No matter which approach is adopted, the key point is that the development of the incident database must have an impact on the manner in which data is both elicited and codified. It is, therefore, extremely difficult to simply bolt-on an existing database to an incident reporting system where either contributors or analysts will have to adjust their behaviour to support the values that are built into a relational schema. In such circumstances, rather than guiding analysts and contributors towards an appropriate classification they can find that a database forces them to 'squeeze' or 'massage' an incident into inappropriate data structures.

- *Efficiency.* One of the key technical benefits behind the relational approach is that it helps to avoid the duplication of redundant information. Ideally, we might store information about a device manufacturer once. Similarly, an optimised system would only ever store a single record about any particular device. This would record the complete service and version history of that item. A link could then be made from an incident report to a device record and from there to the associated manufacturer. An alternative model would be to duplicate manufacturer information each time a new incident record was created. This is not only wasteful in terms of the storage that is required, it can also significantly increase the amount of time that is required to collate incident information. The technical reasons for this relate to the search latencies that are associated with primary and secondary storage. Relational techniques can use indexing so that once a common item of information is stored in main memory then those details do not then need to be repeatedly fetched from slower secondary media [225].

This is a partial summary, however, it is also important to stress that our analysis has identified a number of problems with the use of relational databases for incident reporting. For instance, many of these applications rely upon strong typing to clearly distinguish between the admissible values that can be entered into each field. This creates problems because in the early stages of an investigation it is often impossible to be certain about which values might hold. A good example of this might be the problems associated with any assessment of the consequences of an incident based on a clinical prognosis. This uncertainty results in a proliferation of 'unknown' values that make it very difficult to interpret the accuracy of statistics that are derived from incident databases. Similarly, it can be very difficult for analysts to accurately and consistently distinguish between the numerous values that might be entered into particular fields within a relation. This can result in similar incidents being classified in a number of different ways within the same relational schema. It can also lead to 'not applicable' values being used as a default. The previous discussion has also identified the potential vulnerabilities that can arise from the relationships that often exist between the components of a schema. In particular, problems can arise from the way in which MAUDE links the maximum number of patients and devices in the Master Event record to provide a range for patient and device sequence numbers. Automated support must be provided to ensure that consistency requirements between these linked values are maintained throughout the lifetime of the database. This is a partial summary. the following sections expand on the problems that can affect the use of the relational model for incident reporting databases. Subsequent sections then go on to review further computational techniques that can be used either to replace or augment this approach.

Problems of Query Formation

Previous sections have described how the relational model can be used to reduce the storage requirements and increase the speed of queries that are performed on incident databases. It provides further advantages. For instance, search requests can be formulated a using relational algebra. The operators within these languages have a close relationship to the operators of set theory, such as union, intersection and set difference. This offers a number of benefits. Firstly, the components of the relational algebra should have a clear semantics or meaning. Users can apply the basic ideas in set theory to gain some understanding of the query languages that are supported by most relational databases. There are further benefits. As most implementations exploit set theoretic ideas, it is therefore possible to apply knowledge gained from one relational database system to help understand another. The mathematical underpinning of the approach support skill transfer and a certain degree of vendor independence. Unfortunately, as we shall see, relatively few investigators or safety managers have acquired the requisite understanding of set theory or of relational algebra to exploit these potential benefits. The following sections provide a brief overview of the relational operators applied to an incident database. The intention is both to illustrate the potential application of this approach and also to illustrate some of the complexity that can arise from the relational algebra.

Before discussing the set operators, mentioned above, it is necessary to introduce two additional elements of the relational algebra: SELECT and PROJECT. The SELECT operator is usually represented in the relational algebra by σ and is applied in the following manner:

$$\sigma < selection\_condition > (Relation) \tag{14.1}$$

The selection condition is a Boolean expression that is usually formed from attribute names, operators and constants. The operators include $=, >, \leq, <, \geq$. Attribute names denote particular fields in a relation. For instance, Table 14.12 includes the attributes MDR report key, made when? single use? remedial action, use code, correction number and event type. The constant values include elements of the classification scheme that might be entered into these fields. For example, the Use Code constants include 'Initial use', 'Reuse' and 'Unknown'. We can put all of this together in the following manner:

$$\sigma < Use\_Code = Initial\_Use > (Table\ 14.12) \tag{14.2}$$

This expression would yield all of the entries in Table 14.12 which were associated with the initial use of a device. One of the benefits of this approach is that we can combine elements of the algebra to form more complex expressions. For instance, we might want to SELECT all entries that relate to either the initial use or reuse of medical devices:

$$\sigma < Use\_Code = Initial\_Use > OR < Use\_Code = Reuse > (Table\ 14.12) \tag{14.3}$$

The SELECT operation can be thought of as selecting a row from one of the relations in a database schema [225]. The PROJECT operator, in contrast, can be thought of as selecting particular columns within a relation. The application of this operation can be illustrated as follows. The first sentence denotes the general form of the PROJECT operator. The second sentence shows how it can be applied to the relation in Table 14.12. This would yield a list of all MDR keys together with the date when the corresponding device was manufactured and the type of event it was involved in:

$$\pi < attribute\_list > (Relation) \tag{14.4}$$

$$\pi < MDR\_report\_key, made\_when?, event\_type > (Table\ 14.12) \tag{14.5}$$

There are a number of important features of the SELECT and PROJECT operators. For example, if the projection is applied to non-key fields then it is likely that it will yield duplicate values. For example, if we omitted the MDR Key field in the previous example, we might derive a number of reports in which the devices were made on the same day and produced the same outcome. The PROJECT operation filters these duplicate values because the result must itself be an operation. Brevity prevents any sustained analysis of these more detailed features, the interested in reader is directed to [225]. In contrast, the following paragraphs focus on the main features of the relational algebra. The intention is to illustrate both the power of the approach but also the usability problems that can prevent many investigators from exploiting this language as a means of interacting within incident databases. It is often necessary to combine the operations in the relational algebra to form more complex requests. For example, we might wish to create a list of the dates and MDR keys for all incidents that occurred during the initial use of a device. This can be done in the following manner. Both of the following forms are equivalent, however, the operations (14.7) and (14.8) make use of the RENAME (←) operator to hold the intermediate result of the SELECT operation. This can provide important benefits as users form more complex queries:

$$\pi < MDR\_report\_key, made\_when? > (\sigma < Use\_Code = Initial\_Use > (Table\ 14.12)) \tag{14.6}$$

$$TEMP\_RELATION \leftarrow (\sigma < Use\_Code = Initial\_Use > (Table\ 14.12) \tag{14.7}$$

$$\pi < MDR\_report\_key, made\_when? > (TEMP\_RELATION) \tag{14.8}$$

As mentioned, an important strength of the relational algebra is that it builds upon the relatively well-known concepts of set theory. For example, the UNION operator can be used to produce a relation that is composed from the set of tuples that are in one or other or both component relations. This can be illustrated by the following example. Suppose that an investigators wanted to derive the MDR event keys for all incidents involving devices that were manufactured on or before 1996 or that were in use for more than six years. This can be done in three stages. Firstly, (14.9) and (14.10) identify the devices that were manufactured during or before 1995. Then (14.11) and (14.12) extract those devices that have been in operation for six or more years. Finally, (14.13) forms the union of the two previous stages of the operation:

$$Temp1 \leftarrow \sigma < made\_when? \leq 1995 > (Table\ 14.12) \tag{14.9}$$

$$Old\_devices \leftarrow \pi < MDR\_Report\_Key > Temp1 \tag{14.10}$$

$$Temp2 \leftarrow \sigma < age? \geq 6 > (Table\ 14.13) \tag{14.11}$$

$$Old\_models \leftarrow \pi < MDR\_Report\_Key > Temp2 \tag{14.12}$$

$$All\_Old\_Devices \leftarrow Old\_models \bigcup Old\_devices \tag{14.13}$$

$$\tag{14.14}$$

INTERSECTION and SET DIFFERENCE can be used in a similar fashion. For example, we could identify the MDR Report Keys of devices that were manufactured during or before 1996 but which have not been in operation for 6 or more years using SET DIFFERENCE in the following manner:

$$Legacy\_Devices \leftarrow Old\_models - Old\_devices \tag{14.15}$$

Conversely, we can identify those models that were manufactured during or before 1996 and which have been in operation for six years or more using INTERSECTION. This illustrates the flexibility of the relational algebra. It is possible to use the set theoretic operators to express a range of relatively complex constraints that can be applied to relatively simple relational schemas:

$$Obsolete\_Devices \leftarrow Old\_models \bigcap Old\_devices \tag{14.16}$$

The JOIN ($\bowtie$) operation is used to combine related tuples from two relations to form a single relation. If the first relation has N attributes and the second relation has M attributes then the resulting relation will have N+M attributes. The application of this operation can be illustrated in the following manner. The first sentence illustrates the general form of the JOIN relation. The second sentence illustrates how it can be used together with the *All_Old_Devices* that was derived from (14.13) to extract all of the patient related information for incidents that involved devices, which were either manufactured during or before 1996 or that have been in use for more than six years:

$$Relation\_1 \bowtie < join\_condition > Relation\_2 \tag{14.17}$$

$$Patients\_affected \leftarrow All\_Old\_Devices \bowtie < MDR\_Report\_Key > (Table\ 14.14) \tag{14.18}$$

As before, a number of additional details must be considered when using this operator. For instance, the results of a JOIN operation do not typically include any tuples with NULL values in the parameters of a *join_condition*. This property of the relational algebra can have important consequences for incident reporting databases. The proliferation of null values can mask a large number of candidate incidents that might have been included within the results of a particular query if more information had been obtained about the adverse occurrences that they document. Hence, the results of queries that use the JOIN operator may significantly under-report all possible candidate incident records. They may also mask the number of reports that are omitted by simply ‘dropping’ all candidate tuples with NULL values. These problems cannot arise in the previous example because we have joined the relations on the primary key that is used throughout the MAUDE system, in other words the MDR Report Key cannot contain a null value. Problems would, however, arise if we attempted to perform a JOIN using the Patient Outcome or Source Type fields. As we shall see, MAUDE avoids this problem by constraining the queries that can be performed using these relations. In general, this problem can be avoided by using a form of the OUTER JOIN operation. Unfortunately, such distinctions are often not apparent to those who must learn to use incident databases [472].

A number of pragmatic observations can be made about the relational algebra and its use within incident reporting systems. For those with a mathematical background, it offers a clear semantics and its origins in set theory can reduce training times. For those without such a training, it can appear to be both complex and confusing. There are also other aspects of the language that often irritate both sets of users. A particular feature is that the user must specify a precise ordering for each operation. If they are performed in any other sequence then the result may not be what the user had anticipated. The syntax associated with the relational algebra has also been criticised as opaque and difficult to learn [225]. It is for this reason that the Structured Query Language (SQL) has emerged as a standard means of interacting with many database systems. There are strong differences between the elements of this language and the relational algebra. Duplicate values are allowed within in SQL tables, they are not permitted within relations. Hence the mathematical underpinnings of SQL are based more on bags that sets of tuples. There are further differences. For instance, the SQL SELECT statement has no formal relationship to its counterpart that was introduced in previous paragraphs:

```
SELECT <attribute_list>
FROM <table_list>
WHERE <condition>
```

In this general form, `<attribute_list>` is a list of the attribute names or fields whose values are to be retrieved by the query. A `<table_list>` is a list of the relation names that will be processed by a query. As might be expected, `<condition>` is a boolean expression that identifies those relation components that are to be retrieved. The application of this form can be illustrated by the following query, which retrieved the MDR Report Keys and the date of manufacture for devices that were either being used for the first time or were being re-used when an incident occurred:

```
SELECT <MDR_Report_Key, made_when?>
FROM Table.14.12
WHERE <Use_Code= Initial_Use> OR <Use_Code= Reuse>
```

The SQL query is intended to be declarative. Users should not have to worry about the precise ordering of the individual terms in an expression. It can be contrasted with the corresponding formulae in the relational algebra in which the select must be performed before the projection because the projection would strip out the information about the use code that is used as the basis for the selection operation:

$$\pi < \mathit{MDR\_report\_key}, \mathit{made\_when?} > (\sigma < \mathit{Use\_Code} = \mathit{Initial\_Use} > (\mathit{Table}\ 14.12)) \quad (14.19)$$

SQL offers further benefits. In particular, it is possible to construct complex queries that select elements from several different relations or tables. The following example extracts the MDR report key, any remedial actions and the patient outcome for any incidents involving the CIU panel brand of devices. Notice that the patient outcome is derived from the data that is held in the MAUDE patient file, illustrated by Table 14.14, while the remedial action is associated with a particular device report, illustrated by Table 14.12:

```
SELECT <Table.14.12.MDR_Report_Key, Table.14.12.Remedial_action,
        Table.14.14.Patient_outcome>
FROM Table 14.12,Table.14.14
WHERE <(Table.14.12.MDR_Report_Key = Table.14.14.MDR_Report_Key) AND
        (Table.14.12.Brand_name = 'CIU Panel')>
```

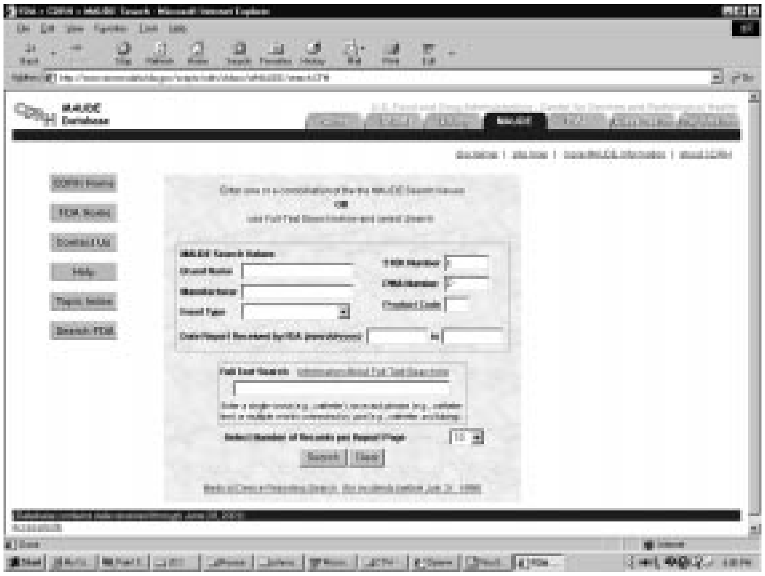
In addition to these mechanisms for complex query formation, SQL also offers a limited range of statistical operations that can be used to support the monitoring functions, which will be discussed in Chapter 14.5. These functions include `SUM`, `MAX`, `MIN` and `AVG`. These can be applied to a set or bag of numeric attributes. The following example illustrates a query to find out the total number of patients that have been affected by adverse incidents, the maximum number affected by a single incident and the average number of patients affected:

```
SELECT <SUM(Number_of_patients), MAX(Number_of_patients), AVG(Number_of_patients)>
FROM Table 14.10
```

The rapid development of SQL as a standard for interaction with relational databases provides a strong indication of its advantages over the 'raw' relational algebra that has been illustrated in previous pages. However, these benefits do not provide a panacea for the implementation of incident databases. Many of the advantages that SQL offers can only be appreciated by programmers and developers who have the necessary background to exploit many of its more advanced features. The majority of safety managers, of regulators and of incident investigators have little appreciation of how to use SQL. This creates considerable practical problems. For instance, the '*query paradox*' arises because those people who can best exploit data about previous failures lack the technical expertise to form the queries that reveal hidden patterns within the data. In contrast, the individuals who have the technical expertise to form appropriate queries often lack the understanding of the application domain that is necessary to identify what questions to ask the incident databases. This paradox does not simply affect incident databases. Its consequences are, however, potentially more serious given the nature of the data that is held in this systems. There are a number of potential solutions. For instance, database experts and information technologists might be trained to have a greater appreciation of the application domain. Alternatively, investigators, regulators and safety managers might be trained to have a greater appreciation of the technical underpinnings of the systems that they use to support the everyday tasks. Unfortunately, neither of these alternatives seems to have been followed in any systematic manner. It is more common to find a lack of understanding between those who maintain reporting databases and those who must use them [423]. This is often revealed in complaints that the system will not provide access to data that the users believe it 'must hold'. Conversely, administrators are often faced with demands to support facilities that cannot easily be provided using relational databases, such as free-text retrieval.

