Pergamon

European Journal of Cancer Vol. 33, Suppl. 3, pp. S44–S47, 1997

Printed in Great Britain
0959-8049/97 $17.00 + 0.00

PII: S0959-8049(97)00023-3

The Design of National Dose Registries

J.P. Ashmore

Bureau of Radiation and Medical Devices, 775 Brookfield Road, Ottawa, Ontario, Canada K1A 1C1

INTRODUCTION

THE DESIGN criteria outlined in this paper are based on our experience in the development and the operation of the National Dose Registry (NDR) in Canada. Many, but not all, of the criteria presented have been incorporated in the current NDR. The criteria described are thus for an 'ideal' registry. While descriptions of existing registries are available in the literature [1–3], there is little on the actual design of large-scale registries for radiation workers [4]. Registers for radiation records have some similar requirements as disease registers (for example, problems related to identification information will be common), but the actual handling of radiation dose records cannot be addressed by these registries.

PURPOSE OF THE REGISTRY

Most National Dose Registries are multi-purpose and in many cases have evolved from dose record-keeping systems used for regulatory purposes. Generally speaking, the applications of a registry include, in decreasing order of impact:

1. regulatory control;
2. legal/compensation/information purposes;
3. dose trends and statistical analysis; and
4. epidemiology.

Quite often requirements concerning the handling of dose records will differ depending upon the application and may actually conflict. For example, in the case of regulatory control or radiation dose control, the approach tends to be conservative: a dose which is most likely non-personal, but cannot be confirmed as such, will be retained on an individual's record. If these data are used for epidemiological purposes, such doses will lead to an underestimation of the risk. For legal or compensatory purposes, it may be necessary to keep fragmented records separate. However, for epidemiological studies, where there is a reasonable likelihood that these records are for the same individual, they should be brought together. The existence of a large number of fragmented records can reduce the SMRs.

The obvious solution is to create different sets of records for different purposes so that there is one set of records for regulatory purposes and another for epidemiological purposes. In practice this is expensive, and having both an 'official' set of records and an 'unofficial' set of records can undermine the credibility of the registry.

One solution is to maintain the records in their original format and create a file that can be used to combine the fragmented records and annotate the questionable personal doses for epidemiological purposes. This process requires fewer resources and has the advantage that there is still only one official record.

OVERALL DESIGN CONSIDERATIONS

The following are overall design considerations that should be taken into account in the development of a national registry.

Cost

In order to provide a long-term database for epidemiological studies and dose trend analysis, it is important that the registry be able to survive periods of resource constraints. Operating costs and the number of staff should therefore be kept at a minimum. In addition, it is important to use upwardly compatible computer hardware and software so that when changes are made, it is not necessary to reformat or restructure either the data or programs.

Availability and timeliness of data

The registry must be able to provide, in a timely manner, information to its clients. These clients may include regulatory authorities, Workers' Compensation Boards, radiation workers, and the general public, all of whom will expect data to be available with a minimum of delay.

The software used must therefore readily extract data and produce a variety of different, professional-looking reports, which may be transmitted to the client by paper, fax, or machine-readable data on diskettes.

Confidentiality and data security

Confidentiality of the data in the registry should be high priority. Without limits on confidentiality, organisations and individuals will be reluctant to provide data. This can increase costs in efforts to obtain the data and cause gaps in the available data.

Legal protection of the data should protect the privacy of the individual and limit the administrative uses of the data to those that have been defined and publicly stated.

The registry should have stated policies on access and should have measures in place to physically restrict access to the data. This latter can be accomplished by use of encryption, passwords, and restricted direct access. Data within the computer should be regularly backed up and a disaster plan should be prepared to ensure that the data are not irretrievably lost in case of destruction of the computer.

DESIGN CRITERIA

The design criteria listed below are intended to address the overall design considerations outlined in the previous section.

Flexibility

The registry data structure and programs must be flexible enough to accommodate changes in monitoring devices, dose models, regulations, and other hazards in the workplace, such as chemicals or diesel fumes.

