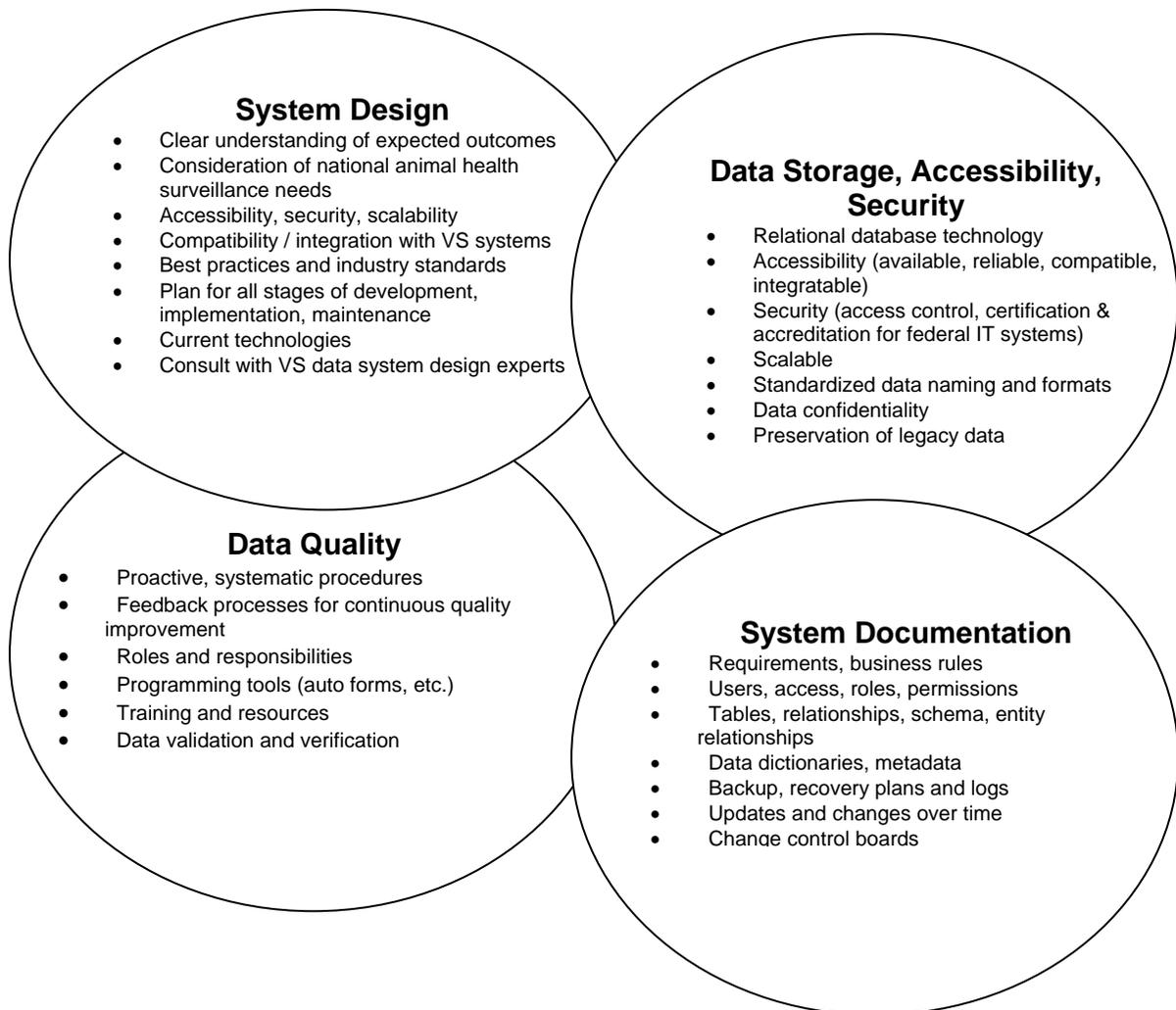The following is an excerpt (Chapter 3) from the *Surveillance and Data Standards for USDA/APHIS/Veterinary Services, version 1.0 (July 2006)* that was developed by the Centers for Epidemiology and Animal Health in July 2006. This Chapter focuses on standards for developing information management systems (database systems) to support Veterinary Services' (VS') surveillance programs.