The tensions that are created by the query paradox focus on the design of the user interface to incident databases. These, typically, attempt to simplify the task of interacting with large collections of data by supporting a limited number of pre-formulated queries. All that the use has to do is specify values for the particular fields that they are interested in. This can create potential conflicts because these pre-canned queries are unlikely to satisfy the diverse requirements of many potential users. Previous paragraphs have emphasised the difficulty of predicting all of the possible queries that investigators and regulators might want to pose to such a system. Phrases such as 'data mining' and 'exploratory analysis' are often used to publicise these systems but these activities are hardly supported by the limited numbers of 'pre-canned' queries that are supported by many incident databases. In consequence, support staff are often faced with continual requests to perform one-off analyses that cannot easily be constructed from the existing interface.

The top image of Figure 14.4 illustrates the screen that provides web-based access to the MAUDE system. The bottom of the two images shows the list of incidents that can be derived from a particular query. If the user selects any one these 'hits' they can view a relatively complete summary of the various fields that have been described in previous paragraphs and which are illustrated in Tables 14.10, 14.11, 14.12, 14.13, 14.14 and 14.15. As can be seen from the top screen, users can either enter specific values for a relatively small number of fields or they can perform a free text search. Subsequent sections will describe the strengths and weaknesses of such 'free-text' searches in greater detail. For now it is sufficient to observe that the standard query interface only provides a very small subset of the potential queries that users might pose of the FDA's data. Recall that the main data file holds a relation with 72 fields. The device file relation provides a further 43 fields. It would be difficult to design a graphical user interface that would enable a untrained user to form the wide range of SQL queries that might be performed on such data sources. The relative simplicity



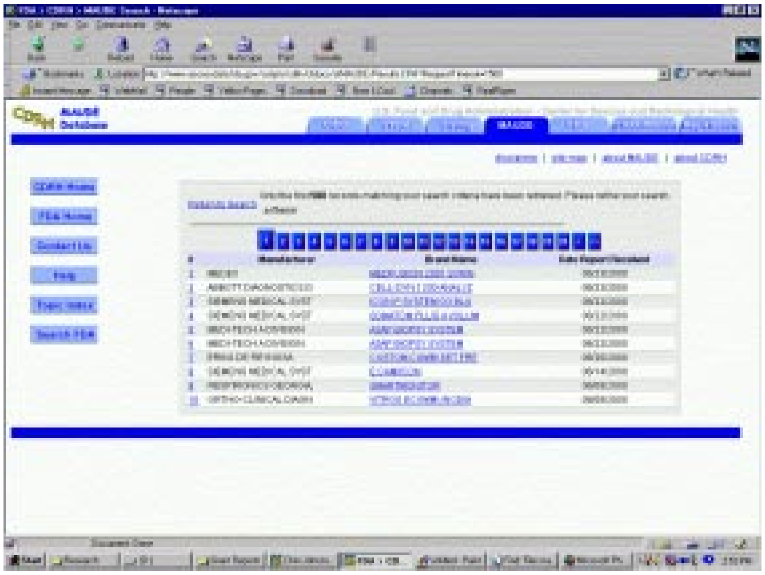


Figure 14.4: The MAUDE User Interface

of the top screen in Figure 14.4 can, therefore, be seen to have strong design strengths. It does not daunt initial users with a vast array of bewildering options. Equally, however, it does not support the more sustained analysis of trends and patterns that might be performed by users who have a more extensive knowledge of the technical under-pinnings of relational databases.

It is important to emphasise the complex nature of the paradox that was introduced in previous paragraphs. Even if investigators are trained to appreciate the concepts and mechanisms of relational databases there are many potential further pitfalls. Many professional software engineers fail to construct 'correct' queries using relational query languages such as SQL [703]. In other words, they rely on queries that will not return the information that they are believed to. There are technical issues, such as the different semantics associated with the JOIN operator, that complicate the application of these techniques. As we have seen, this is a particular problem for incident databases that are likely to contain many NULL values in the aftermath of a near miss or adverse occurrence. These technical issues are, arguably, less significant than the problems of ensuring that incident data is correctly entered into the system in the first place.

Problems of Classification

The previous paragraphs have described how relational databases are constructed around a number of fields. Each of these fields captures particular values. For instance, the event type in Section H of the MAUDE Master Event record can be: death; injury; malfunction or 'other'. The use of such schemas offers numerous benefits. For example, it provides a national framework for the collection and analysis of incident data. Any organisation that contributes to the system must provide their data in a format that can easily be integrated into the relational schema. It is difficult to under-emphasise the practical importance of this. Other industries, have experienced considerable difficulties in exploiting incident data precisely because they lack an agreed format that can be used to structure incident information. For example, Boeing currently receives data about maintenance incidents from many customer organisations. Each of these organisations exploits a different model for the records in their relational systems. As a result, the aircraft manufacturer must attempt to unify these ad hoc models into a coherent database. The GAIN initiative has taken considerable steps to address this problem [310]. At present, however, it can be difficult to distinguish between bolts that have failed through design flaws and bolts that have failed because of over-torquing by maintenance engineers. Sam Lainoff summarised the problems of populating relational databases:

> "There is no uniform reporting language amongst the airlines, so it's not unusual to find ten different ways of referring to the same thing. This often makes the searching and sorting task a difficult proposition? The data we have won't usually permit us to create more refinement in our error typing. But at times it will give us enough clues to separate quality problems, and real human error from pure hardware faults." [472].

The previous quotation blurs the notion of a taxonomy or language for incident reporting and the relational technology that is used to retrieve incident reports. This confusion is understandable. Relational databases implement data models. The popularity of this technology has created a situation in which it is difficult to envisage the development of a taxonomy or data model without some corresponding computer-based implementation. On the other hand, there have been a number of national and international initiatives to develop taxonomies for incidents and accidents without considering the need for tool support. The US National Patient Safety Foundation (NPSF) clarifies some of these important distinctions in its analysis of the Aviation Safety Reporting System (ASRS) . The final sentence is particularly important because it identifies the opposite problem to that faced by Boeing. The previous quotation illustrates the difficulty of synthesising incident data that does not share a common model or taxonomy. The following quote recognises the danger that events will be analysed until they fit the taxonomy even if that taxonomy does not accurately represent the incident under consideration:

> "In the discussion about incident reporting, it was pointed out that the ASRS uses an extensive indexing system, but this is used to collect related subsets of narrative cases from the database that pertain to a theme or question. The indexing system does not

> work automatically but is a tool used by the staff to carry out analyses and to assist outside parties use the database in their analyses. The indexing is used as a tool in analysis; the classification system it represents is not the analysis."

The importance of an appropriate data model cannot be under-estimated. For example, these taxonomies drive the statistical analyses that are often cited as a primary benefit of incident reporting. We have already described how the event type in Section H of the MAUDE Master Event records whether an incident involved a death, injury, malfunction or 'other'. Summing the number of records in each category can provide valuable information about the types of occurrences and near-miss incidents that are reported to the system. Equally, if we were concerned to identify the number of incidents that resulted in a particular type of injury then we must look to other fields in the MAUDE system. If none of them satisfied our information requirement then we might not be able to report on those types of incident.

The incident models that are embedded within relational databases have further benefits. They help to guide the local analysis and classification of incident reports. This is a significant benefit for large-scale systems. Central organisations may lack the necessary local insight to drive the classification of a particular incident. They may also lack the resources that are required to centralise the analysis of every potential report. By devolving the classification process to regional or local representatives, central investigators can focus on responding to higher-criticality incidents or to those exceptional incidents that do not fit into the existing taxonomy. A key point here is that the values in the data model provide powerful guidance to those individuals who are documenting an adverse occurrence. They provide a prompt for the type of information that must be provided. They also indicate the particular values that each item of information might take. It is, therefore, often argued that the data models that are embeded within relational databases help to improve inter-rater reliability during the analysis of incident reports.

It can be extremely difficult to construct a taxonomy that is capable of capturing all of the information that people might want to extract about adverse occurrences and near-miss incidents. For instance, one approach is to rely upon a small number of high-level categories for most of the data fields. Information would be gathered about 'software failures' rather than 'floating point exceptions'. Similarly, incidents might be characterised by 'human error' rather than 'poor situation awareness'. This approach has the advantage that many high-level categories will be resilient to change. The particular forms of software failure that may be reported to a system can be affected by changes in the underlying technology. Similarly, the findings of human factors research are likely to have a more profound impact on detailed distinctions than they are upon broader categories. There are further benefits. For instance, by restricting the number of distinctions that must be made between different types of incident data it is possible both to increase inter-rater reliability and reduce potential training times for local and regional analysts.

Unfortunately, high-level taxonomies suffer from a number of limitations. The elements of these models often fail to capture the particular details that characterise many incidents. They may, therefore, omit information that might contribute to the safety of future systems. In particular, high-level taxonomies tend to support retrieval systems that yield very low precision for many of the queries that users want to pose. For instance, a request for information about 'floating point exceptions' will fail if all relevant reports are classified as 'software failures'. One means of avoiding this problem is to include free-text descriptions that provide additional details about the particular characteristics of each incident. As we shall see, these can be searched using specialist information retrieval techniques. These systems must recognise similar classes of failures in spite of the different synonyms, euphemisms and colloquialisms that are provided in free-text accounts of 'bugs', 'crashes', 'exceptions' and 'run-time failures'. In general, however, users may be forced to manually comb through each recorded software failure to extract those that relate to floating point exceptions.

The US National Co-ordinating Council for Medication Error Reporting and Prevention's Taxonomy of Medication Errors provides an example of a more fine-grained approach to incident classification [582]. This contains approximately 400 different terms that record various aspects of adverse incidents. If this were embodied within a storage and retrieval system then it would enable analysts to pose a number of extremely detailed questions about both the causes and outcomes of adverse medication incidents. Such a sophisticated approach also implies a high-level of training for those who must complete any analysis. Systems that are based on a detailed taxonomy increase the potential for confusion and ultimately low recall because different classifiers may exhibit subtle differences in the ways in which they distinguish between the terms in the taxonomy. In consequence, a number of these systems exploit flow-charting and similar techniques to help analysts identify which fields relate to a particular incident. Figure 11.9 provided an example of this approach by illustrating the the Eindhoven Classification Model. Analysts must first decide whether the causes of an incident are primarily technical, organisational or 'human'. Each of these high-level categories is successively broken down into increasingly more detailed causal factors until the terminal nodes represent the ultimate classification that will be applied to an incident. A recent study of trained staff classifying incidents according to the MEDWATCH codes, described in previous sections, identified a vast array of potential pitfalls. These resulted in a recommendation that either the FDA consider funding the centralised coding of all event reports or that the coding scheme be redesigned to benefit from top-down decomposition techniques similar to those exploited by the Eindhoven approach. MEDWATCH codes could be merged with the coding systems and hierarchical structures available within the National Library of Medicines Unified Medical Language System with an additional hierarchical coding system for device problem coding [264]. Unfortunately, further problems complicate the use of these hierarchical coding schemes. As we have seen, many incidents stem from complex combinations of many different causal factors. It can be difficult to ensure that independent analysts will arrive at the same classification patterns even when they have access to such tools.

The problems of inconsistency in detailed classification schemes can be seen as a slightly esoteric concern. Many large-scale systems face the more prosaic problems of ensuring that staff provide all of the information that is required about an incident. These problems can be exacerbated when staff must search through lists of valid codes to ensure a correct classification for each data field. The US Food and Drug Administration expressed their concern about this issue in their User Facility Reporting Bulletin. This provides feedback to the individuals and organisations who contribute information about device related failures. In an article entitled 'THOSE CODES' they describe how:

> "The final Medical Device Reporting regulation became effective July 31, 1996. Since then, Food and Drug Administration (FDA) staff have observed numerous errors and omissions in the MDR reports submitted by user facilities to report device-related deaths and serious injuries. These errors cause major gaps in FDAs adverse event reporting database, and may also delay manufacturers failure analyses while the manufacturers contact user facilities for additional information. FDA plans to send letters to those user facilities that have submitted incomplete mandatory forms (3500A) to request they file supplemental reports". [277]

One obvious means of addressing this problem is to ensure that analysts receive explicit training in the application of a classification or coding scheme. The APSF operates what is arguably the most elaborate of these systems [36]. Their training scheme is designed to provide experience of each of the 95 different options that can be coded into the AIMS system. The existing database is used to ensure that trainees meet a representative cross-section of scenarios as they learn how to use the classification scheme. They also work under the supervision of an APSF 'accredited coder' even though most of the course is conducted remotely using email and telephone contact. The training consists of an initial orientation session that covers the general motivation behind incident reporting and monitoring. They are taught how to install the associated AIMS+ software. They are then guided through an initial data entry and classification exercise. There then follow three different levels of training. In the first level, the trainee must code a sample collection of fifty incident reports. During this process they can request as much help as they consider to be necessary from the accredited tutor. The results of this classification exercise are then reviewed and graded by their supervisor who will provide appropriate feedback to the trainee

Second level training involves the coding of another fifty incident reports. In contrast to the previous exercises, however, the trainee is encouraged not to seek help from their supervisor unless absolutely necessary. These are again reviewed and graded before feedback is provided. In order to progress to the final level of accreditation, they must achieve a 60% 'pass rate'. This requirement is highly significant. It reflects the recognition that it will not be possible to expect or achieve 100% agreement between different analysts immediately after a period of relatively intensive training in a particular classification scheme. If the trainee fails to satisfy this 60% requirement then they must repeat the second level training with a further bach of fifty incident reports. Trainees are permitted three attempts at this second level before the tutor is required to refer the candidate back to their sponsoring organisation. The third level of training involves a final set of fifty sample incidents. Individual discrepancies in the coding are reviewed by the supervisor and the trainee. The expected pass rate is now raised to 75% before the trainee can graduate from the course [37].

