Chapter 9

Data Management and Security

Table of Contents

9.1 Introduction

This chapter is designed to provide basic guidance on the mechanical and administrative aspects of establishing and operating a birth defects surveillance program, covering a range of topics related to the development of an efficient, effective, and secure program. This chapter is intended to serve as a guide to planning the development of a new surveillance program, as well as to the review of practices and procedures in place within existing programs.

Issues covered in this chapter include computer hardware and software, data capture procedures, transmission of data, data file management, personnel management, physical aspects of the surveillance program office and, finally, data confidentiality and security considerations. We discuss the functionality of a data processing system in Section 9.2, followed by more detailed discussions of hardware and software in Sections 9.3 and 9.4, respectively. Data management is introduced in Section 9.5 on process standards, while the specific topics of data entry (Section 9.6), record linkage (Section 9.7), and record consolidation (Section 9.8) are described in further detail in subsequent sections. The importance of ongoing communication with data sources to identify and correct emerging problems is discussed in Section 9.9, and in Section 9.10 we address physical security and confidentiality issues.

9.2 Functional Data Processing System Features

The functionality of the system for processing and managing surveillance data must be able to support all the necessary processes and activities required by a program. This chapter is intended to develop the key considerations and capabilities that are applicable to surveillance operations. There is a wide variation in case volume, approach to data collection, budgets, and goals of various birth defects surveillance programs. Overall mission, size, and scope will determine the best combination of procedures and features for a given program.

The basic operations of a birth defects surveillance system can be accomplished using minimal computer hardware, software, and systems. The characteristics outlined below provide a broad scope of useful features and capabilities.

Computerized data collection. As case reports on birth defects cases are received, data should be captured within an electronic database designed to maximize a program's ability to manage the surveillance system and utilize the resulting data.

A program's capacity to receive computerized data from reporting facilities and other sources can ease the burden of case reporting and reduce or eliminate the need for recapturing already automated data. Improving the efficiency of data collection can minimize effort in the reporting facility and at the surveillance program, while reducing errors often due to "re-automation." Increased efficiency can also improve relations with reporting facilities and support compliance with reporting requirements. Below we present critical considerations related to accepting and processing electronic reports from reporting facilities.

- *Reporting case data within electronic files, rather than paper reports,* requires the exchange of detailed information on data submission requirements and on the characteristics of the files provided. Design issues, such as file formats and structures, and coding schemes must be understood to ensure accurate data exchange.
- *Any limitations in the reporting facility's computer database must be identified* to ensure that submitted data can meet programmatic needs. Any shortcomings or incompatibilities between the facility's system and reporting requirements must be recognized and addressed. Examples of such concerns include facility data systems that are missing standard (required) data items or that code a given item using coding rules that are not entirely compatible with the program's coding schemes.
- *As the data systems used to generate the data files are revised, any effects that such changes may have on the submitted data should be identified.* Information of this type must be communicated by the reporting facility to the surveillance system.
- *Submitted data must be reviewed for quality control.* This review should compare the data submitted with source documents or files to validate that the data are being represented faithfully within the surveillance database. The review should identify any data distortions caused by differences in processing systems, coding structures or rules, and conversion routines used to build the extract file or to import the data into the surveillance program.

Transmission of electronic data and data telecommunication. Below we present considerations related to transmission of electronic data.

- If case reports are accepted as electronic files, a standard format, file structure, code structure and medium for submission, i.e., tape or disk, must be developed and documented.
- Programs receiving passive case reports may elect to accept data in formats and code structures that follow the reporting facility's database structure and rules. In this case the facility – whether a hospital, a diagnostic laboratory, or other facility – must be expected to refer to the standard for submissions and provide the data in a format and code structure that is compatible with and convertible into the standard format for reporting cases to the surveillance system.
- Whether converted or nonstandard data files are supplied, each facility must identify any compatibility/consistency problems apparent between the source system and the standard.
- Secure methods for delivery of forms and data files need to be recommended by the surveillance program and followed by each facility.

Various modes for data entry. Below we present considerations related to data entry modes.

- To automate the information received in the form of paper case abstracts, a data capture mechanism is required. Approaches to accomplish this task include:
  - Classic data entry by keying data into a fixed format file
  - Optical scanning
  - Data capture through the use of custom screens
- Principles associated with the data entry process, which will ultimately enhance efficiency and data quality, include:
  - Standardizing data review, query, and preparation procedures
  - Verifying the keyed data
  - Editing at the point of entry
  - Editing of the completed input file
- In addition to capturing case data internally, consideration should be given to developing software that enable facilities to report case data electronically. Providing such software to the reporting facilities enables them to use standard file formats and coding and editing procedures for the data they submit. Editing the data at the point of entry, in particular, can reduce the need for later follow-back.

Receipt and integration of new case data while retaining data on reporting activities. Below we present considerations related to the integration of new case data.

- To the degree practical, the program needs to be able to receive submitted data for processing in the form of electronic files, paper submissions, and, potentially, through web-based or other direct data entry across secure connections.
- Integration of data into the system should be done such that integrity of the individual reports received is maintained. This allows the surveillance system to document the data source properly and to monitor reporting quality.

Ability to link surveillance data to new reports and files from other data sources. Below we present considerations related to linkage of surveillance data to new reports and files from other data sources.