Chapter 3 Standards for VS Animal Health Surveillance Database Systems

The success of an animal health surveillance system depends largely on the availability of high quality and accessible data that can be easily and readily used for meaningful analysis and reporting. Data with these characteristics result from expert design of secure, accessible database systems; systematic procedures that assure data quality; and clear and efficient communication among a diverse group of professionals. That group includes, among others, IT experts, animal health surveillance experts, laboratory scientists, statisticians, program managers, veterinarians, and animal health technicians.

This chapter is intended to provide a high-level description of the standards for animal health surveillance database systems. While Chapter 2 focused on the details of standardizing database fields through the use of uniform naming and formatting conventions, this chapter addresses broader topics of database design, data storage, data security, data confidentiality, data quality, and related issues. The chapter describes expectations for designing, implementing, and managing VS data systems developed for national animal health surveillance. These expectations are organized around four broad and interconnected concepts: (1) system design, (2) storage, accessibility, and security, (3) data quality, and (4) system documentation.

System Design
- Clear understanding of expected outcomes
- Consideration of national animal health surveillance needs
- Accessibility, security, scalability
- Compatibility / integration with VS systems
- Best practices and industry standards
- Plan for all stages of development, implementation, maintenance
- Current technologies
- Consult with VS data system design experts

Data Storage, Accessibility, Security
- Relational database technology
- Accessibility (available, reliable, compatible, integratable)
- Security (access control, certification & accreditation for federal IT systems)
- Scalable
- Standardized data naming and formats
- Data confidentiality
- Preservation of legacy data

Data Quality
- Proactive, systematic procedures
- Feedback processes for continuous quality improvement
- Roles and responsibilities
- Programming tools (auto forms, etc.)
- Training and resources
- Data validation and verification

System Documentation
- Requirements, business rules
- Users, access, roles, permissions
- Tables, relationships, schema, entity relationships
- Data dictionaries, metadata
- Backup, recovery plans and logs
- Updates and changes over time
- Change control boards

The chapter does not provide detailed technical information on these subjects; users interested in technical aspects should refer to the wealth of books and Web sites available on these topics, or they may consult the professional experts within VS. The chapter also does not address data system needs for particular end users such as epidemiologists or geospatial analysts. The idea is that if the systems development concepts are properly addressed, then the needs of specific end users will be satisfied.

Table. Summary of Standards For Data Systems

Data System Design Standards:

3.1 Data System Design

3.2 Data Storage and Transfer

3.3 Data Security and Backup

3.4 Security Certification and Accreditation for Federal IT Systems

3.5 Data Quality

3.6 Data System Documentation

3.7 Data Confidentiality

3.1 Data System Design

Standard: The data system for the surveillance program ensures that appropriate data are collected and stored to meet the objectives of the surveillance program.

  a. Integration and compatibility with other VS systems is carefully considered in the data system design.
  b. The data system is designed using industry standards and best practices for database system development.

Supporting Information:

A sound database design creates the foundation of the surveillance system's data collection efforts. Database design ensures that the necessary data will be available to support the objectives and expected outcomes of the surveillance system and ensures the data will be available in a useful format. Resources describing best practices and industry standards for system and database design may be found in many books and Web sites, and the details of this subject are beyond the scope of this document. However, these considerations for VS animal health surveillance database design are emphasized:

- Data systems for VS animal health surveillance programs should be developed in consultation with VS' Center for Animal Disease Information and Analysis (CADIA). CADIA provides leadership in the areas of information systems, technology for collecting data in non-office environments, managerial and epidemiologic reporting, risk analysis methodologies, and computer-based modeling of animal disease.

- Data system design must consider integration and compatibility with other VS data systems including, but not limited to, Emergency Management Response System (EMRS), Generic Database (GDB has been recently renamed Animal Health and Surveillance Management information system [AHSM]), National Animal Identification System (NAIS), National Premises Identification System, and the VS Atlas of Animal Health Information - Spatial Data Library. When integration with existing systems is not possible, justification for this deficiency should be documented.

- Current technologies designed to enhance data communication and exchange between systems should be used to facilitate accessibility, integration, and compatibility (e.g., XML, HL7 messaging, web services). This concept relates to standards 3.1 and 3.2.