Such training undoubtedly provides important support for the codification of incident data. Unfortunately, it can be both time consuming and expensive. These factors act as powerful disincentives for many organisations. There are also doubts about the long-term effectiveness of such training. Even in Sentinel systems, where additional resources are targeted on a few 'case study' organisations, it can be difficult to demonstrate the success of such initiatives. The FDA identified problems including "lack of coding (estimated at 50% of incoming reports), incorrect coding, and use of codes that are too general to be useful (e.g., device malfunction)" [264]. The success rate for organisations who were outside of this select group can be expected to be correspondingly lower.

Chapter 10.4 identified problems that affect the use of coding schemes to inform the causal analysis of adverse occurrences and near-miss incidents. Causal factors change over time as new systems and working practices are introduced. For instance, the introduction of microprocessor controlled infusion devices has created the potential for incidents that could not have happened in the past [183]. Similarly, classification schemes may also change as new ideas are developed about the underlying problems that lead to human 'error' and system 'failure'. For instance, our understanding of the impact of workload on human decision making has changed radically over the last decade [426]. The application and development of a reporting system can also help to identify improvements to existing classification systems. Many coding systems provide analysts with the opportunity to state whether or not they believe that any necessary information has been omitted from their classification. The feedback received from these submissions can be used to distinguish data that cannot be included within a classification scheme from information that was simply overlooked during the coding process. For example, the APSF training scheme, mentioned above, was structured around a Generic Occurrence Classification system (GOC). In 2000, this was was updated to GOC+. The introduction of GOC+ was supported by a computer-based classification system. The interface to this tool leads the analyst through a process that is intended to collect all of the relevant information that is required for each type of event. The changes were intended to increase the scope and content of the system:

> "Since incident monitoring began, the APSF has learned a great deal more about the factors involved in incidents in healthcare. In order for the GOC to remain a relevant classification tool, this additional knowledge has been incorporated into the classification. Another priority in the development process was to improve coding consistency, accuracy and timeliness. By analysing the issues that influence consistency, accuracy and timeliness, the development team was able to focus the development on managing these issues." [34]

Changes to any classification scheme can create considerable problems for the maintenance of an incident reporting system, In particular, it can be difficult to ensure that all data is indexed in a consistent manner. For instance, they may already be incidents in the database that provide examples of new classification concepts. In such circumstances, analysts may be forced to manually reclassify thousands or hundreds of thousands of existing records. This is often impossible. In consequence, many incident collections become partitioned by the coding schemes that were used to compile them. Separate queries may have to be performed for records that were gathered before and after an update. Statistics of the form X% of all incidents were caused by Y will have to be parameterised by the duration of the data-set that supports this analysis, even though there may be data that could confirm this analysis over a longer time period. A number of reporting systems have, therefore, attempted to develop computer-based tools that will guide analysts through the task of converting between coding formats. None of these systems can, however, entirely remove the need for manual intervention when new data is required by revisions to an existing classification system.

14.4.2 Lexical Information Retrieval

The previous paragraphs have summarised the problems of data maintenance that can arise when reporting systems rely upon relational systems. Further problems restrict the utility of these systems for end-users. In particular, it can be difficult to overcome the problems associated with query formation, both in terms of the knowing what to ask and how to ask it. These potential limitations can be addressed through the development of user interfaces that hide the underlying relational model:

> "The semantic query system in AIMS 2 release 2 will enable users to drill down into the data without having to understand the underlying database structure. We will also include some basic data mining facilities that allow users to contrast and compare rates across locations, incidents, etc." [39]

These approaches create additional problems. By hiding the underlying model in a relational system, it can be difficult or impossible for users to learn how to form their own queries. In consequence, they are restricted to those questions that have been anticipated and are supported by the systems developers.

Information retrieval tools provide powerful alternative mechanisms for searching large collections of unstructured data. Brevity prevents a complete exposition of the many different techniques that have been exploited by these systems, for a more complete review see Belew [71]. In contrast, the remainder of this section focuses on lexical information retrieval techniques. This decision is justified by the observation that these techniques have had the most widespread impact on commercial retrieval systems. The following section examines case based reasoning approaches that provide a point of comparison with these more widespread tools. Lexical information retrieval systems, typically, rely upon a three stage process. Firstly, collections of documents are indexed. This process associates one or more keywords with a document. By automating this process it is possible to reclassify large collections as new incident reports are received by the system [792]. This is a significant issue, as we have seen the Food and Drugs Administration's MAUDE system receives between 80,000 and 85,000 reports per year. Automatic classification not only offers the possibility of reducing inter-rater reliability concerns but also can reduce the costs associated with manually classifying each incident within a relational database.

The second stage of the information retrieval process involves processing the user's query or information request. The intention is to identify terms that might be matched against the keywords that were identified for each document. This create problems. There can be a mis-match between the terms that a user exploits when looking for an incident and the keywords used during the indexing phase. There is also a danger that a retrieval system will under-value those terms that the user perceives to be the most significant in their query. In such circumstances a request for 'software failures in surgical procedures' might focus on surgical incidents rather than the more detailed criteria for 'software failures'.

The final stage searches through a collection to identify matches between the terms in the users' query and the keywords that index each report. Unfortunately, users tend to form general queries that match many potential documents even when they have a relatively precise information need. In consequence, it is likely that an initial request may have to be iteratively refined as users search through large scale incident databases. This search process can be supported by relevance feedback techniques. The user indicates which of the proposed documents were actually relevant to their query. This information is then used by the retrieval system to improve subsequent searches. For instance, greater weight can be placed on any future matches between the terms in a query and the keywords of documents that the user has recognised as being relevant to a previous query involving those terms.

Information retrieval tools have supported numerous applications and are ubiquitous on the World Wide Web. It is, therefore, surprising that they have not been more widely adopted to support incident reporting systems. One explanation for this is that they cannot, in their pure form, be used to collate the statistics that are more easily extracted using relational systems. In a relational database, incident reports are classified according to the detailed components of a data model. It is possible to provide particular percentages for the numbers of incidents within each pre-defined category. In contrast, information retrieval systems avoid the pre-defined data models that have been criticised in previous paragraphs. Information retrieval tools make inferences based on the terms in a query and keywords associated with a document to determine whether or not it is relevant to the user [71] Many of these inferences are based on heuristic algorithms that cannot be guaranteed to satisfy the users' information need. Information retrieval systems make incorrect assumptions about the content of the document being retrieved and about the nature of the user's request. In consequence, it is difficult to rely upon the number of items retrieved by a query when generating statistics about the frequency of particular incidents. Further manual analysis must be performed to ensure that the retrieval tool has correctly identified all relevant incidents. As we shall see, this additional analysis can involve two different tasks. It is important to filter out any irrelevant 'hits' from the retrieved documents. This can have profound consequences for incident reporting systems. There may be insufficient resources to manually search through the many spurious matches that can be returned by some information retrieval tools. Conversely, it is important to ensure that any relevant documents have not been missed by the retrieval tool. This is a significant problem because users may fail to recognise a pattern of previous failures if similar incidents are not being detected by a retrieval system.

Information retrieval tools avoid the constraints of rigid data models by focusing on lexical features of documents. Relevant documents can be identified by looking for similarities between the words that are used in a query and those that are contained in an incident report. For instance, if the user issued a request to find 'all incidents of computer failure' then the retrieval system would look for any reports containing the words 'computer' and 'failure'. This example illustrates the potential strengths of this approach. Users can compose queries that do not require any understanding of an underlying relational algebra. This example also illustrates many of the problems that complicate this approach. Firstly, the retrieval system will have to strip out 'noise words' from both the query and the incident collection. This is important if any match is not to be overwhelmed by commonly occurring words such as 'and' or 'the' that occur in almost every sentence. Secondly, any implementation will be forced to process the query in order to recognise lexically related terms in the document set. The query term 'computer' should also match 'computers', 'computerised' as well as 'computational'.

Information retrieval, typically, depends upon the identification of concepts and terms that can be used to discriminate between the items in a collection. Words that commonly occur in all of the documents within a collection are unlikely to provide useful information about these concepts. For instance, 'function words' such as 'it', 'and', 'to', are necessary for the construction of grammatical sentences. They have a relatively high frequency because of their grammatical role but provide little help in identifying the content of a document. Other terms can be regarded as noise within particular systems. For instance, words such as 'clinical' or 'doctor' occur in most medical incident reports. If they were used as document keywords then a significant amount of indexing space and retrieval time would be spent filtering through values that are unlikely to help users discriminate between large-scale collections of incident reports. Unfortunately, we cannot simply strip out 'noise words' based on their frequency alone. For instance, the FDA's MAUDE system yielded more than 3,000 matches for 'software' incidents in January 2002. Such terms cannot be regarded as 'noise' even though they appear in many documents. They provide critical information about the nature of the events that they describe. Many information retrieval systems, therefore, rely upon a *negative dictionary* rather than raw frequencies [71]. These enumerate the words that can be ignored during the retrieval process. These include standard lists of function words. Negative dictionaries can also be supplemented by domain dependent lists provided by the end users of the system. For instance, a variant of the MAUDE system might deliberately exclude 'clinical' and 'doctor' as potential keywords. Clearly, the content of negative dictionaries can have a profound impact upon the performance of an information retrieval system. They must, therefore, be validated in consultation with the end-users of the system. The content of such dictionaries must also be reviewed as the nature of incidents, and hence of the language that is used to describe them, will change over time.

It is important to identify common concept in queries and documents even though they may not contain exactly the same lexical forms. One means of achieving this is through the use of stemming algorithms. For example, a query might contain the word 'error' and an incident might contain 'errors'. Any indexing must be robust to such plural forms. It must also consider variants, for example by deriving 'woman' from 'women'. Fortunately, there are standard techniques, including Porter's stemmer, that can be used to address these potential problems [71]. They extend the ability of the search engine to identify potential matches between lexical terms. They also reduce the number of keywords that are associated with documents. Plurals are stripped out, only the singular 'roots' are retained.

As mentioned, many information retrieval systems exploit the notion of 'inverse document frequency' as a means of identifying useful keywords. Rare words provide better discriminators than more frequent terms. In consequence, many retrieval systems will revise the weightings associated with particular keywords whenever new documents are entered into the system. Changes in the pattern of language used to describe incidents or in the underlying causes of adverse events can be reflected by changes in the weightings associated with particular terms. This creates a paradox in which the increasing frequency of particular incidents might result in lower weightings within the retrieval system. There are further complications. If we consider a document containing the term 'software failure' then this might provide a useful index within a collection of incident reports about medical adverse events. It would not, however, provide useful information in a collection that was entirely devoted to medical software failures. From this it follows that the discriminatory value of any index is determined by its ability to distinguish between the contents of that document and the other items in a collection. An extension of this argument is that the importance of any keyword for a document is determined not by the absolute frequency of that keyword within a collection but by the relative frequency of that keyword within the document compared to the frequency of the term throughout the collection as a whole. A word that occurs frequently within a collection can still provide valuable information about a particular document if it occurs even *more* frequently in that report. In other words, the background or 'noise' frequency of a work can be used to identify a threshold value. This can be distinguished from the signal value of the word if its frequency exceeds this limit [71].

The previous approaches focus on individual keywords. In contrast, a number of retrieval systems rely upon vectors of terms both to characterise queries and to index items in a document collection. In this view, each keyword represents a different dimension along which to compare a document to a query. The following vectors illustrate this technique using binary values. A keyword is either present or absent from a document. Variants of this approach rely upon weightings to indicate how often a particular word appears or how 'significant' that word might be in determining relevance within a collection of incident reports:

	Keyword 1	*Keyword* 2	*Keyword* 3	...	*Keyword N*
Document 1	0	0	1	...	1
Document 2	1	0	1	...	0
Document 3	0	1	0	...	1
...	1	0	1	...	1
Document N	0	1	0	...	1
Query	1	0	1	...	0

The simplest approach would be to take the inner product of the query and document vectors as a metric of similarity. However, vector-based information retrieval systems can go beyond the isolated use of keywords to look for patterns in a document collection. Matches may be based not simply on the direct relationship between a query and the document that matches it best but also on the transitive relationship between that document and other similar reports. The components of a query can be expanded to include keywords that are not explicitly mentioned by the user but which are also common to those documents that best match the users query. For example, a user might issue a request to identify incidents involving 'catheter' and 'lines'. It might be observed that many other reports which contain the word 'catheter' do not contain the word 'tubing' but do contain terms such as ‘tubing’. These partial matches might therefore be offered to the user during subsequent interaction with the system.

Unfortunately, vector-based approaches also suffer from a number of problems. Most users queries yield very few keywords and so their vectors can potential match a large number of documents in the collection. One solution to this is to construct a query vector both from the users most immediate request and from the keywords that have been extracted from previous search tasks. This creates the problem that documents which were incorrectly returned during previous sessions will continue to be returned during future interaction. Vector based techniques must also account for the problem of document length normalisation. This arises because longer documents are more likely to contain more keywords that shorter ones. Hence there is a greater likelihood that they will be returned in response to most queries. In incident reporting systems this relates to a tension between scope and verbosity. It can be difficult to distinguish between lengthy accounts that describe a large number of complex failures and those that simply use ‘more words’. The Swiss Anaesthesia Incident Reporting System (CIRS) provides good examples of this tension where reports of similar incidents involve IV lines range from under 100 words to over 1000 [756]. There are a variety of potential solutions to the problems of document length normalisation. In the CIRS system, this is less of an issue given the relatively small number of additions each month. In larger-scale reporting schemes, lengthy documents can be divided and indexed separately to ensure that keyword vectors reflect the changing content of each section. If the resulting vectors are very similar then they may be merged to prevent the generation of unnecessary indices. In practice, however, this leads to a host of further problems [71]. Brevity prevents a full analysis of techniques for document length normalisation and the interested reader is directed to Singhal et al [744].

The previous paragraphs have introduced some of the main issues that arise during the development of information retrieval systems. It should be apparent that this term covers a broad range of different approaches. Many of these techniques have still to be applied to support the development of incident reporting systems. There have, however, been a number of recent attempts to extend the benefits that these systems provide to support search and retrieval tasks amongst large collections of occurrence reports. In particular, the FDA have pioneered the use of some of these approaches in the web-based interface to their MAUDE reporting system for incidents involving medical devices [270]. As we have seen, this system is based around a relational database using techniques that were described in previous section. It also provides access through the Verity free-text search engine. This relies upon a lexical analysis that has much in common with the information retrieval techniques described in previous sections. From the users’ perspective, they can issue restricted free-text queries rather than being forced to compose more complex sentences using SQL syntax. Initially, users of the Verity interface to MAUDE are encouraged to enter either a single word, such as Catheter. This will yield only those incident reports that contain the exact spelling of the word that is entered. Alternatively, users can enter an exact phrase, such as Catheter line. This will yield records in which those words appear in the exact order specified by the query. Users can also perform searches involving multiple words connected by the AND operator, such as Catheter AND tubing. This retrieves records that contain both search words in any order and any location in the text being searched. The initial Verity interface provides users with information about how to perform these relatively straightforward lexical queries. The FDA also provide guidance on how to perform more complex retrievals. The OR operator can be used to find reports that contain one of two search terms. For instance, pregnancy OR folate returns documents that contain the word pregnancy or folate but not necessarily both. Parentheses can be used to form more complex queries. Quotation marks can also be used to explicitly denote that a literal match should be performed. Users can select documents that contain both pharmaceutical companies and stock by entering AND (‘pharmaceutical companies’, ‘stock’). The , comma operator returns documents containing at least one of the words specified using a ranking approach. The FDA’s implementation returns an ordered list; incident reports that contain the most occurrences of the keywords are given the highest rank. There is, however, no attempt to exploit the length normalisation algorithms mentioned in previous paragraphs.