- The program must have appropriate software and system computing capacity to screen incoming data against existing surveillance data in order to identify accurately duplications in reporting.
- The program must have the ability to implement a variety of strategies to link surveillance data with data from other sources in a manner that allows cases within other data sets to be identified as important. Through this means the program can augment the surveillance data base with information on cases identifiable through other data systems, as well as acquiring new case reports. This capability is essential if the surveillance program is to facilitate research studies.

Ability to modify system easily and inexpensively. The system used must provide flexibility with respect to systems modifications, edit specifications, and other data handling processes, as well as permitting modifications of code structure and data set variables within the database. When possible, programs should avoid developing the database using software systems that require considerable time and expertise to modify. It is best if modifications are controlled and can be made by birth defects surveillance staff.

Ability to handle updates. The system must provide an easy way to update registry data as new information on cases is received. Updates may include information on additional hospitalizations, further diagnostic work, or corrections to earlier reports.

Editing data. The system must support data editing at various steps throughout the data collection process. Data editing should be carried out as data are collected, processed, and incorporated into the program's database. Conducting edit checks as early as practical in the data collection process is an efficient way of improving data quality. Key stages for data editing include:

- At the point of data abstraction
- During data entry
- As files of new report data are prepared
- As case files are updated with new cases
- As additional data on known cases are added

Common edit procedures include field code range checks, table look-up of diagnostic and other codes, inter-field consistency checks, and editing across records for individual cases.

Report preparation. Staff must have the computer capability and training to conduct statistical analyses; to interpret the resulting statistical information; and to prepare text, tables, and graphics in the form of reports. Examples of statistical analyses include establishing basic case counts and rates, developing summary data on treatment information, reviewing prevalence trends, adjusting rates, calculating variance components and standard errors, and developing measures of the observed and expected prevalence of specific conditions.

On-line case queries. The database should be readily accessible to staff using different types of information to identify specific cases.

Easy maintenance of reference tables/files. The various reference files used to process and edit incoming reports and to develop statistical data on those reports must be easy to maintain and update. Such reference files may include tables of diagnostic and procedures codes, code groupings, geographic code dictionaries, hospital and laboratory code dictionaries, among others.

Extracting files/subsets. The capacity should exist to generate readily subset files of the surveillance data. This capacity should allow inclusion of data on cases selected using a variety of criteria and inclusion of specific variables for selected cases. File subsets often are needed for statistical analyses, quality control work, field site visits, and other uses.

Quality control information on data sources, amounts, and quality. To monitor case reporting timeliness and quality, the system will need to store sufficient information to support calculations of reporting timeliness and other data reporting quality measures. The system must allow for assessment of reporting quality overall and by reporting source.

Systems security, administration, and backup. The system must include features to protect data and programs from loss due to systems failure or user error and to maintain the confidentiality of patient and provider data. The computer system must provide a secure environment with security features designed and enabled to protect data from inappropriate access. Such measures must include a system of user name and passwords, along with a system to control the access of users to the computer server and drive locations where data are stored. These must be updated promptly with staff changes.

Program staff must be able to control or oversee these system administration activities. The system must also provide redundant back-up procedures to protect against system failure. This should include back-up and recovery procedures with regular and reliable copying of existing surveillance data and systems to tape or disk.

Archiving of data and systems. The surveillance data and the systems used to develop and maintain the data must be archived according to a predetermined schedule to protect against catastrophic loss. Archiving procedures must ensure appropriate preservation of submitted case abstracts and the routines used to process abstracts and to analyze case information. Archiving should also encompass statistical analyses, special studies, and the procedures used in those studies.

Cost effectiveness. The computer system and hardware used must be selected to fit both the needs and the budget of the surveillance program. The initial cost of the system and the cost to maintain and support both the operating computer system and specific programming requirements are critical considerations in selecting an appropriate system.

Adequate performance. The system selected must be responsive and provide adequate computing speed, disk storage, and working memory space to address the needs of the surveillance program.

9.3 Hardware

Computer hardware. Individual work stations and overall processing platforms should be selected to handle the work of the surveillance program and allow simultaneous on-line use of the data by multiple users and the various software packages used by the staff. Systems speed, number of concurrent users, active memory, disk capacity, robustness, and compatibility are all important considerations.

Systems back-up hardware matched to size of system. Systems back-up strategies must be complemented by hardware of sufficient size and speed to generate systems back-up on a prescribed schedule without eroding systems performance.

Printers, printing capacity, and quality/variety. Hardware that permits printing in the volume required by the program and that will produce high-quality printed tables, charts, and reports is important. Printing capabilities may be required for high-volume printing of envelopes or other specialized printing. If a surveillance program has a large-scale follow-back, the ability to print in-house materials and mailings that carry names and addresses significantly enhances security of the information at relatively low cost.

Graphics and slide production capabilities. Hardware that can be used to develop Microsoft PowerPoint presentations or slides is important. Slide makers and LCD projectors should be available to surveillance program staff.

Communications hardware and links. The staff of the birth defects surveillance program must be able to send and receive e-mail and to access the Internet. Data collection through hospitals and use of data by staff must also be supported by appropriate computer communication systems.

Strategy for planned obsolescence. The hardware used by the surveillance program must be able to operate software and systems that are actively supported by the software or systems suppliers. Planning for replacement of existing hardware should be an ongoing process. This should ensure keeping pace with changing software and systems requirements, enabling staff to manage the surveillance database effectively.