- Data system design should consider accessibility, security, scalability, maintenance/support, risks and impacts. The system documentation should reflect these issues (see standard 3.5, system documentation)

- As described in Chapter 1 (standards 1.10 and 1.11), a clear understanding of the objectives of the surveillance system, expected products of the surveillance system, and how data will be analyzed and reported should be established before data modeling is initiated. Application and database development teams should work closely with the surveillance planners and managers to assure that the data system is aligned with goals and expected outcomes.

- Data system design includes standard design practices such as:
  - o Development and evaluation of conceptual and physical data models
  - o Development of schema and entity-relationship diagrams that define relationships and reflect data flow, events, and business rules
  - o Databases are normalized to reduce data redundancy and the associated pitfalls; ideally, databases are normalized to a modified third normal form.
  - o Restricted use of user-defined or miscellaneous fields or tables

- Data system design follows standard stages of the database or software development life cycle, including planning, analysis, design, construction, testing, implementation, and maintenance.

3.2 Data Storage, Accessibility, and Transfer

Standard: Data for the surveillance program are stored in an electronic relational database using standardized naming and formatting conventions for data fields. The database system allows access to different types of authorized users and allows for easy data transfer and integration with other systems.

   a. Electronic, relational databases are used to store data and are designed using the industry standards as described in standard 3.1.

   b. To facilitate the flow of information among systems, the database uses the standardized data naming and formatting conventions described in Chapter 2.

   c. The data are stored in a manner that facilitates the flow of information across systems.

   d. The database should be scalable to allow for appropriate expansion of the system and reliable for continuous use.

   e. The data storage system should maintain and make accessible legacy data for the program.

Supporting Information:

Relational databases. Electronic, relational databases provide the most accessible, reliable, scalable, and secure form for storing surveillance information. Spreadsheets are typically not appropriate for storing animal health surveillance system data. Relational databases provide these advantages:
- Tools for efficient and secure data storage, retrieval, and update
- Descriptions of the stored data items and their formats (i.e., metadata, data dictionaries)
- Tools for transaction and concurrency control that allow multiple users to access and use the database(s) simultaneously
- Tools for data recovery
- Tools for database security, including user authorization and backups.

Standardized data formatting. One of the challenges of database design and development is deciding what data to include (i.e., what fields or elements), and how to format the data

so it is compatible with other systems. Use of standardized naming and formatting of data fields greatly increases the ability to integrate information from multiple data collection systems. Chapter 2 addresses the naming and formatting of data elements.

Accessibility. Data and information accessibility is critical to a successful surveillance program. Animal health surveillance data must be accessible, in a timely manner, to a variety of authorized users who may need this information for analysis, reporting, and decision-making, particularly in emergencies. Tools should be available that allow access to data by authorized users so this information can be queried or extracted to other software for analysis (e.g., statistical or geospatial analysis).

Exchange of Data Among Systems. More information can be obtained from the same amount of data if the data are stored in a manner that facilitates the transfer of data to other systems or integration of data with other systems. Extensible Markup Language (XML) is commonly used for this purpose; XML facilitates information exchange among systems by encoding data with meaningful structure and semantics that computers and humans can understand (see <www.orafaq.com/glossary>). Other technologies or industry practices may also facilitate data flow among systems. This concept relates to standards 3.1 and 3.2.

Scalable. The database technology and architecture should allow appropriate levels of expansion of the system without compromising system efficiency.

Reliability. Reliability refers to the user's ability to consistently and efficiently use the database system for data entry or retrieval. Animal health surveillance systems must be able to consistently provide many users with simultaneous access to high quality data.

Legacy data (or historical data). Older data from some surveillance programs may have been collected and stored using outdated computer technologies. These data are often still a vital source of national animal health surveillance information and should be maintained for use by analysts and decision-makers. Legacy data should be accessible, and clear documentation should be available to describe these data to ensure usability in the future.

3.3 Data Security and Backup

Standard: Data for the surveillance program should be accessible to authorized users and be secure, recoverable, and reliable. The database is backed up to minimize intentional or accidental loss of information.