It is also possible to create queries using the NOT operator. Ideally one might like to pose queries of the form NOT software to return every non-software related incident. The unrestricted use of such queries would create considerable computational overheads. This undermines variants of the indexing strategies described in previous paragraphs and would result in a form of exhaustive search over several hundred thousand records. Verity will, however, attempt to execute unrestricted queries involving the NOT operator. Issuing a request for NOT software in January 2002 returned more than 7,000 MAUDE records before the system ran out of resources and stopped the request. In contrast, the FDA recommend that users form queries that restrict the negated search term. The NOT operator 'finds documents containing the word that precedes it but that do not contain the word(s) that follows it'. For instance, pregnancy NOT folate yields incident reports with the word pregnancy but excludes any document that also contains folate.

The NOT operator demonstrates that many free text search facilities are not as 'intuitive' as they might first appear. They do, however, support the notion of proportionate effort. It is possible to perform literal keyword searches with minimal assistance. More complex query formation involves some additional thought. It might be argued that the implementation problems surrounding negated queries demonstrate that lexical forms of information retrieval offer few benefits beyond those provided by relational databases. This argument can, however, be challenged. In the case of relational databases, users must consider both the semantics or a range of relatively complex operators and the underlying data model that will be different for each database. In the case of lexical information retrieval tools, the user only has to understand the underlying concepts associated with particular operators. The proponents of these systems also argue that, in contrast to relational databases, most of the key concepts can be formed inductively without explicit training. Over time users will learn about the efficiency problems associated with unrestricted negation as they experience significant delays in processing their queries.

Previous paragraphs have described how the Verity retrieval tool searches for literal matches with the terms used in a query. This can create significant problems for many users. In particular, it can be difficult to search for all incidents involving particular manufacturers. The FDA acknowledge that 'when searching on company names, the search does not include variations of spelling or use of symbols such as hyphens, slashes, etc' [271]. However, the problems associated with exact literal match algorithms are exacerbated by the difficulty of data validation in large scale incident reporting systems. During the preparation of this book, I found numerous instances in which the names of manufacturers had been misspelt in the sections of the incident report that were searched by the Verity system. In consequence, users must exploit literal search facilities to identify incident reports that contain the correct spelling for a device manufacturers. They must then form additional queries to check whether any reports have been missed because those names were mis-spelt. This problem can be avoided in relational systems where manufacturers must be associated with one of a number of pre-defined attributes. The Verity system does, however, provide additional operators that can be used to address this limitation of using free text data during the analysis of incident reports. The ? question mark provides a wild-card that can represents any single character. For instance, the query ?ietermans would locate documents containing the words Viertermans, Fiertermans, Giertermans and so on. In contrast, the * asterisk represents one or more characters. A query of the form corp* would return documents containing corporate, corporation, corporal and corpulent.

The Verity interface to MAUDE also provides users with access to some of the stemming facilities that have been described in previous paragraphs. Queries that exploit this facility must include key terms using single quotation marks. For example, the query cath' finds catheter, cathlab, cathode and cathodic among others. This explicit approach to query formation using stemming can be combined with the <MANY> operator to count word densities in FDA incident reports. For instance, the query <MANY> cath' produces a ranked list in which the first document contains the most occurrences of words with the cath stem. In contrast to the comma operator introduced in previous paragraphs, Verity's <MANY> queries do perform length normalisation. Hence the FDA advise that 'a longer document that contains more occurrences of a word may score lower than a shorter document that contains fewer occurrences'. Verity offers a range of more complex operators that can be used to search for words within particular sections of an incident report. For example, the < NEAR/N> operator can be used to find documents that contain words within a specified distance of each other. For example, the query, balloon < NEAR/10> rupture, would locate all documents with the terms balloon and rupture within ten words of each other. Similarly, < SENTENCE> and < PARAGRAPH>

will find documents in which the specified terms are in the same sentence or paragraph.

The previous paragraphs have focussed on the facilities that the FDA's Verity tool provides for lexical information retrieval across the MAUDE incident collection. These facilities are built upon partial or literal matches between keywords in a document and the terms in a query. There are, however, a range of information retrieval techniques that make inferences about potential matches that go beyond the keywords that appear in a document. Many of these approaches rely upon thesauri that represent the relationships between keywords. In consequence, if there are few literal matches between a query and the documents in an incident collection then retrieval tools can look for matches between a query and other keywords that are in some way related to those in the document. Alternatively, the users' query can provide the basis for additional searches using terms related to those in the original request. Thesauri have been extended to include the following relationships:

- synonymy. Two expressions are synonymous if the substitution of one for the other does not change the interpretation of a sentence. For instance, cardiopulmonary resuscitation is synonymous with artificial respiration and heart massage. This relationship can also be used to connect acronyms to their associated terms. Hence CPR is related to cardiopulmonary resuscitation. Synonymy can also be used to capture conventional or authoritative keyword that replace less favoured terms used within a document or query. For instance, amyotrophic lateral aclerosis or ALS might be preferred to Lou Gehrig's Disease. Such relationships are critical in natural language queries that search for similar incidents describe from different perspectives. A variety of terms can be used to describe the same concepts depending on the geographical location of an incident or the functional role of the reporter.

- antonymy. This relationship is less commonly supported than the other forms in this list. Antonymy represents a pair of words which are related by an associative bond. These associations are often validated in empirical studies, or word associated tests, involving the potential end-users for a retrieval system. Antonyms are often revealed to have an opposite semantic relationship to the probe terms used in th studies. Hence many people will respond with the term victory when promoted with the word defeat and vice versa. As we shall see, there have been few attempts to apply this form of relationship to support information retrieval within an incident reporting system. It can, however, be argued that such techniques might be used to identify successful instances of a procedure rather than previous failures.

- hyperny/hypony. A hypernym designates a class of specific instances. Y is a hypernym of X if X is a (kind of) Y. In contrast, a hyponym describes a member of a class. A hyponym inherits all of the features of the more general hypernym and adds at least one feature to distinguish it from the high-level concept. For example, Lymphoma can be treated as a type of Neoplasm [71].

- meronymy/holony. Meronyms are constituent parts or members of something else. Hence, X is a meronym of Y if X is a part of Y In contrast, a holonym is the whole of which the meronym names a part. For instance, the cecum is the first part of the large intestine and is hence a meronym for the larger structure. Any query about incidents involving procedures on the large intestine might also return procedures involving the cecum.

This partial list only provides an indication of the relationships that can be used to expand on the keywords derived from a query or used to index a document. These techniques have not, however, been widely applied in either incident or accident reporting systems. Carthy has recently begun the first systematic examination of thesaurus-based retrieval techniques for incident reports in a project funded by the Irish Government. His work builds on automated topic detection and tracking systems. These applications enable their users to identify common threads amongst the publications of news and broadcast media. It also exploits the development of domain specific taxonomies, such as National Library of Medicine's Medical Subject Headings Thesaurus and the National Co-ordinating Council for Medication Error Reporting and Prevention's Taxonomy of Medication Errors [582]. Although these systems have been developed to support other applications, they can be directly applied to support the retrieval of incident reports [152]. It is also hoped that Carthy's work will encourage the development of techniques that are specifically intended to detect patterns of failure in reporting schemes. In anticipation of this research, the FDA's Verity interface provides access to more general facilities. The <THESAURUS> operator expands a search based on synonyms of the word(s) in a query. The example provided in the FDA documentation is that the query <THESAURUS> altitude will yield documents that include the terms height, elevation and altitude. The <SOUNDEX> operator expands the search to include words that 'sound like' the term(s) in a query.

The previous analysis of the FDA system again illustrates two key issues. Firstly, that advanced information retrieval systems can make powerful use of thesauri and similar techniques to make inferences about relevance that go well beyond the terms contained in either a query or a particular document. Secondly, that some training may be required if users are to fully direct or control the facilities that these techniques provide. In particular, the use of either the <SOUNDEX> or <THESAURUS> operators can lead to a rapid rise in the number of hits that are detected by the system:

> "If your full text search is broad, you may be attempting to retrieve more then the system limitation. If this happens, you will receive a message indicating that your record retrieval is incomplete. The system is not capable of retrieving any missing records over the limit". [271]

The following paragraphs will describe ways of measuring the adverse consequences of such information 'overload'. It is, however, important to consider recent attempts to control this problem by integrating information retrieval techniques and relational databases. For instance, Chapter 4.3 has already described how many incident reporting forms combine check-box questions with more open-ended questions that can be answered using free-text. Computer-based systems can, therefore, be developed to implement the strongly typed check-box information using relational techniques; each check-box represents a value for one of the attributes in a relation. The same system might combine this with lexical techniques for information retrieval so that users can search in a less directed fashion over the free-text descriptions of adverse events and near-miss occurrences. In theory this combine approach can offer numerous benefits. For instance, statistical returns that require deterministic answers to particular focussed queries can still be conducted over the data that is stored using a relational database. Less directed 'information mining' operations can exploit the free-text areas of each report. This complementary approach can also help to control the information 'overload' problem. If a thesaurus is used to expand query or document keywords then the mass of potential returns can be filtered by restricting the search to incident reports that match particular relational attributes. For example, the use might direct a request to find all incident reports relating to <THESAURUS> CATHETER so that it was only evaluated over reports filed by device manufacturers or, alternatively, by end-user facilities. If the report contributor were recorded as a field in the relational component of the system then users might be relatively confident that their query was restricted in the desired manner. Such a filtering would be less easily achieved using lexical variants such as NOT END-USER because the free-text accounts might not all have used the term END-USER and thesauri-based techniques do not guarantee to find all possible forms of synonym that may have been used by every contributor.

The FDA's MAUDE system provides a partial integration of their relational model and the Verity retrieval system. The lexical analysis is restricted to free-text areas of the device reports and does not cover any fields that are 'encoded' using numeric or other identifiers. The database elements that are examined by the Verity facility include: MDR Report Key; Manufacturer Name; Distributor Name; Brand Name; Generic Name; Model Number; Catalogue Number; Product Code and Adverse Event or Product Problem Description At first sight this might appear to offer the form of integration mentioned in the previous paragraphs. It is possible to use Verity to search over the attributes of relations within the MAUDE database. The FDA note that 'the Full Text Search cannot be combined with any other search options on the MAUDE search page' [271]. It is not possible to apply Verity to a subset of incident reports that have been filtered using the query language provided by the relational system. Hence it can be argued that the FDA provide a form of data-level integration rather than a full system-level integration. Information can be shared between

Verity and the relational format but both systems cannot easily be used to construct hybrid queries.

This section has argued that the information 'overload' problem might be overcome by using relational queries to filter the incident reports that are examined using lexical approaches to information retrieval. Unfortunately, this can only be a partial solution to what is a more complex problem than has previously been suggested. In large scale systems, relational filtering may still yield enormous numbers of incident reports in response to thesaurus-based free text queries. It is important to recall that MAUDE includes almost 400,000 records at the start of 2002. Analysts would still have to invest considerable time and energy to identify common features even if a query returned only 1% of the reports in the system. One solution to this problem would be to develop more precise data models within a relational system so that users could filter on more detailed features of an incident. This is unsatisfactory because increased discrimination tends to be achieved at the price of increased complexity. Alternatively, lexical analysis can be focussed more tightly to filter out spurious matches. For instance, by restricting the use of a thesauri it is possible to focus on a narrow selection of synonyms. Unfortunately, this increased precision will also typically result in worsening recall rates. The lexical analysis will miss reports that contain related terms and concepts, which were excluded by the narrow associations provided in the thesaurus.

Precision and Recall

Precision and recall are concepts that are used to assess the performance of all information retrieval systems. In broad terms, the precision of a query is measured by the proportion of all documents that were returned which the user considered to be relevant to their request to the total number of documents that were returned. In contrast, the recall of a query is given by the proportion of all relevant documents that were returned to the total number of relevant documents in the collection [221]. It, therefore, follows that some systems can obtain high recall values but relatively low precision. In this scenario, large numbers of relevant documents will be retrieved together with large numbers of irrelevant documents. This creates problems because the user must then filter these irrelevant hits from the documents that were returned by their initial request. Conversely, other systems provide high precision but poor recall. In this situation, only relevant documents will be returned but many other potential targets will not be retrieved for the user.

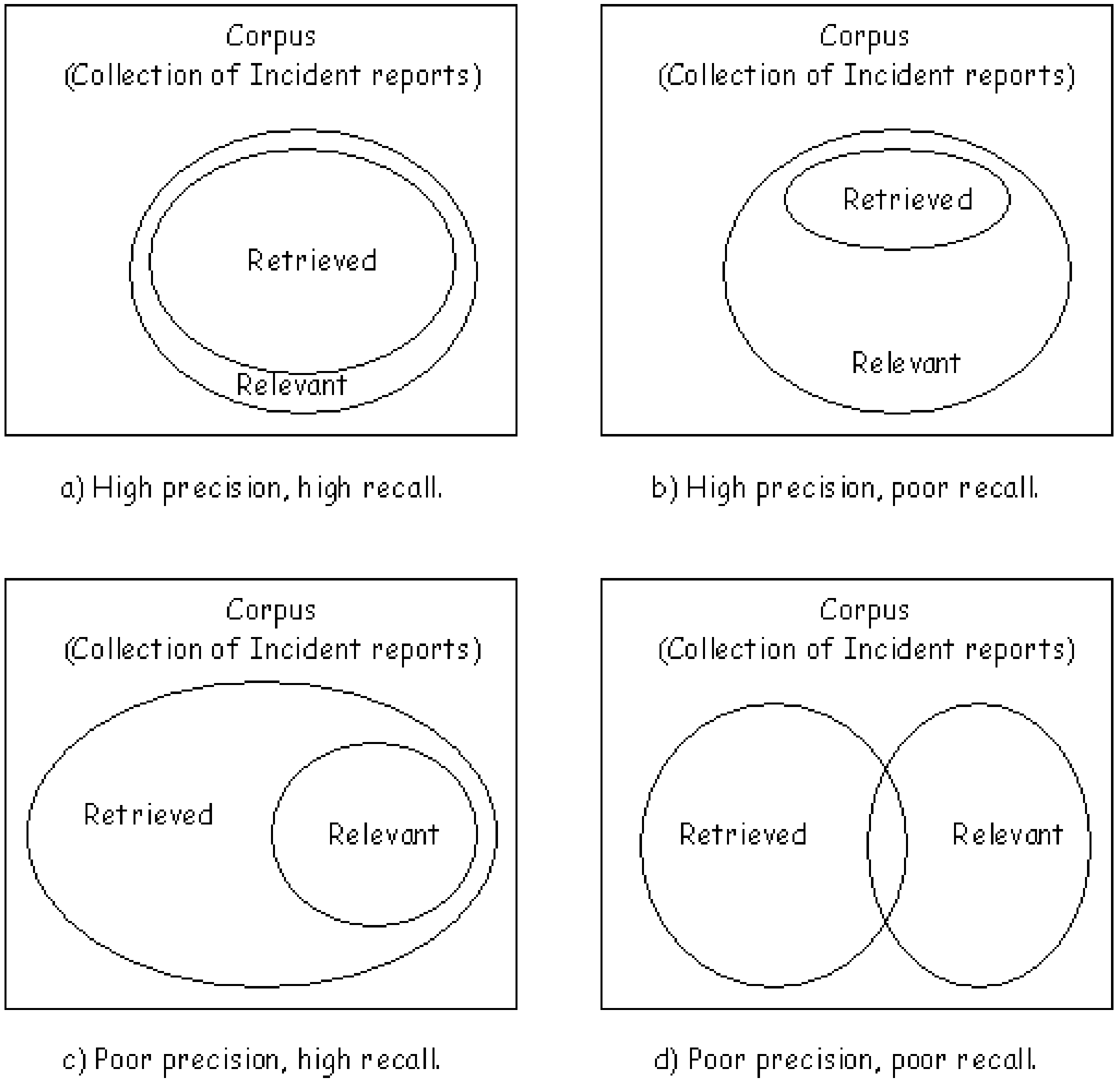
Belew [71] defines precision and recall in terms of the intersection between two sets:

$$Recall \equiv \frac{\mid Retrieved\_documents \cap Relevant\_Documents \mid}{\mid Relevant\_Documents \mid} \tag{14.20}$$

$$Precision \equiv \frac{\mid Retrieved\_documents \cap Relevant\_Documents \mid}{\mid Retrieved\_Documents \mid} \tag{14.21}$$