For this reason, programming should be well documented and modular in structure so that areas of the program requiring changes can easily be identified.

The use of codes that link to look-up tables can provide flexibility for items that may change with time. Parameters can then easily be added or changed and input data can be validated against the look-up tables. Typical look-up tables are listed in Table 1. They include: dose record types to distinguish between the different kinds of radiation records; birth place codes to identify countries of birth; monitoring frequency to distinguish different monitoring frequencies currently in use, as well as those used in the past.

Automated data processing

Much data validation, editing and processing can be automated with the computer checking. Bulk dose records can be entered in batch mode and erroneous records written to a suspense file. Printouts of these records can be used for follow-up and corrections made to the suspense file. Records that have been corrected can be automatically retrieved from the suspense file and reprocessed.

Even with the use of a national identifier such as the Social Insurance Number or National Health Number, it is still necessary to be able to merge records. For example, if temporary identification numbers are given to visitors or immigrants and a permanent number granted later, it will be necessary to merge the records together. If no national identification number is used, it would be prudent to have an automated search for 'new' registrants to ensure that they are not already in the database. A merging facility is essential to ensure that records can be efficiently merged when fragmentations are identified.

Table 1. Examples of look-up tables

Radiation type
Record type
Dose units
Organ codes
Raw data units/conversion factors
Measurement type
Birthplace (country) codes
Geographical codes
Employer type
Regulatory authority code

Candidates for merging can be selected according to preset criteria and presented on a screen. After the operator has confirmed that the merge should proceed, the computer can then combine the dose records and update the personal information.

If complete identification information is included on each submitted record, then updates to the information in the registry can be made as required. For new individuals, the computer can automatically enter the identification information. In cases where the record already exists in the database, there is an opportunity to compare the incoming information with the existing. Printouts of conflicts can be generated for follow-up.

Automatic surname updating by the computer can save considerable time and can be used for surname changes arising from spelling variations or from legal changes due to marriage, adoption, etc. If the other identifying information such as given names and date of birth match, the computer can automatically update the surname and produce a printout for review. In most cases, the updates will be correct and can be readily confirmed from the printout. The original surname can be stored in a previous surname file for future reference.

Reporting

Extraction and reporting of the data must be accomplished in a timely manner. A fourth generation language (4GL) can be useful because it provides the rapid extraction of data and the production of professional-looking reports with a minimum amount of code. However, the 4GLs do require more CPU time and space. Furthermore, since the manipulation of data for complex reporting can be difficult using 4GLs, it may be necessary to continue to use a compatible third generation language for the more complex data manipulations.

A number of routine reports, which should be an integral part of any system, are listed in Table 2. Daily operating statistics should be produced showing the number of records processed, the size of the files, and the file capacities to allow the monitoring of the data input and of the growth of the files.

Daily error reports should be produced for invalid data and characters in the input data. Listings of individuals who are not in the database and have missing identification information should be produced in a format that allows direct mailing to the organisations for follow-up.

When an individual's dose exceeds the regulatory limits, a notification letter identifying the individual, the dose limit

Table 2. Reports

Routine (daily)
Operating statistics
*High exposure notification
Invalid data in record
Surname automatically added
Identifier non-match
*Missing ID info (for new registrants)
Request (paper/screen)
*Dose history
Company personnel report
Ad hoc reports

* Formatted for direct mailing

exceeded and the appropriate regulatory authorities should be produced in a format ready for signature and mailing.

In addition to the above routine reports, a number of standard reports should be available upon request. These include a radiation dose history summary and a company personnel report listing monitored workers at that organisation.

A number of specialised programs will be required to respond to requests from the regulatory agencies. These could include routine summaries of individuals and organisations that have doses exceeding specified limits in a given time period.

An *ad hoc* query screen can be used to create and access specialised programs retained in an *ad hoc* library so that over a period of time, programs developed in response to specific requests and applications can be retained for future use.

Search capabilities

Powerful search capabilities are necessary for identifying fragmented records and the processing of current records. Searches can also help locate dose records when requests for information do not always have complete identifying information.