Supporting Information:

Security. The database should be strictly controlled and managed to prevent intentional or accidental loss of information. Security measures should include, but not be limited to, authorization and authentication of users, limiting privileges of users based on roles, encryption, data backups, and data storage in multiple locations. Security measures should address loss of data, loss of confidentiality, loss of privacy, loss of integrity, and loss of availability. The official process of data system security certification and accreditation for federal IT systems is described in standard 3.3.

Database backup. Backup is the process of periodically taking a copy of the database and moving it to an offline storage media[1]. Complete backup protocols and systems for animal health surveillance databases should be implemented to prevent loss of data that would effect daily operations. The backup system should consider how fast data recovery needs to occur, how long data must be available online, and methods for maintaining permanent backup media.

3.4 Security Certification and Accreditation (C&A)

Standard: All federal IT systems must be officially authorized, from a security standpoint, to operate.

Supporting Information:

Numerous directives explicitly state the requirement for owning agencies to perform a security certification and accreditation (C&A) of their IT systems. The C&A process involves a comprehensive evaluation of the technical, operational, and management controls in place to properly secure both the data and infrastructure of a given application. Accreditation is a formal acceptance, by the agency in question, of any residual risks once the security measures identified are in place, and is the official authority to operate the system as described in the associated documentation.

C&A standards require these steps:
- Establish and document access controls
- Maintain configuration management plan
- Perform a risk assessment
- Assess impact of data to personal privacy
- Ensure existence of a disaster recovery plan
- Create and maintain a system security plan

Critical components of the C& A process include:
- The system is certified and accredited every 3 years or following any major change to the system.
- The security classification of the system is properly assessed
- Required security documents are in place and updated as related attributes of the application evolve.
- The C&A process is integrated into the system development life cycle of any new system.

Information about the C&A process for federal IT systems is available through the VS Application and Information Management (VS-AIM) team; developers unfamiliar with the C&A requirements and process should contact VS-AIM.

---

[1] Connolly T, Begg D, Database Systems: A Practical Approach to Design, Implementation, and Management 3rd Edition. Addison-Wesley, 2002

3.5 Data Quality

Standard: The database system is designed, implemented, and maintained to ensure that data are of high quality.

    a. Systematic and proactive procedures, used at multiple points in the data flow for the surveillance program, are implemented to ensure data quality; this includes feedback processes designed to not only inspect data for quality but also to address and fix the root causes of data errors.

    b. Those responsible for data quality are identified and have a clear role in the data quality feedback process.

    c. Programming tools are used to prevent data entry errors (e.g., code that enforces business rules or filters data that is out of range for particular fields).

    d. Individuals responsible for data entry have training and resources to perform their job correctly.

    e. Data are periodically validated and verified to ensure that the information is sufficiently accurate and complete to meet the needs of the surveillance program, and feedback from this process is used to improve the procedures used to ensure data quality.

Supporting Information:

Data that are complete and error-free are essential for a surveillance program to provide accurate reports and meaningful analyses in a timely manner. Data quality is defined as the ability of the data to meet the requirements of all internal and external customers. The level of accuracy, completeness, and timeliness may depend on the goals of a particular surveillance program, but animal health surveillance data collection efforts within VS also need to address the information needs described in the VS Strategic Plan [2]. In particular, VS' strategic goals for national animal health surveillance require data that are accurate, complete, accessible, available, and compatible with data from multiple systems. To achieve this, all animal health surveillance data systems within VS must demonstrate commitment to data quality. High quality data is not something that "just happens;" it is the result of systematic and proactive quality management efforts throughout the surveillance program and data system.

As with Data Systems Design (Standard 3.1), the measurement and improvement of data quality is a broad subject, and there are many books and Web sites on this topic. However, these considerations are of particular concern to data quality in VS animal health surveillance systems:

- Animal health surveillance systems within VS, and their underlying databases, should reflect commitment to implementing systematic and proactive procedures to

---

[2] "Veterinary Services Strategic Plan FY 2004 to FY 2008," http://www.aphis.usda.gov/vs/pdf_files/strat_plan.pdf. Accessed March 20, 2006.

assure data quality. These procedures should go beyond mere data inspection/correction and reflect a system-wide strategy to identify and eliminate causes of inadequate data quality. The allocation of resources within a surveillance program should reflect this commitment to data quality.