9.4 Software

The basic software selected to run the surveillance database must provide the features required to meet programmatic needs. It must have the capacity and robustness to conduct required procedures, be compatible with other similar data management systems, and be supportable.

Data analysis software. Standard statistical software should be available to analyze surveillance data. Statistical packages must have a full range of capabilities for developing standard statistical tables (including counts and rates) and conducting more complex analyses (such as standard error calculations or observed-to-expected ratio estimates). In addition, the software must support the design of tables, as well as presentation features such as titles, footnotes, graphics and, potentially, mapping. (See also Chapter 8 on Statistical Methods).

Record locking and file locking. The data management system must provide for data security and confidentiality as well. Systems should be considered with confidentiality and security features that enhance the proper protection of data. Depending upon the types of direct database access various system users are permitted, the data management software may need to control data access at the level of the file, the record, and the individual variables. This may require various levels of file access, which could be handled by using data management software with these capabilities.

File security software. Software that can regulate access to file servers and to specific computer drives or computer files, and maintain various levels of file access rights, is essential for storage of data on a Local Area Network (LAN)-based or mainframe system. Staff should either manage the administrative features within this software, or these activities should be under their direct supervision.

Multi-user capability. Software used to access and manipulate the data may need to have multi-user capabilities, allowing access by multiple users during most, if not all, file management routines. The need for this capability will depend on staff size and the scope of the surveillance program.

Integrated and stand-alone utility programs. The database must be accessible to program staff. Software and skills needed to develop ad hoc and specialty software routines must be available, as necessary, to manage and maintain files or to conduct specialized analyses. Such custom routines may be required on an ad hoc or a routine basis.

Record linkage software. Software is needed that supports data linkage. This capability is essential to de-duplicate new report data and to link cases to corollary files, such as birth or death files. Record linkage capabilities are also essential to the conduct of cohort studies that can link cases to files of study subjects. The ability to link data, with a high degree of accuracy, is critical to data quality, to conducting basic surveillance functions, and to research. Some states have developed their own custom – designed programs to meet record linkage requirements (e.g., Colorado).

Linkage capability can take several forms, ranging from on-line case-by-case queries to electronic comparisons of large databases. Data can be linked electronically through either pre-programmed routines or ad hoc routines and can be based on deterministic or probabilistic linkage procedures. The specific strategy and approach used by a surveillance program will depend on its size, overall mission, and resources available.

- ***Deterministic record linkage*** procedures, which involves the literal comparison of fields or columns within fields for exact matches, can be developed relatively easily and can be supported by most database management software. If this approach is used, it is essential to audit and refine the procedure painstakingly to ensure a high degree of matching accuracy.

- ***Probabilistic linkage*** bases record linkage decisions on determined probabilities that two records are likely matches. This technique generally is accepted as quite reliable when applied appropriately. However, it is dependent upon costly proprietary software packages that may not interface well with other data systems used by the surveillance program.

Regardless of the approach used, the results obtained through record linkage must be reviewed periodically for quality. The presence of unidentified duplicates within the case data and combining report data for different children into a single record are two obvious hazards of improper linkage. These false positive and false negative rates must be minimized by reviewing linkage quality regularly.

9.5 Process Standards

In this section we discuss the following aspects of process standards: inputs into the surveillance system (Section 9.5.1), instructions for reporting facilities on proper submission of data (9.5.2), procedures for initial review and query of submitted data (Section 9.5.3), procedures for receipt and logging of shipments (Section 9.5.4), and forms and batch control procedures (Section 9.5.5).

9.5.1 Inputs

Proper management of data within the surveillance program needs to begin through careful coordination with those providing the data and through following appropriate internal practices and procedures. These must be designed to promote accurate reporting and complete processing, ensuring a trackable system where processed data can be re-traced back through to the data originally submitted.

The procedures developed need to accommodate the various forms in which data are reported and the sources from which the data are derived. Data coming in to a surveillance program can vary widely with respect to the way they are transmitted and their content. This is true across states and within a state. The data can be provided in the following ways:

- Paper/electronic abstracts for reportable conditions
- Hospital discharge data
- Medicaid data
- Early intervention program data
- Data on services to children with special needs
- Birth and death record data
- Medical examiners reports

9.5.2 Instructions on Proper Submission of Data

Clear and concise instructions must be developed and distributed to all those involved in reporting cases to the surveillance program. Necessary components of these instructions include:

- Precise definitions of what constitutes a reportable condition/case
- Item-by-item explanation of information to be reported
- Timelines for reporting
- Acceptable reporting methods – paper, electronic, File Transfer Protocol (FTP)
- How and where to ship reports
- Procedures recommended to ensure secure shipment
- Procedures for handling corrections and updates to previous reports
- Sample of abstract
- Detail on electronic submissions, if appropriate
- Definition of terms, as appropriate

- Name of a contact person in case of questions

These instructions should be readily available to those with a 'need to know', should be prepared to minimize any anticipated potential misunderstandings, and should be updated routinely. Instructions need to be customized and targeted to specific data sources – such as laboratories, hospitals, physicians, or medical examiners – to reflect differences in what is expected from each.

Complete documentation of the receipt and preparation of case data from internal sources is also required. Examples include information from data systems for programs that provide specific information on children with reportable conditions. The nature of the data system and the schedule for providing or obtaining data, the format and technical specifications of the data all need to be documented. This is necessary to ensure coordination between the birth defects surveillance program and other data systems and/or sources.