This is illustrated in Figure 14.5, which provides a high-level sketch of the relationship between precision and recall in information retrieval systems. Image a) reflects a query that achieved both high precision and high recall. Most relevant documents and no irrelevant documents were retrieved. In contrast, image b) represents high precision but poor recall. Only relevant documents were returned but many potential 'hits' were missed by the system. Image c) shows poor precision and high recall. Many irrelevant documents were retrived and hence the system is imprecise. In contrast, the query yielded all of the relevant documents so the system showed good recall. Finally, image d) shows both poor precision and poor recall. The query yields many irrelevant documents and retrieves very few that provide the required information.

Although the concepts of precision and recall are widely used in the evaluation of information retrieval systems there remains considerable disagreement about how to measure them in practice. These measures do not simply relate to system performance, they also relate to the corpus or collection that is being used. A system that achieves good recall rates on one set of documents may not achieve the same level of performance on another. This is particularly true for systems that rely upon thesauri. The meaning of key terms may differ considerably between domains and hence the system will have to be tailored to reflect differences in usage. For example, in the time series analysis of cardiovascular data the term 'leakage' is used to describe a loss of power from a


Corpus
(Collection of Incident reports)
Retrieved
Relevant
Corpus
(Collection of Incident reports)
Retrieved
Relevant
Corpus
(Collection of Incident reports)
Retrieved
Relevant
Corpus
(Collection of Incident reports)
Retrieved
Relevant

a) High precision, high recall.

b) High precision, poor recall.

c) Poor precision, high recall.

d) Poor precision, poor recall.

Figure 14.5: Precision and Recall

frequency band to several adjacent spectral lines which is typically due to the finite data set over which the periodogram is estimated. This is very different from more general applications of the term and hence appropriate relationships between synonyms would have to be explicitly encoded into an information retrieval system. Even if these relationships were encoded in a way that ensure good performance in the medical domain, there is no guarantee that the same system could easily be ported to, for instance, aviation. More research is urgently required to determine whether the linguistic characteristics of incident reports within these different fields can be used to support many of the retrieval techniques, mentioned above. Even within a topic, performance can vary depending on the nature of the documents that are contained within a collection. It is for this reason that information retrieval tools are, typically, evaluated using standard collections that provide a 'gold standard' for performance comparisons. This creates some problems for safety managers who want to exploit this technology. It is far from certain that the recall and precision values that can be obtained from a 'standard' corpus in information retrieval research will be mirrored in the operation of an incident reporting system.

There are further complications. The previous description of Figure 14.5 depicted precision and recall as properties of a particular query. For example, image a) shows a query that results in both high precision and high recall. It is important to recognise, however, that recall and precision vary dramatically depending on the query that is evaluated. For instance, if a thesaurus based system recognised a number of synonyms for a keyword then it is likely to provide high recall values for any query involving that term. In contrast, poor recall rates might be anticipated if the same system were presented with a query that did not contain any recognisable keywords. In consequence, comparisons between the precision and recall rates for particular systems must often be made in terms of specific queries on a particular data-set. If this were not the case then misleading values might be presented for carefully chosen requests. This raises the very practical concern that safety managers identify a 'realistic' test suite when attempting to evaluate the relative merits of these search engines. They must also identify an appropriate set of queries that reflect the likely information requirements for the end-users of the system. As we have seen, these can be difficult or impossible to predetermine given that the nature of incidents will change over time.

There are further complications. The images in Figure 14.5 assume that it is possible to unambiguously determine whether documents are either relevant or irrelevant to a particular query. There is no 'fuzziness' in the membership of *Relevant_Documents*. This reflects a strong assumption within the information retrieval research communittee that cannot easily be maintained for most 'real world' systems [422]. In particular, it does not characterise search tasks involving collections of incident reports. In many cases, it is difficult to be sure whether or not a particular document is relevant to a particular query. To illustrate this point, Jeffcott has recently conducted a study in which Risk Managers in Scottish hospitals were asked to read 8 reports of medical adverse incidents ranging from a problem in the use of a Doppler Fetal Heart-Rate monitor through to a morphine overdose [397]. Each incident was selected by a consultant and a senior nurse to provide a broad cross-section of the incidents reported to their Unit. The Risk Managers were asked to associate each incident with a number of broad categories that might correspond to retrieval requests using a variant of the 'expressed preference sampling' procedure developed by Fischhoff, Slovic, Lichtenstein, Red and Combs [251]. In simple terms, they were asked to rank whether or not they agreed with particular statements about an event using a 7 point scale. The results of this study showed a marked reluctance to use the extremes of the scale. The Risk Managers were unwilling to state that particular incidents did or did not exhibit a strong relationship to the questions that were posed. These responses undermine the binary distinction between relevant and irrelevant documents that is often assumed to exist in validation techniques for information retrieval systems .

It is difficult to under-estimate the importance of precision and recall to the application of advanced search techniques within incident reporting systems. In most other areas, including web-based retrieval, the trade-off between precision and recall can be characterised as either a performance or usability issue. In incident reporting schemes, these characteristics have safety implications. Low-recall results in analysts failing to identify potentially similar incidents. This can lead to litigation in the aftermath of an accident. Failure to detect trend information in previous incident reports can be interpreted as negligence. Conversely, low-precision leaves analysts with an increasing manual burden. They must filter the irrelevant documents that have been identified as hits by the search engine. This will result in omissions and 'errors' if fatigue or negligence undermine the manual filtering.

In spite of the problems in assessing the performance of lexical information retrieval systems, it is likely that these applications will play an increasingly important role in the computer-based dissemination of incident reports. The reasons for this centre on the need to provide technological support for the sharing of incident information between and within heterogeneous organisations. As we have seen, lexical information retrieval systems support the notion of 'proportionate effort'. Simple queries can be formed in a relatively flexible manner with only a limited understanding of the underlying data representation. More complex queries can be formed providing users understand the basic mechanisms involved in lexical retrieval, such as the use of a thesaurus to identify synonyms for keywords. It is not, however, necessary for users to learn the specific data representation that are associated with different incident databases. This contrasts strongly with the use of relational systems where it is necessary to understand the underlying data model before users can construct well-formed SQL queries. Such learning overheads would be of limited importance if safety managers only had to access a single incident database. Over time, novice users will gain experience in using the relations that lie behind systems such as MAUDE. Unfortunately, the lack of standardisation within many industries has combined with the increasing availability of web-based information resources to create a situation in which safety managers may have to understand the underlying data models associated with several different reporting systems. For instance, the UK MDA uses a relational model that is quite different from that used to describe US medical incidents. A small number of international initiatives are beginning to address this problem. We have mentioned the GAIN programme within the aviation industry in previous chapters [310]. This is, however, focusing more on the analytical techniques and underlying technological infrastructure necessary to support information sharing. Limited progress has been made towards the development of integrated data models for incident reporting that might enable users to exchange and search information from competitor companies in a convenient manner.

Several further factors increase the likelihood that lexical information retrieval systems will provide the technological infrastructure to support the dissemination of incident data between different reporting systems. The commercial impact of the world wide web is arguably the most important of these factors. Rapidly identifying relevant documents amongst a mass of other data is a key business requirement for many of the organisations and individuals that use the world wide web. In consequence, many companies are investing heavily in the technologies that support these tasks. This has produced tools that enable users to perform interactive retrieval tasks involving many millions of documents. These commercial developments offer further benefits. It is important to recognise that most of the documents that are placed on the web are unstructured. They owe more in common to the natural language accounts that are amenable to lexical information retrieval than they do to the more rigid relations within a database model. If a user issues a request for information about a medical product, they cannot expect that every device manufacturer will format the pages about their products in exactly the same way. This analogous to the situation facing safety managers and regulators looking for patterns of failure across several incident reporting systems. They cannot assume that all of these systems will exploit the same relational model. In consequence, lexical information retrieval systems offer a flexible means of analysing incident reports produced in many different formats by many different agencies. As we have seen, however, these systems do not yield the deterministic results that are typically required by statistical analysis. The precise number and nature of incidents returned by any query will depend upon the thesaurus that is being used and upon the discriminatory value of keywords that will change over time [71]. In consequence, common interchange formats for relational databases still offer considerable benefits for the exchange of incident data. The development of such common formats or schemas will not, however, resolve the problems associated with inter-rater reliability in the assignment of particular values to the attributes in a relational model. In consequence, I would argue that lexical retrieval tools will continue to provide the only feasible means of creating multi-national incident databases within the near future. Many safety managers and regulators already use mass-market retrieval systems to search for mandatory occurrence reports that are routinely placed on the web sites of the

CAA, NTSB and similar organisations. As we have seen, however, these more general tools do not support the domain specific thesauri that can be used to extend the scope of particular searches to achieve improved recall and precision. Similarly, it can be difficult to ensure that these mass market tools only retrieve potential hits on recognised sites. A search on catheter and incidents returns advertising material from manufacturers, research advertisements from government organisations, general news items from publishers, collections of papers published by particular individuals and so on. In order to address these problems, we have developed a series of web-crawlers that restrict the keyword indexing of documents to incident and accident reports on named sites. The terms that are used in the indexing and retrieval process are based on interviews with safety managers and regulators within the domains that we are investigating, principally rail and aviation safety, and are tuned according to the lexical frequency of terms within the collections of incident reports that we are studying. For instance, queries involving the term 'CRM' will yield incidents mention 'Crew Resource Management', 'Communication Failure' and so on. The intention is that these systems will provide feasible means for safety-managers and regulators to search for patterns of failure across the pages of incident reports that are increasingly being published via the world wide web [287, 529].

14.4.3 Case Based Retrieval

The previous section has identified a number of limitations of relational databases and lexical retrieval systems. Relational systems, typically, use strictly defined data-models to structure the information that is recorded about an incident. The many different individuals who enter or retrieve data from these systems often only have a limited understanding of these models. Further problems arise when changes are made to the components of a relational model; it may be necessary to manually reclassify hundreds of thousands of existing records. Alternatively, lexical search engines can be used to identify related terms in many different incident reports. Stemming techniques and thesauri can be used to expand queries or documents so that retrieval does not depend on literal matches. These approaches also avoid some of the problems associated with relational data-models. Users can enter natural language descriptions of each incident. Requests can be expressed as (pseudo) natural language queries. Unfortunately, as we have seen, the matching processes depend upon the frequency of terms within a collection. It may also be affected by relatively small changes within a thesaurus. In consequence, lexical approaches cannot easily provide the types of statistical returns that are required by regulatory organisations. A number of further problems relate to the precision and recall provided by these retrieval techniques. Precision is defined as the proportion of documents that the user considers being relevant within the total number of incidents that are retrieved. Recall is defined as the proportion of relevant documents that are retrieved against the total number of relevant documents within the entire collection. Hence an information retrieval system may have high recall and poor precision if it returns a large number of the relevant incidents in the entire collection but these incidents are hidden by a mass of irrelevant incidents that are also retrieved. Another system can have good precision and poor recall if it returns very relevant incidents but only a small proportion of those that pertain to the topic of interest. Many users have great difficulty in composing free-text queries that achieve a desired level of precision or recall. Most searches provide a small number of appropriate documents with many more irrelevant references. This poor level of precision can be exacerbated by inadequate recall. It is rare that any single query will yield all of the possible references that might support a user's task. These limitations can be frustrating for the users of mass-market retrieval techniques, such as web-based search engines. They can have more profound consequences for incident reporting systems. There are clear safety implications if a search engine fails to return information about similar incidents. A pattern of previous failure may be hidden by the poor precision or inadequate recall of some retrieval tools.

Case-based reasoning techniques relax some of the strict classification requirements that characterise more traditional databases. They do not avoid the concerns over precision and recall that affect other information retrieval tools. However, they often provide explicit support for users who must issue queries to identify similar classes of incidents within a reporting system. In the past these systems have been used to support fault-finding in computer systems, the design of wastewater treatment systems and route planning for mail delivery [455]. Ram provides an overview of this approach; 'case-based reasoning programs deal with the issue of using past experiences or cases to understand, plan for, or learn from novel situations' [693]. Most of these systems are based around a four stage process. Firstly, problem descriptions are used to identify previous similar cases. Secondly, the results achieved by attempts to address these previous cases are passed to the user. Thirdly, some attempt is made to extrapolate from the results of previous cases to the likely outcome of a similar approach being applied to the current problem. Finally, a generalised representation of both the old and new solutions are entered into the system so that future problems might benefit from any insights obtained during the analysis of the current problem. Ram's general analysis of case based reasoning can be applied to illustrate some of the potential advantages that this technology might offer for the analysis and retrieval of incident reports. Each 'case' can be thought of as an incident report. The attempts to resolve those cases can be seen as the recommendations that were made following those previous incidents.

The central problem of case-based reasoning is how to generalise from the specifics of a new incident so that it is possible to recognise any underlying similarities with previous cases. This is not as straightforward as it might appear. It is possible to identify at least three possible outcomes for any search:

1. *Exact match.* Two incidents are identical. In particular, we might be interested in those incidents that share both common causes and consequences [456]. Such similarities should not be discounted as unlikely given the increasing scale of many reporting schemes.

2. *Local divergence.* We might also want to identify partial matches between a new incident and previous cases. Two incidents share the same causes but an additional event or circumstance during one of the incidents led to divergent consequences. Alternatively, two incidents might have the same outcome but different causes. This reflects the causal asymmetry noted by Hausman [315] and described at length in Chapter 10.4.