The system should be able to search by surname, and a combination of given names, sex and date of birth. The use of different combinations of identifiers can be used to narrow down or widen the search. Searches of a more specific nature can be readily done by means of special programs using the 4GL.

Compatibility

The layout for data submitted in a machine-readable format from other organisations should be standardised. Documentation specifying the standardised format and outlining administrative procedures should be available.

Compatibility of hardware, software, and data layouts is necessary if data are to be received from other organisations, otherwise considerable time can be wasted in trying to process the data.

Some attention should be given to ensure that the hardware and software used are compatible with early records, otherwise the reformatting of the early records into compatible formats may be costly.

DATA MODEL AND DESIGN

There are basically three kinds of information that are stored in a national dose record system. These are: individual information; dose information; and employer's information.

The relationship between these three different types of data is illustrated in Figure 1. There is a one-to-many relationship between individual information and dose information, and between the organisational information and dose information. The relationship between the individual information and organisational information is many-to-many. These relationships must be taken into account in establishing the data flow in the system.

The three kinds of information are stored as keyed files or databases and are linked together by common keys. The individual information and dose information can be linked by the individual's identifying number (a national identity number). The dose information can be linked to the organisational information by a numerical organisational code. Additional keys can also be used to allow rapid searching on a variety of items. These keys should be reviewed carefully in the early stages of the design, since it may be difficult to create new keys once the system is in place.

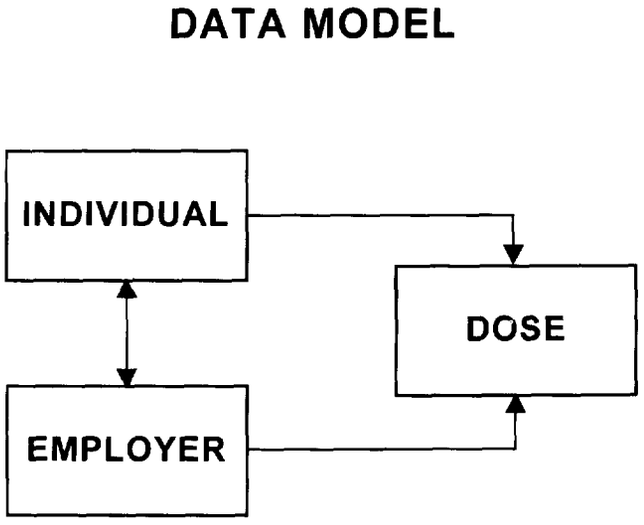

Figure 1. Data model.

The information included in the three data sets is summarised in Table 3. This table also indicates validation checks, which can be performed and the appropriate action recommended if the check should fail. For serious errors the records should be written to a suspense file for future follow-up and correction. In the case of other errors, the records can be processed and a printout produced showing the error.

The national identity number or surname are sufficiently important that records missing these items should be rejected to the suspense file. Records can also be rejected if codes cannot be validated against look-up tables. Further checks can be made to ensure that the incoming data matches or is consistent with data already existing in the database. For example, incoming given names or dates of birth can be checked against existing data for mismatches.

Figure 2 shows the physical data flow model used by the NDR for routinely updating information in the database. Updates are written to a holding file LDHSRANS, which is

Table 3. Dataset summary

Variable	Check	Action
Numeric ID	Arithmetic	Suspense
Surname	Alpha/present	Suspense
	Match with previous	Print/update
Given names	Alpha/present	Suspense
Sex	M,F,	Print/no update
	Present	Print/update
Birth date	Valid day, month, yr	Suspense
Birthplace	Look-up table	Suspense
Monitoring frequency	Look-up table	Suspense
Radiation type	Look-up table	Suspense
Dose	Numeric/range	Suspense
Geographical code	Look-up table	Suspense
Employer's ID number	Look-up table	Suspense
Employer type	Look-up table	Suspense