9.5.3 Initial Review and Query

As reports are received, and during the intake process, it is important to establish procedures to screen incoming reports and data. These screens should be designed to avoid unnecessary work and to identify and resolve quickly any gross problems with the submission. These screening activities might result in a submission being returned and not processed. Other types of screening may occur at various points within intake processing of reports.

For paper reports, potential screens could include very basic things, such as examining the mailing for physical damage, proper addressing or, perhaps, tampering. Paper reports could be screened, as they are inventoried, to be sure they are completed adequately and that the case is truly reportable. If the paper report is primarily a case-finding tool, it might first be screened against cases in the database to determine whether it is new.

Pre-screening of data submissions should include checking for possible computer viruses, determining whether the file is readable and in an appropriate format and file structure, and establishing whether the count of records within the file is correct.

9.5.4 Receipt and Logging of Shipments

As data are received in the form of paper reports or automated files, forms- and data-control procedures need to be followed. Procedures should be designed to ensure that all data submissions are processed. They must also provide a mechanism for rechecking the status of the surveillance program database to validate that all information has been processed properly and appropriately. This can be accomplished by developing a log to record receipt and processing of reports by facility. Such a log could contain basic information about each submission. This might include: date received, reporting facility, number of reports, date span for the reports, format of the reports, date prepared, report numbers assigned to the batch, file or batch name assigned to the data, and the date data processing was completed. Maintenance of a log serves as a control point for the data. It can also be designed to permit monitoring of the reporting status of individual facilities.

Depending upon the type of report being processed, other approaches may be appropriate to ensure completeness. For example, data that serve primarily for case finding may need to be screened first against the surveillance program's database to determine whether the report is for a new case, thereby permitting previously reported cases to be quickly dropped from the case-finding data.

9.5.5 Forms and Batch Control Procedures

For data control and tracking, it is important to use a systematic procedure to inventory and to identify unique reports. Classically this is accomplished for paper reports through use of a series of sequential numbers, with each form assigned a unique number. Forms are then organized sequentially into batches of manageable size. A similar procedure can be used for automated data submissions to assign each electronic record a unique identifier and to maintain a record of the file name assigned to each batch of reports submitted.

These procedures allow staff to locate a specific report easily and provide a mechanism for data inventory. Missing report numbers can be listed and resolved as report file completeness is evaluated. Data edits for each report can reference these numbers to identify and resolve any concerns with the data in that report.

9.6 Data Entry

The most basic aspect of developing a surveillance database is preparing an electronic file of reported information. While the proportion of paper case reports received varies widely across surveillance programs, each program must have reliable mechanisms for entering data from manual reports into an electronic file. In addition, many programs can provide reporting sources with software that can be used for submitting cases. Applying some simple concepts to the automation of information can help provide data files of consistent quality.

Interactive edits. Developing a process for capturing data destined to reach the surveillance provides an opportunity to build functional editing of entries into the operating procedures. This is especially effective when edits are used to question incoming data at the point where the patient's chart is available for review. Editing data at the point of origin is the most efficient method to ensure high quality.

Interactive edits can be very simple checks, such as ensuring that only numbers are entered into a numeric field, ensuring a date entered is a valid date, or preventing a required field from being left blank. More complex edits might involve providing links to a database of valid codes for diagnoses or procedures, editing for consistency across fields, or screening each case to determine if the child was reported previously.

In designing and developing interactive editing procedures, it is important that the objectives be kept in focus. A process is needed for producing high-quality electronic data efficiently and effectively. Interactive editing needs to be functional. It should be designed to screen for impossible or improbable entries. It must also be efficient, providing the operator with a clear explanation of the perceived problem and a ready mechanism for resolving the issue.

Verification. Verification procedures are another tool for controlling the quality of incoming data. The key data processing steps of information coding and the actual process of data entry are candidates for verification. Verifying data is an old and time-tested method of monitoring and controlling errors introduced into data through data processing procedures. These practices do not improve the quality of the reported data, but they do minimize degradation in data quality during data processing. The purest example of data entry verification is blindly re-keying previously entered data using software that compares the newly keyed data, key stroke by key stroke, to that entered earlier. Any discrepancies are identified and resolved by the verifier.

To verify data is to double check the data to ensure it is captured accurately. Verification procedures can take two basic forms, namely, independent and dependent verification. There are also two basic strategies relative to the scope of verification: verifying each incoming case or verifying a sample of cases. In ***independent verification***, the verifier is not provided with the previous work and must essentially redo the work. The two versions are then compared and any discrepancies resolved. For ***dependent verification***, the verifier has access to the original work and reviews the entered data, comparing it to the source document; in the case of data entry, essentially proof reading the work.

Focusing on verification as a tool for efficiently developing data files of consistent quality, verification can be developed incorporating the quality and skill of the processing staff with efficient methods for screening and resolving processing errors. As an example, dependent verification of all the diagnostic coding done by new staff might be done by experienced staff and continued until a "qualified" level of accuracy is consistently demonstrated. Once the new staff member has qualified, only sample independent verification might be done.

Information obtained through verification can provide important insights into staff training needs. These results can also ensure a consistency of understanding and interpretation across staff involved in data preparation, highlighting any inconsistencies.

Forms/record and batch controls. Since data arrive in a variety of forms and from numerous sources throughout the year, effective methods to inventory all incoming data are important. As a corollary to logging the receipt of data shipments, control of individual records is very important.