3. *Global divergence.* Two incidents have no apparent similarity. They stem from different causes and result in different outcomes.

Case-based reasoning exploits some of these distinctions. For instance, an exact matching offers considerable efficiency gains because two cases can effectively be treated as a single more general case during the final stage identified by Ram, described above. Local divergence can be used to generate new indices that distinguish between cases with, for example, different causes or consequences.

Cases can be represented in a number of ways. Keyword or feature vectors, introduced in the previous section, can be used to represent whether or not particular terms are relevant to an incident. For example, the following narrative describes an incident from the MAUDE collection. This individual case might be represented by a vector that indicates the presence of indicative terms such as 'software', 'upgrade' and 'package':

> "During in-house software testing (of an ultrasonic analysis package), the manufacturer discovered unexpected software behavior in the generic tool kit when waveforms were inserted, resulting in correct calculation for the tricuspid valve regurgitant orifice area measurement.
>
> The software problem was found during in-house software development. It occurs when the user attempts to make a specific calculation in the cardiac calculations package and is related to a formula error. The software error has been identified and was corrected in a subsequent revision of the system software. Actions taken include customer notification of problem and installation of software upgrade to affected systems. It is important to note that there was no reported adverse event to a patient as a result of this event."

Stereotypes can be used to identify patterns of failure between the individual incident vectors. Each stereotype can be represented by the terms that a domain expert or contributor might use to describe particular incidents. For example, a stereotypical report of a software failure might include terms such as 'bug', 'crash', 'program', 'error', 'upgrade' and so on. If an incident report contained these terms then the associated similarity measure would be incremented each time that they appeared. In the previous example, the software stereotype score would be incremented for the terms 'program', 'error' and 'upgrade' because these are mentioned in the MAUDE account. The case-based reasoner returns a ranked list of stereotypes based on these similarity metrics. It is important to stress that the ranked list might return high scores for more than one stereotype. This is appropriate given that a software failure might be compounded by operator error or another form of adverse event. Each stereotypes can also be associated with particular remedial actions. For example, if software failure was returned as the highest ranked stereotype then the user of the system could be prompted to consult a guidance document on recommended procedures for resolving such incidents. This illustrates how lexical retrieval techniques can be intergrated into a case-based reasoning system.

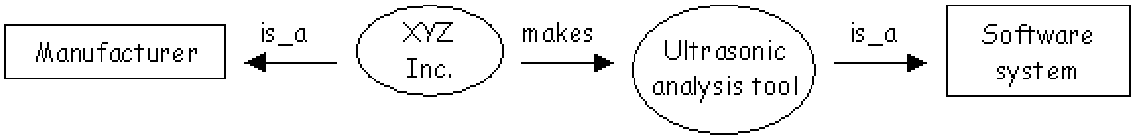


Figure 14.6: Components of a Semantic Network

The use of term-vectors is only one of several alternative approaches that can be used to represent and reason about common patterns in individual incidents. For instance, semantic networks can model an incident and more general aspects of the domain in which a failure occurs. In their simplest form, a semantic network can be thought of as a series of nodes and edges. The nodes represent objects and concepts in the domain of discourse and the edges represent relationships between them. For instance, Figure 14.6 represents part of the MAUDE incident report for the software failure that was cited in previous paragraphs. This diagram includes two different types of node. Rectangles are used to denote higher-level abstractions that may be common to many different cases. Software systems and manufacturers are likely to be involved in a more than one incident. In contrast, elipses represent particular instances of those abstractions. XYZ is a particular manufacturer, the company named in the report has been anonymised here. Similarly, an ultrasonic analysis tool is a particular type of software system.

Figure 14.7 extends the semantic net for the ultrasound software failure. As can be seen, the MAUDE narrative provides information about several different aspects of this incident. The failure mode was detected during a wave form insertion test. The problem was remedied by notifying the customers and by issuing a software upgrade. The fault might also have resulted in a patient injury. This diagram does not provide any information about a particular outcome for this incident. Figure 14.7 might, therefore, be revised to explicitly denote that nobody was injured as a result of this incident. This illustrates how the high-level abstractions in a semantic network can be used to provide an alternative to the lexical stereotypes, mentioned in previous paragraphs. Rather then relying on work frequencies to cluster similar incidents, semantic networks can be used to describe common relationships that characterise particular types of adverse event. Figure 14.8 shows how this can be done by removing all of the instance information from the previous semantic network. This leaves a high-level description not simply of the MAUDE incident that we have analysed but, more generally, of many different software-related failures.

Case-based reasoning systems can use the abstractions in Figure 14.8 in a number of different ways. They can be used like the components of a relational model to prompt uses for particular information whenever they enter information about an incident that seems to match with a particular stereotype. If, for example, the system determined that the new incident included a software-related failure then it might prompt the user to provide information about the detection method. This process can, in turn, contribute to the development of more appropriate abstractions. If a new incident was detected by an end-user facility then Figure 14.8 would have to be amended. The existing abstractions only consider Test Procedures as a means of detection. This process of case-based generalisation represents an instance of the final stage in Ram's taxonomy of case-based learning, mentioned above [693]. It also illustrates how the development of semantic networks from

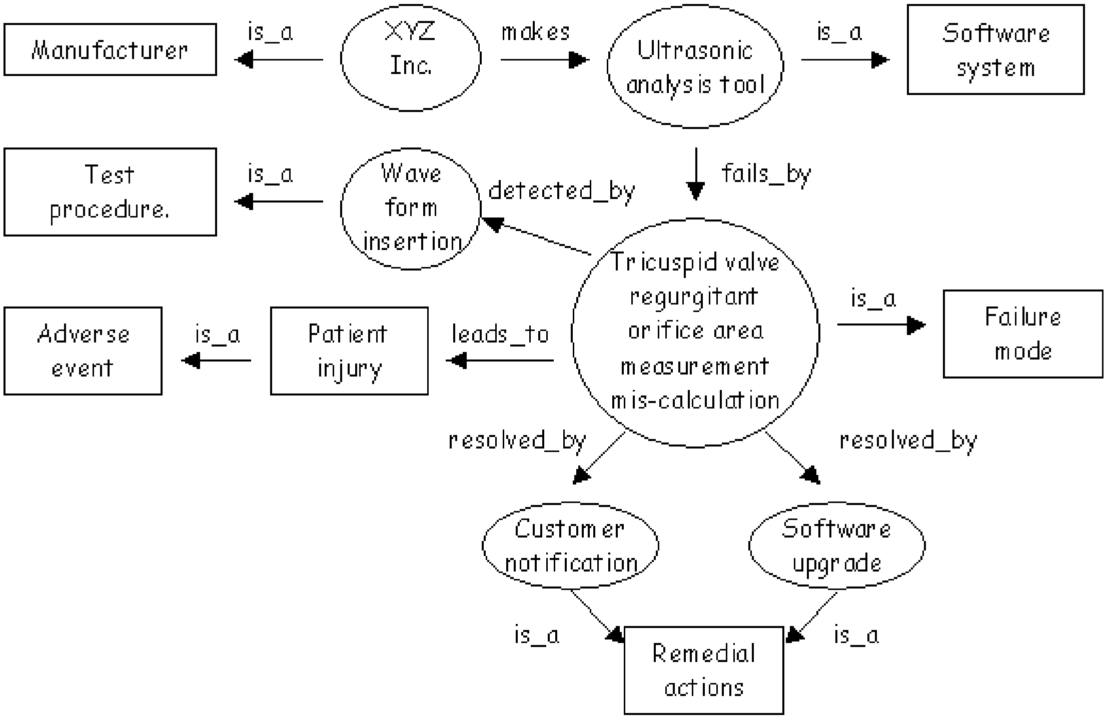


Figure 14.7: Semantic Network for an Example MAUDE Case

individual cases can help to create an ontology for particular types of incident. These ontologies provide a common reference point for the kinds of objects and relationships which characterise certain failures.

Figure 14.8 illustrates similarities between relational schemas and high-level abstractions from case-based reasoning systems. There are, however, strong differences between these two approaches. As we have seen, case-based learning systems are explicitly designed to cope with changes in the high-level models that represent previous failures. This contrasts with the costs that arise from changing the data model in a relational database. As we shall see, case-based reasoning systems also typically hide the detailed components of the underlying networks. Users are not expected to form complex queries that depend both on the underlying model and components of the relational algebra. Further differences stem from the matching algorithms that are used to determine whether a new case is similar to a previous incident. Both case based-reasoning systems and relational databases support instantiation or literal substitution. Similar incidents can be identified by looking for previous records with identical attributes. In addition, many case-based systems also exploit knowledge-based search techniques. These approaches extend the semantic networks shown in this chapter to support the thesauri-based approaches described in the sections on lexical retrieval techniques. For example, a search might be made through the previous cases to find incidents that were detected by tests which are synonyms of wave form insertion, such as wave form addition or wave form introduction.

There are many more complex variations on the general approach described in previous paragraphs. For instance, Kolodner pioneered many of the initial case-based reasoning techniques using a Dynamic Memory Model that was based on 'generalised episodes' [454]. These episodes form a hierarchical structure. For instance, at the highest level the MAUDE system describes episodes that relate to the failure of medical devices. These can be further sub-divided into episodes that describe software failures, human error and so on. Each 'generalised episode' is described in terms of norms, cases and indices. Norms are common to all of the cases indexed under a generalised episode. For instance, a normal expectation of all MAUDE reports is that they refer to medical devices. Indices discriminate between the cases in a generalised episode. For example, the components of Figure 14.7

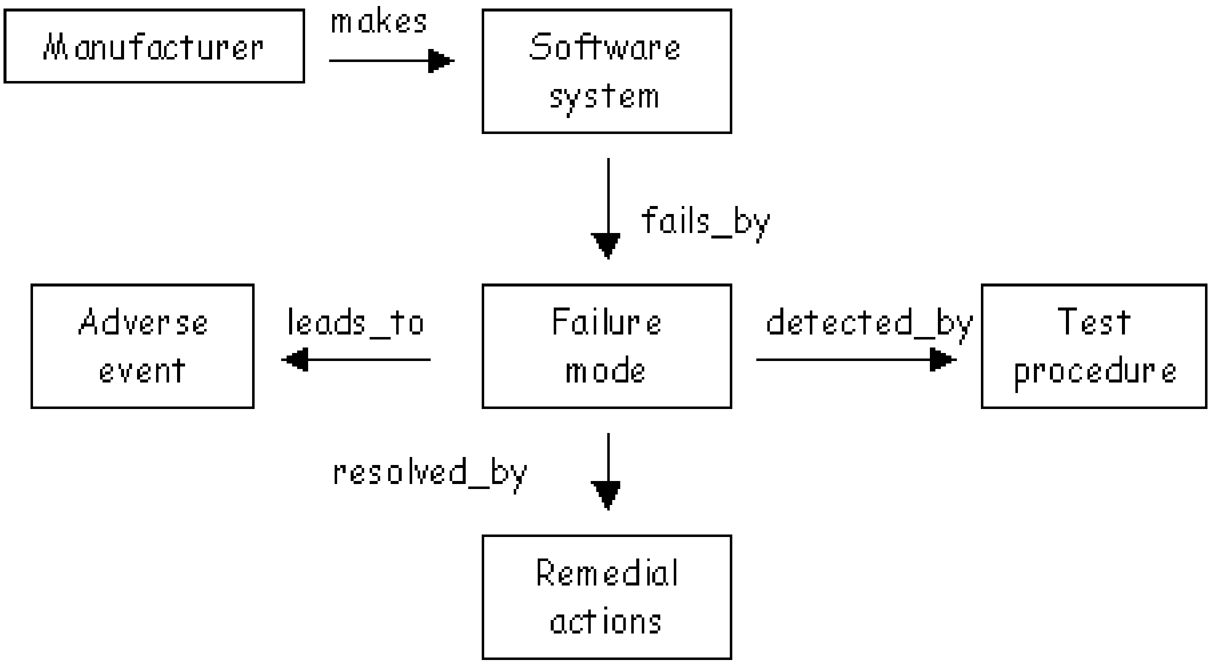


Figure 14.8: Using a Semantic Network to Model Stereotypes

might be used to index individual cases of software failure. A particular incident report could be identified by the software system involved, by the failure mode, by the detection method and so on. It is instructive to draw parallels between such system architectures and the distinction between general and particular causes that was introduced in Chapter 9.3.

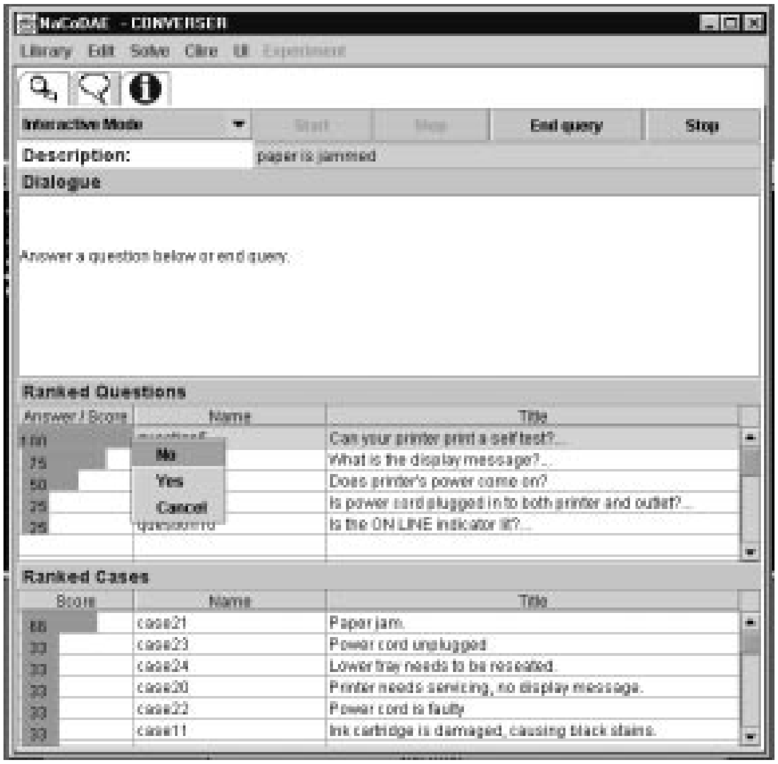
Kolodner goes on to describe how the hierarchical structure of generalised episodes can be used to search for similar cases. The system begins at the top of the structure by examining whether or not the new incident obeys the norms associated with the episode. For instance, a retrieval task with MAUDE might begin by asking whether or not the new incident involves a medical device. If the norms are satisfied then the system examines the indices associated with that episode. These point to successively more detailed episodes. For example, a MAUDE search task might go on to consider the generalised episode associated with software failures. As before, the system examines the norms and indices associated with this form of failure until eventually an index is found that points to a matching case. The match can be computed using a 'nearest neighbour' algorithm which associates measures of similarity with each of the values that are assigned to an index. For example, if the generalised example of software failure were indexed by device manufacturer then lexical similarity might be used to identify a potential match. In this case 'Arclights technology' might return a high similarity value for 'Arclights systems' and so on. This matching process can result in the extension of the 'case memory'. If a feature of the new case matches a feature of an existing case then a new generalised episode can be created. The two cases are discriminated by creating new indices within this generalised episode. This implements a dynamic memory structure because similar parts of two case descriptions are dynamically generalised into a further episode. The significance of this should not be underestimated. Implementations of the Kolodner approach will continually update their equivalents of the semantic networks introduced in previous sections. This is done automatically as new incidents are entered into the system and hence the approach avoids many of the problems associated with 'static' data models in relational systems.