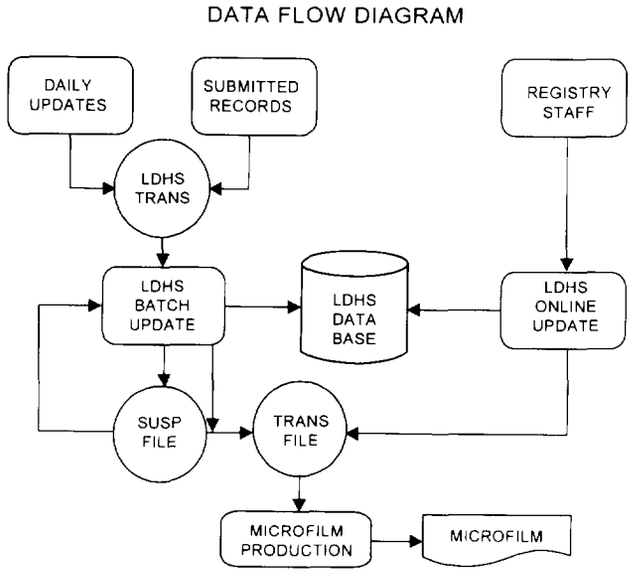

Figure 2. Data flow diagram.

processed daily. Records for LDSTRANS are checked for errors and any records with errors are written to a suspense file and an appropriate printout is produced for follow-up. Records are edited in the suspense file by a screen and the suspense file is automatically processed with the regular routine run the following day. Records that do not contain errors are written to the LDHS database and a copy of the updating record is written to the HISTRANS file. This file contains on-line transaction records that can be periodically removed and placed on microfilm.

In addition to the daily batch updating, records can also be edited and corrected on-line. All on-line transactions are written to audit trails that contain the original data, the modified data, the time and date of the update, and the name of the clerk who performed the update.

The on-line updating can include adding and deleting doses and updating individual identification information. A special screen that allows the changing of any fields without edits or checks is useful for cases where changes to records cannot be made through the regular screens because of the built-in edit checks.

HARDWARE REQUIREMENTS

The registry described in this paper requires an interactive computer system with random access. Five data access or entry terminals with slave printers are used. A laser printer is used for printing the follow-up letters and forms.

Table 4. Dataset capacities

Dataset	Number of records	Space (mb)
Individual	486,424	56.8
Alt — surname	60,054	5.3
Alt — ID — Number	413,776	18.8
Annual dose records	2,686,524	305.0
Dose trans	546,420	186.6
Employer address	23,440	15.5
Tables, programs, etc		86.7
Total capacity		682.1

The current file sizes and estimated space requirements are listed in Table 4. These figures are based on the current National Dose Registry, and can be used to get rough estimates of sizes. It can be seen that most of the space is required for the annual dose summary records, although the dose transaction records, if kept permanently on the system, would take up a large amount of space over time. The current National Dose Registry is growing at the rate of approximately 20 megabytes (mb) per year.

FUTURE DIRECTIONS

With the continuing decrease in the cost of storage, different approaches to the design of registries may be possible in the future. For example, optical disk storage of machine-readable data offers very large storage capacities at reasonable cost. It may be feasible to keep only the basic discrete input records on optical disk and to retrieve the necessary records and calculate the doses when required. Such a system would consist of an optical reader capable of reading automatically a number of optical disks, and a work-station served by a smaller local area network of work-stations for input and updating the information. This system would be in contrast to the design presented here where the main database consists of annual summary records that are supplemented by the discrete transaction records stored eventually on microfilm.

1. Ashmore JP, Grogan D. The National Dose Registry of Canada. *Rad Prot Dosim* (2) 1985, **11**, 95–100.
2. Ashmore JP, Davies BD. The National Dose Registry: a centralized record keeping system for radiation workers in Canada. In *Applications of Computer Technology to Radiation Protection.* IAEA-SR-136/58, Ljubljana, J. Stefan Institute, 1989, 505–520.
3. Darby SC (ed). Protocol for the National Registry for Radiation Workers. NRPB-R116, Chilton, 1981.
4. Newcombe HB. The design and future uses of national dose registers for regulatory control and epidemiology. *Health Phys* 1980, **39**, 783–796.