As reports – both paper and electronic – are received and early on in their processing, a report number needs to be assigned to each report to serve as its unique identifier. This identifier provides a ready mechanism to inventory the incoming reports and, later, the consolidated files of processed reports. This report identifier also enhances coordination of the work during later stages of file editing and processing.

There are a variety of schemes for assigning a report identifier. The most basic is a sequential number that begins with the year the report was received followed by a simple sequential number. By including record type/source codes within the prefix for the sequential number, the type of report or information source can be incorporated into this identifier. Such information is often useful in developing management information regarding database status.

A system for numbering data entry work files needs to be developed and employed to properly control and inventory work batches. Each work batch needs to be assigned a unique batch identifier. A log should be established to record the report identifier numbers within each batch. The log should include the date completed, the individual completing the batch, the individual verifying the batch, and the date the batch was processed into the surveillance program. This information will aid in assuring all reports are processed and in tracking down any discrepancies. Information in the log will help assess processing issues, such as timeliness and staff accountability.

Procedures appropriate for a variety of data inputs. It is important to map out the proper handling and intermeshing of data from each data source carefully to ensure data quality. As mentioned earlier, sources of information can vary widely, both in type and quality of data. In designing the data entry process, the form in which the incoming data are presented can create a need for customized procedures.

Tailoring the procedures to match the data source and data format can add to efficiency and enhance final data quality. These adjustments might take many forms, including facilitating data entry through customized data entry screens for certain report types. Specialized editing to match the data source and, perhaps, to screen for code conversion errors may be required. Some data sources might be considered primarily as sources of case ascertainment. The first processing step might be screening cases against the program's database to determine if the case has been reported previously.

Training/certification and instruction for data preparation. Program staff members involved in data collection and processing must have the skills required to accomplish their work accurately. The skills required vary across key functional activities, namely abstracting case data, coding the information, and entering the data. Data management and editing routines will not correct data quality problems that occur if staff members are not properly trained.

Surveillance programs need to have a strategy for training new staff that allows them to learn the new job; measures their understanding of the work; provides feedback on problems and progress; and determines, in some objective way, that the new staff member's work has reached an acceptable level of quality. Staff skills and the rigor with which work is reviewed will vary among surveillance programs. Whether a surveillance program utilizes active or passive case ascertainment influences the skills needed. Hiring staff with training and experience in health information management may prove important. By the very nature of birth defects surveillance, there will always be a need to train new staff in a number of areas that are unique to the program and where it is not possible to hire experienced staff. There must be a strategy to ensure that staff assigned to a task have the skills the task requires.

As a component of continuous training, detailed manuals are needed that document procedures and serve as a reference source for staff. Staff should be encouraged to refer to these manuals and to identify errors, inconsistencies, and misinterpreted sections. Updating these guides periodically ensures that the manuals/instructions remain functional and current and able to serve as training guides for new staff.

Future editions of *The Surveillance Guidelines* will address training issues for surveillance programs in greater depth.

Input file processing functions. The management of data quality within electronic birth defects data files is important as well. The procedures and processes for handling the quality of processed electronic data are similar to those used for paper reports. The tools available to a birth defects surveillance program are basic data processing and management practices that are not unique to these types of data. Electronic data files readily lend themselves to editing and clean-up. Standard computer routines can be used to screen files for obvious errors or inconsistencies, to spot problems with the data efficiently, to summarize findings, and to organize results in ways that allow the efficient correction of any errors.

Key components of input batch processing are outlined to provide an inventory of the tools available for functional data quality control. The combination of practices employed by a given surveillance program needs to match the methods and procedures used for data file development.

Editing. Development of data editing procedures is a standard activity in any database development effort. As with interactive editing during data entry, computer routines can be developed to identify a variety of data problems. Standard edits often include:

- Field range checks
- Report number range checks to identify missing records
- Inter-item consistency checks
- Field validity checks
- Code validation through table look-up, i.e., diagnostic or procedures code tables
- Consistency across multiple reports for the same case
- Hard versus soft edits and use of edit flags

The organization of the results from edits requires the same care in design as do the edit criteria. The results of an edit run need to be organized to make error resolution and file correction as efficient as practical.

Tracking information. As potential problems with data are identified, it may be necessary to ask the reporting source for clarification or for additional information. A basic procedure to monitor outstanding requests for clarification or correction should be used.

Printed case abstract. In conjunction with efforts to correspond with staff at the data source about reports, a ready mechanism to print an abstract of a report can improve the effectiveness of communications and may enable correction of other errors in a report that cannot be identified by the edit routines. The capability to print a case summary easily can prove useful for multiple purposes.

Error correction. It is important to have effective and efficient procedures for error correction. Reports that identify edit exceptions can be linked to the edit results to pull up rapidly, or to ***queue***, the records needing attention. A well-designed process can minimize the potential for introduction of errors in the course of record correction.

Case-by-case and multi-record correction. Mechanisms to correct records one at a time are important. The capability to update multiple records simultaneously can also be useful. When used judiciously, multi-record correction can save time and reduce the potential for error.

Add/delete. The capability to delete spurious or redundant records can prove to be very useful.

Linkage and assignment of case identifier. As input files are processed and screened for duplication, a system for uniquely identifying each case is necessary. While a program may choose to number and retain all reports received, it is critical that a specific child's reports all have the same case identifier. This is necessary for record and file linkage. In a program where data are consolidated immediately, an identifier for each child is still a critical component of the system. A mechanism for assigning identifiers to newly reported cases is necessary. In the case of electronic submissions, this process should be automated.