The approach advocated by Kolodner has been elaborated by a number of other researchers. For instance, the 'category and exemplar' approach distinguishes between problem descriptors and the cases that are stored in the system. Users are assumed to be looking for previous cases that describe potential solutions to the situations characterised by a problem descriptor. This approach provides three different types of indices [67]. Feature links point from problem descriptors to cases or categories. These indices are called 'remindings' because they remind users of previous solutions. Case links point from categories to associated cases. These are known as exemplar links because they indicate those cases that provide examples of the higher level category. These exemplars are ordered in terms of how well they represent this category. Finally, difference links relate similar cases that only differ in a small number of features. Unlike Kolodner's approach where there is a strict hierarchy between generalised cases the 'category and exemplar' approach uses a semantic network to link higher level categories. This supports the generation of explanations during 'knowledge-based pattern matching'. For example, Figure 14.8 supports inferences about partial reports of similar incidents. Both reports might identify the same manufacturer and the same software failure mode. If only one report named the software involved then a partial match might be made because, from Figure 14.8, the same manufacturer makes software that has previously failed in the same manner.

The previous paragraphs have describe how many features of case-based reasoning systems can be used to support search and retrieval tasks in large-scale incident collections. There are a number of further benefits [4]. For example, conversational case-based systems address the problems of poor precision and recall that frustrate the users of probabilistic information retrieval systems. In this approach, users interactively answer questions that are intended to guide them along the indices that lead to previous cases. By providing feedback about the numbers of cases that match particular responses, users can iteratively refine their search tasks in an interactive manner. For example, an initial search on the MAUDE data might prompt the user to specify who was responsible for submitting the report, whether the report addressed a hardware failure, software failure or an operator error and so on. Associated with each possible response would be an indication of the resolution provided by the question. If for example, there were only four software related incidents in the system then the user would see that by selecting this possible answer then their search would be refined down to a relatively small number of candidate cases. If, in contrast, 'operator error' indexed several thousand cases then the user could be alerted to the potential need to further refine their search task. As can be seen, this interactive approach does not directly address the underlying problems of precision and recall. The case-based reasoner may still fail to return a previous case that the user might consider to be relevant to their query. Conversely, it might return a previous incident that the user does not consider to be related to their current search task. Conversational case-based reasoning does, however, enable users to interactively control the granularity of their search task. The iterative presentation and answering of questions guides the users towards similar cases and avoids the need for users to create valid queries using a relational algebra.

The efficiency of any interaction with a case-based system can be assessed in terms of the amount of information that a user must provide in order to identify similar incidents. An inefficient system might request a mass of contextual data that does little to focus the search process. For example, a MAUDE implementation might prompt for details about high-level 'norms' within Kolodner's Dynamic Memory Model, described above. These details are likely to provide only limited benefits during any retrieval task because they will be shared by all incidents in the system. A number of algorithms exist for increasing the efficient of case-based retrieval. For instance, decision tree techniques often assign relatively high priorities to indices that partition candidate cases into a number of near equal groups. Selecting any one of the available answers will exclude a large number of cases from the other groups. If partitions are of different sizes then there is a risk that the user will continually select the index with the largest number of remaining cases and the partition will be less effective. Other algorithms have been implemented to ask questions based on their frequency of use to discriminate previous cases by other users [4].

The US Naval Research Laboratory has exploited conversational case-based reasoning techniques in the development of their Conversational Decision Aids Environment (NaCoDAE) [641] Figure 14.9 illustrates how this system supports fault-finding tasks. In this example, NaCoDAE is being used to diagnose a problem with a printer. After loading the relevant case library, the user types in a free-text description of the problem that they are faced with. The tools uses this to perform an initial search of the available cases using a form of lexical search. NaCoDAE responds with two ranked lists. The



Description: paper is jammed

Dialogue

Answer a question below or end query.

Ranked Questions

Answer / Score	Name	Title
100	question5	Can your printer print a self test?...
75	[illegible]	What is the display message?...
50	[illegible]	Does printer's power come on?
25	[illegible]	Is power cord plugged in to both printer and outlet?...
25	question10	Is the ON LINE indicator lit?...

Ranked Cases

Score	Name	Title
66	case21	Paper jam.
33	case23	Power cord unplugged
33	case24	Lower tray needs to be reseated.
33	case20	Printer needs servicing, no display message.
33	case22	Power cord is faulty
33	case11	Ink cartridge is damaged, causing black stains.

Figure 14.9: US Naval Research Laboratory's Conversational Decision Aids Environment

first contains cases that are ordered using similarity measures that are based on the free-text query and the vector-based techniques that were described for probabilistic information retrieval. Each NaCoDAE case is composed of a problem description, some associated questions and, if appropriate, a description of remedial actions. The second list, therefore, presents a series of questions that are associated with the cases in the first list. The user can choose to select a possible answer to one of these questions as a means of further filtering their search. For example, they might indicate that they were only interested in cases for which their was a positive answer to the question 'was the incident detected by an end-user facility?'. The list of matching cases would then be revised to exclude those that were not detected by end-users.

In Figure 14.9, the user has typed 'paper is jammed'. The system has responded with a list of questions headed by 'Can your printer print a self-test'. As mentioned, this question guides the user in their retrieval task. If they did not understand the question then they can double click on the question to reveal a further explanation:

> "To perform a self-test, make sure that the printer is off-line and while holding down the ALT key, click the TEST button'.

If the user cannot follow these instructions they can continue the search by answering another question from the list. In Figure 14.9, the user has indicated that the self-test procedure failed. The cases displayed below can then be revised in the light of this additional information. This co-operative exchange of questions and answers will also help improve recall because the user can continually review the list of 'relevant' cases being retrieved at each stage of the process. If the user selects the 'Paper jam' (Case 21) then they will receive further information on corrective actions. The information encoding used by NaCoDAE can be illustrated by this example:

```
BEGIN QUESTION QUESTION5
   TITLE 'Can your printer print a self test?...'
   TEXT ‘To perform self test, make sure printer is OFF-LINE, and while
      holding ALT key, click the TEST key.’
   ANSWERS
      TYPE : YES_OR_NO
   WEIGHT
      MATCH : 10
      MISMATCH : 2
   AUTHOR david_aha
   CREATION DATE 7/30/91 TIME 15:18:33
   LAST_UPDATE DATE 7/30/91 TIME 15:18:34
END QUESTION
```

As can be seen, the initial self-test question includes information about how to reach a potential answer. It also states that the type of the answer must be a YES_OR_NO. The run-time environment provided by the case based reasoning tool interprets this information and presents the user with a drop down menu which constrains them to a ‘yes’ or ‘no’ answer. The weighting information can be used in a variety of ways. The simplest approach is to increment the weighting of any cases matching the users’ selected answer and a penalty for cases that do not match the selected response. The encoding also includes information that supports the maintenance of a case base by denoting the identity of the person who entered the question into the case-base and the date of last modification. Individual cases can be encoded in a similar fashion. As can be seen, the developer explicitly states the responses to particular questions that will increase the weighting associated with a particular case. In this instance, if the user selected ‘no’ in response to question 5 ‘Can your printer print a self test?’ then the match would be incremented in the manner described above. Conversely, if a self test was completed then the weighting would be decremented:

```
BEGIN CASE CASE21
   TITLE ‘Paper jam.’
   QUESTIONS
      Question5 :  ‘No’ (MATCH_WEIGHT : + MISMATCH_WEIGHT : -)
      Question25 :  ‘13 Paper Jam’ (MATCH_WEIGHT : + MISMATCH_WEIGHT : -)
   ACTIONS Action23
   CREATION DATE 8/15/91 TIME 10:56:51
   LAST_UPDATE DATE 8/29/91 TIME 18:42:1
   LAST_USED DATE 8/15/91 TIME 10:56:51
   NUMBER_OF_CALLS 0
END CASE
```

NaCoDAE’s encoding of individual cases identifies potential solutions. The paper jam case number 21, illustrated above, is associated with remedial action number 23. This is represented by the following formalisation. As can be seen, the proposed intervention is identified by a short title ‘Clear paper path and reseat paper cassette...’ as well as a more sustained series of instructions. These end with a final recommendation that if the problem persists, users should contact a service engineer:

```
BEGIN ACTION ACTION23
   TITLE ‘Clear paper path and reseat paper cassette...’
   TEXT ‘Jamming can be caused by crooked cassette, wrong paper type,
      wrong side of paper up, or sometimes by a dirty print bar or worn
      tractor wheels.  If the problem persists, contact a service
      representative.’
   AUTHOR david_aha
   CREATION DATE 8/15/91 TIME 10:56:40
   LAST_UPDATE DATE 8/15/91 TIME 10:56:42
END ACTION
```

Previous sections have described how NaCoDAE represents each case in terms of a free-text description, a set of appropriate actions and the answers to questions that help to classify the case. The MAUDE data-set readily provide descriptions for each incident in the form of the free-text reports that were associated with each record. MAUDE does not provide access to detailed information about the response to individual incidents. We cannot, therefore, directly encode MAUDE records within NaCoDAE. Our initial studies overcame this problem by referring the user to a range of documents provided by the FDA about appropriate responses to general types of device failures, including recall and emergency response guidelines [259], reporting delegation procedures [254] and risk management documents [278].

It is harder to identify appropriate questions that might be used both to partition the data set and to guide the users' search. The most straightforward approach is to derive questions directly from the existing relational data model. For instance, users might begin a search by answering the question 'what was the outcome of the incident?'. They could then select an appropriate response from the alternative answers 'death, injury, malfunction, other'. This might result in the retrieval of a number of cases that contained either positive or negative answers to the question 'was the anomaly reported by a manufacturer?'. This question would then be presented to the user as a way of further refining their search using information that was common to the cases from their initial query. All of this data is readily identifiable from the existing MAUDE database. However, it is important to ask whether this encoding would offer any benefits over the traditional database approach? The first benefit is that NaCoDAE does not associate an answer for every question with each case in the system. This is appropriate because, as we have seen in Chapter 4.3, there can be considerable uncertainty about the causes and consequences of some incidents. For instance, the person submitting the form may not know how it was resolved. NaCoDAE actively exploits the absence of information because it helps to distinguish between different cases. If a user decided not to answer a question then their search will retain cases with these 'unknown' values. However, if they select a definite answer then these cases will be excluded along with cases that are associated with the alternative answers to that question. In contrast, most relational implementations specifically prohibit 'absent values' from the fields of a record. Many relational systems, therefore, resort to using 'other' as a potential value that can be recorded. This does not resolve the problem, however. There is an important distinction between other' which implies that a definite response is known but is not supported by the system and 'unknown'. Some relational systems support the distinction by including both 'other' and 'unknown'. Unfortunately, this creates frequent problems during the training and appraisal of coders who must be reminded of the difference between these two potential values.

The previous paragraphs have described how our initial application of NaCoDAE was restricted to the information that was included in the original MAUDE reports. The relational data model that supports the existing database only provides limited causal information. Most of this detail is embedded within the natural language accounts that are associated with each incident report. Previous sections have argued that lexical retrieval techniques can be used to identify common features in the language that is used in these accounts. Unfortunately, there is no reliable automatic means of extracting causal information from these natural language accounts. We, therefore, decided to re-code our MAUDE sample data to demonstrate that case based reasoning tools, such as NaCoDAE, can be used to support the direct search for common causal factors. This builds on previous work in the application of case-based reasoning to 'small-scale' incidents by Koornneef [456]. The first stage of this new work was to perform a causal analysis of the incident reports. This followed the Eindhoven classification technique described in Chapter 10.4 [841, 845]. The causal analysis associated each incident with a number of the leaf nodes shown in Figure 11.10. Fr instance, the following natural language description provides an informal account of the potential causes of an adverse event involving an insulin infusion pump:

> "Patient treated at hospital for hyperglycemia. Pump not returned for evaluation... Manufacturer could not evaluate the pump, as the patient did not return it. User error likely caused event. Continuous insulin infusion therapy requires that the patient continually assess the impact of such factors as their caloric intake, activity levels and other medical conditions and/or treatments on their blood glucose level. The ther-

> apy also requires periodic self-testing of actual blood glucose levels. Failure to monitor and/or adjust the insulin amount appropriately will result in erratic blood glucose levels. Extreme excursions from normal blood glucose levels can result in conditions such as hypoglycemia or hyperglycemia. Patients experiencing these conditions may require hospitalization and medical intervention to preclude serious medical conditions including death."

Previous chapters have identified a number of criticisms that might be made both about the style and content of this account. Strong assumptions are made about the patient's role in the incident. For instance, we are told relatively little about the information and guidance that the physician offered to support the use of the device. Chapter 10.4 has presented techniques that can be used to address these concerns. For now, however, it is sufficient to observe that the causal factors related to this incident might be categorised using the HRM (Human Behaviour: Monitoring) and PRF (Patient Related Factor) nodes from Figure 11.10. The incident was caused by a failure on behalf of the patient and clinician to monitor their use of the device. The incident was also caused by specific patient related factors, including their underlying medical condition that led to the hyperglycemia. We then encoded this analysis as positive responses to the questions 'was there a failure in human monitoring?' and 'was the incident exacerbated by patient related factors?'. Conversely, there was no evidence of a device related failure (TD, TC or TM). This was encoded as negative responses to the questions 'was there a problem with the device design?', 'was there a problem with the device construction?' and 'was there a problem with the device materials?'. As mentioned, we did not have to indicate whether or not each element of the Eindhoven classification was a causal factor for every incident. Answers were only provided when there was definite evidence for or against certain causal factors. Instead of questions about the facts known for each incident, such as the name of the device or the manufacturer, these changes support the classification of cases or incidents by the results of the causal analysis. Not only does the NaCoDAE application support direct queries of the form 'who reported the incident?' but it also supports searches that look for complex combinations of causes such as 'what incidents were not reported by manufacturers but were caused by a lack of monitoring on behalf of the device user or clinician?'. Such queries cannot easily be satisfied using conventional databases and information retrieval engines.

The previous paragraphs have described initial attempts to apply case-based reasoning as a partial solution to the problems identified for relational databases and lexical information retrieval systems. As we have seen, however, many case-based systems draw upon ideas that were originally developed to support these more common applications. For instance, NaCoDAE relies upon a lexical analysis to perform the initial identification of candidate cases and questions. Similarly, the semantic networks of many case-based systems can be thought of as dynamic versions of the data models that underly relational systems. There are also strong differences. In particular, most case based systems do not require an initial domain model. The classification emerges over time as new cases are added to the system. This is a significant benefit given the widespread disagreement that exists over appropriate incident classification schemes [419].