- Surveillance systems should identify teams of individuals who are tasked with data quality management; this group should include database managers as well as those involved with data entry, analysis, and reporting.

- Data entry user interfaces may be designed to minimize data entry errors. This may be achieved by restricting data entry in particular fields to acceptable values, cross-validating information within and between records, minimizing the use of free-text fields, using standardized terminology, using standardized lookup tables to control data entry, forcing users to double-key critical information, and constraints on data entry that assure data integrity.

- Training those responsible for data collection and data entry is recognized as a critical component to the data quality management process. The implementation portion of a surveillance plan should describe the process for initial and ongoing training of data collection and data entry personnel (Chapter 1, standard 1.12). Data entry personnel should be provided with resources (e.g., documents, protocols, user's guides, web-based resources, help desk contacts) to facilitate data entry and address questions or concerns. Feedback processes should be established to help data collectors and data entry staff identify and correct errors.

- Data validation and verification processes should be established to periodically review the data for accuracy, integrity, and completeness. These processes can identify systematic problems with a data system and avert data integrity problems that can affect quality reporting and analysis. These processes should not only identify and correct errors, but should also identify causes of data quality problems and provide solutions. Further, mechanisms should be in place to implement solutions to data quality problems when they are identified.

3.6 Data System Documentation

Standard: Data system documentation must be readily available and clearly describe:
- The system requirements and business rules
- The system users, including their access and roles
- The system components and their relationships (database schemas, entity-relationship diagrams, and descriptions of tables, views, and stored procedures)
- Data dictionaries that explain what data are stored in the database(s), how these are formatted
- Details regarding data backup and recovery procedures
- Changes to the data system over time, including changes to data elements that may impact analysis and reporting
- The activities of the system's Change Control Board (CCB), if applicable

Supporting Information:

Comprehensive system documentation provides critical information to IT developers, data end users, and other stakeholders about what, where, and how data are stored. System documentation must be developed for all animal health surveillance data systems. Examples of documentation for VS data systems can be obtained from VS CADIA.

Documentation of Changes to the Data System. Changes to a data system must be carefully considered and fully documented. Changes to any data system can have far-reaching ramifications both to developers of the system and stakeholders who use the data in the system, especially when data systems are integrated and pull or receive data from multiple sources. It is important to have a critical review process of all changes that occur to a data system's architecture and design. This includes, but is not limited to, changes in data collection, table design, table relationships, and coding systems used within tables.

Critical Components:
- A Change Control Board (CCB) should be established that has members representing the IT team and officials involved in planning and implementing the surveillance system. The CCB approves the original system, the system documentation, and system changes.
- Periodic reviews of the data system should be conducted to evaluate new technologies and modernization of the current system.
- New data systems within the agency should be evaluated for integration and contribution to the existing system.
- The data system should be evaluated periodically by unbiased experts to evaluate system efficiency and design.
- Any changes to the data system should be documented completely, including updates to the data dictionary, entity relationship diagrams (ERDs), and user guides.

3.7 Data Confidentiality

Standard: The data system ensures the confidentiality of sensitive or private information.

      a. Appropriate authority is obtained to collect and maintain data that may be sensitive or confidential.

      b. Sensitive or confidential data elements are identified and a protocol is developed and implemented to restrict access to these elements; access is limited to authorized individuals.

      c. Staff are trained on data confidentiality policies and protocols.

      d. Information-sharing policies for databases are documented and adhered to, ensuring confidentiality.

      e. Federal and State information sharing laws and practices are carefully evaluated before linking, transferring, or sharing data with any organization.

Supporting Information:

Data confidentiality and privacy is a top priority for animal health surveillance systems. Surveillance systems must include approved protocols for collecting, storing, restricting access to, and working with confidential information. In addition, confidential information may only be shared when the situation complies with Federal and State laws.

First steps for addressing data confidentiality issues include:
- o Obtain authority to gather and maintain private information
- o Identify data that require restricted access
- o Identify groups of users (roles) and the data they can access
- o Restrict user access to data based on group membership (roles)
- o Determine if information on individuals or groups of individuals can easily be obtained
- o Ensure the information stored is used as authorized

For VS data systems, refer questions on data confidentiality and information sharing to APHIS Legislative and Public Affairs.