Facility reports. Summaries of data quality relative to screening and editing of incoming reports is important for maintaining an accurate picture of the quality of submitted data. Summary reports that permit the tracking of report quality over time and across facilities can be designed. Such information is very useful in identifying facilities that are candidates for data quality reviews and/or in-service training. These reports can complement efforts to work with facility staff to correct any persistent problems.

9.7 Record Linkage

The proper operation of any birth defects surveillance program depends on developing and following procedures for efficient and effective record linkage. These procedures should be developed carefully. The accuracy of the procedures used to link individual case reports needs to be measured and monitored. Instances of the same child being in the database as different children and different children being presumed to be the same child must be estimated. Developing and monitoring linkage procedures carefully is as crucial for programs that manually search for potential matches as it is for those that use electronic linkage.

Not only is it important to link incoming reports accurately to the historic file to locate previously reported cases, but the ability to link to other databases is also essential. Procedures need to be tailored and evaluated specifically for each type of linkage. Goals for linkage completeness that reflect these expectations need to be established.

Linking birth defects case data with files from other sources may be done to meet a number of objectives. These include:

- Deleting duplicate data
- Case-finding
- Augmenting the information available for a case
- Conducting special studies or program evaluations

The level of precision and efficiency that can be expected from a matching process are functions of several factors. Key among these are:

- Quality of data within the files to be linked
- Number of fields common to both databases
- Logic employed to compare the files
- Time available to review and assess each link

The matching strategy developed should maximize the results and minimize the resources employed to obtain those results. Estimating the level of precision for any linkage procedure can be used to assess the advisability of revision. These estimates are also important for evaluating the suitability of using the linked data for specific purposes.

With respect to assembling the required data on each case, linkages to birth certificate files and death certificate files are extremely important. These sources can provide the surveillance program with valuable information on each case. For example, linkage to the birth certificate file has the added benefit of identifying reports for a single child that may not have been linked properly during the processing of incoming data. Linkage to birth and death records can also provide the ability to track changes to a child’s name over time. This can assist in collating data on a single child that might otherwise be treated as distinct cases. In some jurisdictions, access to this kind of information will depend on legally prescribed restrictions.

9.8 Record Consolidation

When multiple reports are received on the same case, differences can be expected in some of the information across reports. By developing a summary of the information on each case, consolidating the information across reports into a single summary, the information about the case can be enhanced. A number of issues must be addressed in any information consolidation effort. The categories of information that could be consolidated or summarized, and the key issues relative to summarization, include (1) demographics and identifiers and (2) diagnostics.

Demographics and identifiers. Most demographic items are constants and do not change with the age of the child. These include date of birth, race or ancestry, mother's age. Updating missing data fields using data from subsequent reports is generally appropriate. Conflicts across reports for these fields can be difficult to resolve, but may be predicated on the source of the data, prioritizing data from specific files or facilities. Changes can be expected to occur over time in identifier fields such as name, parents' names, and address. Selection of the appropriate data to be included in a summary needs to be based on the purpose of the summary. For example, data for referral or outreach efforts need to be current, while data for auditing birthing hospital records should represent information at birth.

Diagnostics. As multiple reports for a child are received, collecting diagnostic information across all reports can result in significant redundancy. When the same diagnoses are reported repeatedly, this redundancy is simple to manage. As diagnostic data change across reports, there are three possible causes. Each of these raises specific issues relative to proper management:

- *New conditions being diagnosed.* Newly diagnosed conditions clearly must be included in any summary for the child.
- *Previously diagnosed conditions reported with greater or lesser specificity.* Redundancies in diagnostic codes caused by differences in specificity can be problematic. In the absence of accepted guidelines for doing so, to eliminate redundancies (increasingly specific diagnostic codes) using intuitive logic can be problematic. The logic must be thought through clearly, with the intended use of the resulting summary in mind.
- *Actual changes to a previous diagnosis.* Changing a diagnosis can reflect a revision based on better information or an alternative diagnosis, for example, a difference of opinion. A key problem with such changes is the need to differentiate a changed diagnosis from a condition that has been newly diagnosed. Having an effective mechanism in place for facilities to report corrections to diagnostic information can help reduce confusion in interpreting subsequent reports. Making changes to diagnostic data should be done carefully and through close coordination with facility staff. Some programs may prefer not to change original information, but rather flag it as inaccurate or no longer valid. In this way, the integrity of the database remains intact: inaccurate information is not "counted", data quality evaluations can be conducted, and new or confirmed information is accepted. This is especially useful when comparing reported information with the results of medical records review by surveillance staff.

Procedures data. Redundancy can be expected in data reported for procedures, since it is not uncommon for a child to have some treatment or corrective procedures performed multiple times. Therefore, such data can be consolidated reliably only if the date each procedure was performed is reported, or available through medical record review, to identify specific procedures by date received.

9.9 Feedback to Data Sources/Abstractors

When data problems are identified during report processing, it is important to communicate those problems to staff at the data source. Facilities and/or individuals providing data should be interested in learning of problems the surveillance system encounters with respect to the completeness and accuracy of their data. With passive reporting systems, communicating errors, resolving inconsistencies, and reviewing apparent discrepancies represent effective feedback mechanisms. An efficient mechanism is needed to provide feedback on problems, although the necessary corrections to the information may seem obvious to surveillance program staff. Staff of the reporting facility will benefit from the feedback and may need to correct the information within their own records.