A number of further issues must be addressed before case based reasoning techniques can be widely applied to support the storage and retrieval of incident reports. For instance, it is unclear how to provide the system with feedback when users disagrees with the matches that are proposed for particular incidents. This is complicated because such matches often depend on indices that have been automatically inferred by the system. It is, therefore, important to provide the user with information about the reasons why the system identified a target incidents as being similar to the one under consideration. Some systems address this issue by simply showing the user a trace of all of the factors that match between the situation that they are describing and the one that has been retrieved. Under such circumstances, the user can then either revise their interaction with the system or alter the labels associated with the case that was erroneously retrieved. The user might provide additional indices to distinguish the new incident from the case that was incorrectly matched. This can create problems if arbitrary users are permitted to alter the indices that are generated by the system. Different users are likely to disagree about the appropriateness of a particular match. It is for this reason that NaCoDAE records authoring information with the insertion of new cases and questions into a case library.

This chapter has focussed on the use of case based reasoning systems to identify patterns during the retrieval of incident reports. There is a danger that this focus will obscure the main motivation behind the development of this technology. Case-based reasoning systems were originally intended to help users solve problems and make decisions. Most previous applications of this technology, therefore, also include some assessment of how effective a proposed intervention was in response to previous cases. We have not been able to introduce this information into our initial studies using the NaCoDAE system because MAUDE does not assess the effectiveness of interventions following individual device failures. If this data were to be made available then regulators and analysts could use the case-based retrieval facilities of NaCoDAE to ensure that they respond to situations in a consistent manner. Users could also determine the circumstances in which a particular intervention had previously been effective. Without such assistance, there is a danger that the system would consistently advocate the wrong intervention. The next chapter, therefore, focuses on techniques that can be used to monitor the effectiveness of incident reporting systems and the recommendations that they produce.

14.5 Summary

Previous chapters have described the elicitation and investigation of adverse incidents and near-miss events. We have also considered a range of different techniques for presenting the findings of these investigations. In contrast, this chapter has looked at the issues that arise when regulators and safety managers must disseminate information about these safety-related occurrences. It is important not to underestimate the scale of this task. For example, the UK MDA provides information on approximately 7,000 incidents each year [540]. The US FDA's MedWatch program generates well over 300 incident-related publications each year [272]. The tasks associated with disseminating this information are exacerbated by the tight deadlines that must be met if safety managers are to be provided with the information that is necessary to respond to adverse vents in a timely manner. The MDA has a commitment to issue Hazard Notices within 20 days of notification, safety notices should be issued within 90 days. There are also financial pressures. The MDA are expected to meet these targets while at the same time achieving 2% efficiency savings per annum.

The pressures of time and of economy have led many reporting agencies to carefully consider who should receive the information that they disseminate. Some systems operated closed distribution policies where reports are only passed to a few named individuals within an organisation. Horizontal systems distribute to safety managers within other companies in the same industry. Vertical distribution schemes disseminate reports widely within the same company. Parallel reporting systems distribute reports to companies that operate similar processes in a range of different industries. Open distribution policies place few restrictions on the recipients of incident reports. Although we have identified these general approaches, many organisations operate hybrid techniques. For example, the MDA distribute Safety Notices through the Chief Executives of Health Authoritities, NHS Trusts and Primary care Trusts as well as the directors of Social Services in England. This represents a parallel approach to dissemination because each of these individuals may be responsible for similar healthcare systems that operate in very different contexts. However, each of these individuals is then responsible for further disseminating the Safety Notices widely to 'all who need to know or be aware of it' [536]. Hence this second stage dissemination opens up access to a far wider audience.

The carefully designed distribution policies that have been devised by many reporting agencies are often undermined by alternative communication channels. For example, informal anecdotes and 'war stories' provide both a powerful means of self-help and a dangerous source of rumour depending on the information that is conveyed and the context in which they occur. These informal channels are becoming increasingly important as technological innovation is increasing individual access to wider distribution media. In particular, the development of Internet chat rooms and of 'special interest' web pages has led to the dissemination of many 'alternate' accounts for incidents and accidents. The press and broadcast media provide further means of disseminating information about adverse events. They may be used to publicise the findings of an official investigation. They can also disseminate the results of journalistic investigations which are, typically, triggered by members of staff who feel that safety-related information must be disseminated to a wider audience.

Incident reporting agencies can recruit a range of technologies to implement the distribution policies, described above. These range from conventional paper-based publications through to increasingly complex, computer-based storage and retrieval systems. Paper based resources have numerous benefits. They are accessible to a wide audience and impose few additional technological requirements either on the publishers or the recipients. Unfortunately, they can be costly to produce and are difficult to disseminate in a timely fashion. There are also limitations in the types of information that can be captured in books, pamphlets and journals. Some information can be better conveyed using more dynamic media such as video images of incident locations and computer reconstructions of likely events. Finally, it can be difficult to ensure that all readers receive a copy of periodic updates to paper-based incident reports. There is a danger that some safety-managers may retain printed documents that contain obsolete recommendations.

Many of the distribution problems that are associated with paper-based documents can be addressed through the use of fax and telephone based mechanisms. For instance, pre-recorded messages can be accessed by telephone so that the potential recipients of an incident report can determine whether or not to request a printed copy using more conventional means. Alternatively, fax machines can automatically send updates to many thousands of telephone subscribers. This can be done over-night or during periods when the necessary equipment is likely to be idle. Unfortunately, the low resolution of most fax devices and the relatively unreliable infrastructure can create problems if these approaches are used as the primary means of dissemination. Increasingly, reporting agencies view this form of technology as an interim measure while the intended recipients of their documents acquire the necessary support to access computer-based resources.

A range of issues must be considered by any organisation that is considering using computer-based systems as a means of disseminating safety-related information. They must consider whether machines will be isolated from the security concerns that are associated with many local and wide area networks. They must consider whether the information that is held on a machine is to be disseminated by transient media, such as email, or by more durable forms of secondary storage, including CD-ROMS. They must consider the way in which individual reports will be formatted. For example, variants of the HyperText Markup Language (HTML) and Adobe's proprietary Portable Display Format (PDF) are both emerging as standards for the transmission of incident reports over the web. Each of these approaches offers radically different support for the dissemination of incident reports. PDF provides better support for the local generation of printed documents. HTML is more easily indexed and searched by a wider range of automated systems. Investigation authorities must also approve the access control mechanisms that are intended to secure their information resources. They must consider who has the right to read an incident report. They must also consider whether those initial readers have the right to disseminate the report more widely once they have received it. These access control mechanisms must be implemented using techniques such as public and private key cryptography. These techniques can be used to establish that the information has been sent by an official source and that the recipient has the correct permissions to access any data. Digital watermarks can also be used to ensure that incident information has not been altered by a third party. Finally, reporting agencies must also consider accessibility issues. Many schemes operate within regulatory and legal frameworks which help to ensure that the use of particular technologies does not prevent potential recipients from reading an incident report. Most commonly this is interpreted as a requirement to provide information in a format that can be accessed by individuals with a visual disability, for instance using a screen reader. Some legal and regulatory requirements have wider implications, including the need to perform usability evaluations to establish that computer-based resources can be operated by a wide cross-section of potential users [687].

These issues are generic because they affect the application of computer-based technologies to support the dissemination of incident-related information. There are, however, a number of more specific concerns that stem from the use of particular computational techniques in this domain. For instance, most existing systems rely upon relational databases. These applications structure the storage and retrieval of information using a static data model that must be carefully designed before the system is implemented. These models can be refined to improve efficiency both in terms of the storage space that may be required and in terms of the access speed for individual reports in large scale systems. Relational data models also help to ensure that data is not omitted or needlessly repeated. There are further benefits. Relational data models associate particular fields of information, or attributes, with key entities in the application domain. For instance, the FDA's MAUDE system is structured around manufacturer, device and patient 'entities'. The attributes associated with these key entities help to define the minimum information that must be recorded about each incident. This increases consistency between individual incident reports. Unfortunately, many of the benefits of relational databases can also be interpreteted as potential weaknesses. For instance, these is often considerable confusion about the values that must be entered into the individual fields of a relational system. The MAUDE system supports the distinction between a 'generic name' and a 'brand name' that can be confusing without further explanation. Such problems can also frustrate information retrieval using relational systems. Queries must, typically, either be pre-formatted or composed using a variant of the relational algebra. If queries are pre-formatted then it can be difficult for designers to anticipate all of the questions that users might need to pose of the incident data that is collected. If 'raw' queries are to be constructed from the relational algebra then users must not only be very familiar with the underlying data model but they must also have some understanding of the particular operators supported by their database management system.

A final set of limitations stem from the static nature of many relational schemas. As mentioned, most database applications structure the storage and retrieval of infromation around a number of tables or relations that are 'optimised' to improve the efficiency of a resulting application. These tables must be 'pre-programmed' into the system. In consequence, it can be difficult to develop appropriate models if safety managers or regulators are unsure about the precise nature of the incidents that will be reported or the information that they wish to capture. This might seem like a trivial requirement; the operators of a reporting system should have a clear idea of the information that they wish to elicit before starting a scheme. Unfortunately, things are rarely this simple. Even if it is possible to identify information requirements before a system is established, those requirements are highly likely to change over time. For instance, changes in production techniques may lead to new questions being asked about the circumstances in which an incident occurred. This would force programmers to refine the attributes in a relational data model. Similarly, if a relational model were to include causal information then changes might have to be made whenever new causal factors were identified. This would raise particular problems if previous incidents in the database were not re-classified using the new causal model. For instance, if a system added 'high workload' as a new cause in January 2002 then all the system would only recall incidents after this date even though there may have been 'high workload' incidents received before this date. These earlier failures would not have been recorded in this way because the relational model operating before 2002 did not support this distinction. For systems, such as MAUDE, that contain several hundred thousand records the maintenance issues associated with relational systems can impose a considerable overhead upon systems administrators.

Many incident reporting systems avoid the limitations of static relational models by restricting the information that is encoded within the fields of the database. These fields only record information about entities that will not change over the lifetime of the system. Every medical device will have a manufacturer, every incident report will be submitted by a contributor and so on. The remainder of the contextual and causal information that will change over time is recorded in a free-text description of the adverse event. This approach is adopted by the MAUDE system. Free text information retrieval offers numerous benefits for the maintenance of large-scale incident reporting systems. Many of these techniques do not assume that users have any knowledge of the underlying implementation techniques. Nor do they require the use of complex relational algebras. In contrast, users are encouraged to form natural language queries. Typically, key terms are extracted from these queries. There terms are then compared against the indices that point to individual narratives. Stemming techniques and thesauri can be used to ensure that lexical retrieval systems detect matches even though users do not enter exactly the same literal terms that are indexed by the system. Hence, 'fail', 'failed', 'failure' and 'fallible' can all be recognised as referring to similar concepts. If there is a sufficient match between the terms in the query and the index terms in the document then the system will propose a potential match to the user. This lexical retrieval clearly depends upon there being a minimal distance between the language used in the query and the index terms. If users' continually uses words that the system does not recognise then it will be difficult to identify appropriate documents.

Expert advice can be used to guide the selection of appropriate indices. This advice can be validated against records of queries performed with previous versions of the system. Index terms can be identified by a lexical analysis of word frequencies. This has the benefit that index terms can be revised to reflect changes both in the language that is used to describe incidents and, ideally, in the nature of the incidents themselves. Any changes in word frequency will be accounted for each time the indexing program is run. One side effect of this is that free text retrieval systems avoid the limitations of more static relational schemes. A potential problem with this approach is that requests will not always return the same results because the indices depend on the changing frequency of terms used in the collection.

Further problems arise because it can be difficult to ensure both high precision and high recall over a broad range of user queries. Precision refers to the proportion of relevant to irrelevant documents that are returned in response to a query. Recall refers to the proportion of relevant documents that are returned to all relevant documents held in a collection. Both of these concepts are well illustrated by the current generation of web technology. Many search request now provide hundreds of potential 'hits'. Many of these will not be relevant to the users query. The manual process of sorting through these many irrelevant matches stems for poor precision. Similarly, the same request may not return all of the potentially relevant information. There may be sites that could have provided exactly what the user required but which were not recognised as being relevant. This illustrates poor recall. In the context of incident reporting, each of these concepts has considerable significance. A failure to retrieve a similar incident in the past may mean that safety managers fail to detect an emerging pattern of failure. Poor recall can, therefore, lead usrs of a system to underestimate the potential risks of any recurrence. Similarly, if a request returns many dozens of incidents that the user does not consider to be relevant then they may be dissuaded from performing the necessary manual filtering that might have identified previous similar incidents. Such poor precision can impose considerable burdens upon the finite resources of many incident investigation agencies.

A range of solutions have been proposed to avoid the limitations of relational databased and lexical information retrieval systems. As mentioned, the FDA's MAUDE system implements a hybrid interface that provides access both to a relational database for directed search and a free-text retrieval system for broader queries. Case-based reasoning tools have recently been identified as a further alternative [456]. These systems, typically, avoid any predetermined data model. Instead, they will automatically reconfigure indices as new cases are entered into the system. There are several ways of achieving this. For instance, the Dynamic Memory Model distinguishes between 'generalised episodes' that collect together similar cases and indices that are used to distinguish between each of the particular cases that represent instances of a 'generalised episode'. Alternatively, 'category and exemplar' approaches introduce several different types of indices some of which indicate the degree of closeness between a higher level category and the incident reports that are exemplars of that category. These approaches offer significant benefits to incident reporting systems. Not only are they based upon dynamic classification techniques, most of these tools have been deliberately designed as information support systems. Individual cases are usually formulated as descriptions of problems. These are then associated with remedial actions. Hence it is possible to find out what other circumstances might prompt similar interventions. It is also possible to determine whether the same incidents are provoking the same reaction. These standard features of many case-based reasoning tools must be explicitly designed into mass-market relational databases.

The chapter has closed by describing initial attempts to apply the US Naval Research Laboratory's Conversational Decision Aids Environment (NaCoDAE) to store and retrieve reports of device failures from the MAUDE collection. NaCoDAE is a conversational case-based reasoning tool. Users provide an initial free-text query. This is used to identify an initial set of matching cases. This initial match is then analysed to identify a series of questions that might best be used to discriminate between these individual incidents. The user is then prompted to answer a list of these questions. For example, the user could choose to select the answer 'yes' to the question 'was the incident reported by an end user facility'. Each time they select a response, the system will automatically revise the set of matching cases and the list of questions. The efficiency of the entire system can be judged in terms of the number of questions that must be answered before the user is satisfied that they have identified a potential match. This ‘conversational’ approach helps the user with the problems of query formation. They are continually prompted to answer questions that are intended to guide their search task. This approach also helps the user to control the number and nature of potential matches. It is a trivial task to ‘cancel’ a response to a question if it narrows the number of potential matches too rapidly.

As mentioned, an important benefit of case-based tools is that they explicitly support the association of incident descriptions and recommended remedial actions. They do not, however, guarantee that those remedial actions will either be effective or appropriate. The following chapter, therefore, described ways of monitoring the effectiveness of the interventions that are identified in response to adverse events and near-miss incidents.