9.10 Security

Many of the data assembled by a birth defects surveillance program are extremely sensitive. For this reason, a program must initiate and maintain a comprehensive strategy for data security that ensures data are protected from improper access or inappropriate use. Developing a security plan that establishes and demonstrates a commitment to data protection is essential in reaching the program's long-term goals and objectives.

9.10.1 Personnel Issues

Four aspects of security management fall under the category of personnel issues.

- *Hiring practices.* Attention needs to be paid in selecting new staff members to screening candidates to ensure they can be relied upon to handle confidential data appropriately. A work history that includes responsibly handling confidential data is an example of desirable experience. It is important to request and check references for all prospective employees. If possible, security background checks should be conducted prior to making a hiring decision.
- *Written procedures on security and access.* New employees need to be informed clearly of the procedures regarding appropriate access to and use of data, particularly any files that include personal identifiers. Written materials that describe the nature of the data and the rules and policies relative to data handling must be reviewed with employees. These materials must cover all aspects of employees' actions for which they are accountable. These materials need to be discussed with each employee to ensure that the employee has every opportunity to ask questions so that they understand the policies explicitly. A written policy on the release of identifiable or potentially identifiable data must be included. It is essential that all aspects of data release be identified within the policy, along with who has the authority to authorize a release. Such a policy must include the "business activities" of returning data diskettes and corresponding on data editing problems with data providers, as well as release of identifiable data for research use or in conjunction with child-find referral activities.
- *Security and confidentiality agreement/oath.* Each employee who has access to the program's data must sign a confidentiality pledge. The pledge should be in the form of a comprehensive statement that outlines the confidentiality policy in broad terms. In addition, this document should include a statement that the employee understands the confidentiality policies and the potential consequences for violating these policies. Finally, the document must include an oath on the part of the employee that they will abide by these policies. As significant changes to the confidentiality policies are made, each employee must sign a new pledge that reflects the new policies.
- *Disciplinary policy.* Whenever there is an allegation of mishandling confidential data or where unauthorized access is suspected, the incident must be investigated. Such an investigation must be conducted carefully and in a manner consistent with existing employment laws and personnel practices. Appropriate disciplinary action must be taken if it is established that an employee has violated the confidentiality policy. Any deliberate violation of policy that results in the inappropriate release of confidential data should be grounds for dismissal and for potential criminal action, depending upon the law governing these data.

9.10.2 Transportation and Information Handling

Basic security concepts that should be considered relative to shipping and handling information are listed in this section. Standard office practices and procedures for handling materials that include confidential data need to be developed and followed. These are necessary to avoid problems due to inappropriate or inadvertent access to these data.

The privacy regulations developed as part of the Health Insurance Portability and Accountability Act, or HIPAA, place significant responsibilities on hospitals, physicians, and others to properly safeguard confidential data on their patients. These regulations place strict procedural standards on health care facilities, heightening concerns about patient privacy held by health care facilities and providers. It is important to adopt information exchange practices with data sources that do not create a potential liability under the provisions of HIPAA (see Chapter 2 on Legislation). For example, common and efficient methods for exchanging information, such as fax or e-mail, need to be avoided or used with great attention to appropriate security. This is because faxed images can be intercepted and printed, can be inadvertently sent to the wrong fax or to a fax that is unattended or otherwise not secure. E-mail shares all these problems in addition to the fact that e-mailed materials will become part of the e-mail back-up systems and copies of sensitive materials will become interspersed with other documents that may well be public information. The existence of these back-up files means a loss of control over the data and access to the data. The problem of potentially intercepting e-mail only compounds this problem.

With these thoughts in mind, key considerations relative to good data handling and transporting practices are provided below.

- *Instructions to data sources for addressing and shipping of incoming reports and information.* All facilities and individuals who ship abstracts or data to the surveillance program must be provided with current and precise address information. Data sources should be encouraged to ship data in a secure manner where chain of custody signatures are required, such as certified mail or FedEx. If a shipment is received that was misaddressed, the data source reporting the cases should be contacted promptly by telephone, with a follow-up letter, and be advised of the correct addressing of shipments. In addition, standard practice should include prompt acknowledgement of shipment receipt. As staff at the data source learn to expect an acknowledgement of each data shipment, failure to receive an acknowledgement will alert them to the possibility that a shipment has been lost or delayed.
- *Use of fax or e-mail for forwarding or receiving sensitive data is not advisable.* These methods of transmission should not be considered secure unless the sensitive information is encrypted and password protected. Fax machines that both send and receive such materials need to be attended during transmission.
- *Managing the work station.* Employees need to be trained to manage their desktops. Confidential materials should not be on a desk if they are not being used actively and should not be left unattended during breaks or lunch periods. Similarly, passwords should be required for staff access to any personal computer that holds confidential data or that allows access to confidential data through a network or other computer connection. Such equipment should not be left unattended with connections in place that would permit unauthorized access. Care must be used in displaying confidential data on the computer monitor in order to ensure that persons who do not have authorized access cannot read them. All materials must be filed properly and locked when not in use. Following these common sense practices reduces the possibility of inappropriate access and should be applied conscientiously to the desk, the personal computer, and the files assigned to each staff member.

- *Physical access to abstracts, other documents.* Limiting direct physical access to files and other materials that include confidential data is a basic step in reducing the likelihood of someone seeing information to which they are not privy. The design and layout of offices can be done in a way that enhances staff members' ability to carry out their responsibilities without exposing confidential material to others. Careful planning in this regard, combined with good desktop management practice, will minimize inadvertent access.
- *Procedures and furnishings to lock up documents and diskettes.* Employees must be provided with the office furnishings needed to adequately secure documents. Locking desks and locking file cabinets are essential, with thought given to assignment and management of keys for these locks and the organized storage of extra keys.
- *Procedures for shipping reports and information from the program.* Program staff need to follow secure practices, as when they send any confidential materials to data sources. It is essential that such shipments are addressed properly, and the address should be confirmed if there is any doubt about its correctness. Materials should be shipped using a method/carrier that obtains a signature to verify receipt. Confidential data should not be shipped through e-mail or FTP unless the security of the connection is ensured or an adequate encryption technique is used to disguise the data.
- *Shredding and destruction.* Considerable care must be taken to avoid any potential for disclosure of data when confidential material is discarded. Employees must be conscious of the need for care when discarding any program-related materials that include identifiers or that would be considered confidential. Computer listings, correspondence, and other materials need to be screened to be sure that confidential data are handled appropriately. Staff should be provided with access to a shredder, and paper abstracts or printouts with confidential data should be shredded promptly. Any large volumes of confidential materials that need to be disposed of must be destroyed in a secure way.

  Similar standard precautions must be established for computer storage devices. Diskettes should be reformatted, rather than simply deleting files. As hard drives on personal computers are replaced, the old drive must be reformatted or any data remnants otherwise destroyed, for example, by storing them between strong magnets for a period of time.
- *Transportation of data.* When staff members are in the field, all confidential data must be carefully safeguarded. Documents should be transported in locked brief cases or otherwise protected. The security of portable computers must be ensured. Confidential materials must be kept locked in the vehicle trunk while traveling. During overnight stays these materials should be removed from the vehicle and placed in a hotel room rather than left in a vehicle overnight.

9.10.3 Physical Security

Physical features of the worksite can enhance information security significantly. There are two specific ways the facility housing a surveillance program can maximize security:

- *Restrict physical access to the work area.* To the degree possible, access to the surveillance program's work area should be controlled. Ideally, it should be isolated with a card entry access system. Reducing or eliminating travel into and through the work area translates directly into reducing or eliminating opportunity for inappropriate data access.
- *After-hours security.* At a minimum, the office area must be locked securely after working hours. Ideally, the office area should be protected against unauthorized access through use of an alarm system that includes motion detectors and that is monitored centrally and continuously. If possible, no janitorial services should be carried on after hours.

Periodic maintenance work should generally not be conducted after hours unless surveillance staff are alerted and have an opportunity to take any and all extra precautions to ensure appropriate security of the data.

9.10.4 Computer Security

Proper data security requires a comprehensive approach to computer security. A number of key aspects to any plan designed to protect electronic data files are listed below.

- *User ID and password.* A system for unique user IDs and passwords is a cornerstone of computer network security. Staff should not be allowed to share ID and password information. Departing employees must be deleted from the system promptly. Periodic outdating and changing of passwords should be standard. Employees need to understand the importance of these activities and know that their personal login is critical to protect. This is because activities on the system will be traceable to the user's ID and password.
- *Virus scan – current.* In receiving electronic data, it is essential that diskettes and other electronic files be scanned for viruses prior to loading onto the personal computer or the network. A comprehensive and continuously updated virus scanning package should be used for this purpose.
- *Control of user access to data.* Careful management of user rights to the network or other computer system can significantly enhance data security. It is important to minimize to the extent possible unnecessary access to files. Steps must be taken to decrease or eliminate both potential misuse of data and inadvertent damage or destruction of data that are accessed inappropriately. Planning the architecture for data storage can complement limiting access to the various data files and greatly enhance security in the process. Much like user IDs and passwords, this level of security must be continuously maintained, with access modified as staff work assignments change over time.
- *Discarding of old personal computers, hard drives.* There are a number of special considerations regarding the security of electronic files. Simply deleting a file from a hard drive or diskette does not actually erase the data. This problem is not always properly addressed as old computer equipment is swapped out or discarded. There must be procedures developed for disposal of computer storage devices that ensure none of the data are recoverable.

9.10.5 Policy on Release of Data

Written procedures must be established that describe the proper mechanisms for release of information from the surveillance program. Written procedures are necessary to provide surveillance staff with a clear understanding of proper data handling and release. The process for obtaining approval for access to the data, and authorization for release of the information, must be described in detail. There must be no confusion among the staff on this critical topic.

Confidential data release procedures should include the specific practices required for proper preparation of tabular statistical data, as well as de-identified micro data files. These micro data files must be designed to guard against inadvertent disclosure of confidential data. The procedures must delineate clearly the approval process that governs and regulates release of identifiable or potentially identifiable data. Issues related to sending identifiable information to data sources and to other sources of information about cases of birth defects must be covered. Providing access to confidential information for research purposes must be discussed, describing the types of research projects that may gain access to these data and the system's process for reviewing and approving such projects. Finally, the conditions under which the information can be used for administrative purposes must be covered. This should include using the data to ensure that children and families are referred appropriately for needed services, if this is part of the surveillance system's objectives.