

# THE ESSENTIALS OF A DATABASE QUALITY PROCESS

*Dorothy M Blakeslee<sup>1\*</sup> and John Rumble, Jr.<sup>2</sup>*

<sup>1</sup>*National Institute of Standards and Technology, Standard Reference Data Program, 100 Bureau Drive, Stop 2310, Gaithersburg, MD 20899-2310, USA*

*Email: dorothy.blakeslee@nist.gov*

<sup>2</sup>*National Institute of Standards and Technology, Standard Reference Data Program, 100 Bureau Drive, Stop 2310, Gaithersburg, MD 20899-2310, USA*

*Email: john.rumble@nist.gov*

## ABSTRACT

*Many steps are involved in the process of turning an initial concept for a database into a finished product that meets the needs of its user community. In this paper, we describe those steps in the context of a four-phase process with particular emphasis on the quality-related issues that need to be addressed in each phase to ensure that the final product is a high quality database. The basic requirements for a successful database quality process are presented with specific examples drawn from experience gained in the Standard Reference Data Program at the National Institute of Standards and Technology.*

**Keywords:** Data, Database, Interface, Quality, Review, Scientific data, Software, Technical data

## 1 INTRODUCTION

The Standard Reference Data Program (SRDP) at the National Institute of Standards and Technology (NIST) has been providing critically evaluated scientific and technical reference data to scientists and engineers for over 30 years. Today approximately 26 PC databases and 26 online databases are available in a variety of disciplines, including chemistry, physics, and materials science. The core of any NIST Standard Reference Database is a collection of evaluated data, that is, data that have been selected and reviewed so that their quality is evident to the user community, most of whom are not expert in how the data were generated. The evaluation process is a labor-intensive activity, often spanning years of work and involving world-leading scientists. Data evaluation is the key to NIST data programs and provides the primary motivation for their existence.

NIST SRDP is customer-driven. The emphasis is on serving customers with the highest possible quality data through both computer databases and publications, where the term *database* is used in a generic sense to include data files, calculational programs, and expert systems, as well as database systems with interfaces for query and display. In recent years, computer databases have become increasingly important as the primary way the data are distributed. Thus, for NIST SRDP to provide high quality data, it is essential that the databases themselves be of high quality. In this paper, we present a process that is designed to ensure high database quality, a process that is applicable to both PC databases and online data systems.

## 2 PREVIOUS WORK

The last fifteen years have seen an emphasis on quality in every area of technology. Every type of industrial activity has been affected, and the quality of measurements is an integral part of this focus. However, as pointed out by Fox, Levitin, & Redman (1994), although data quality has become increasingly important with the rapid growth of computerized information systems, there is no agreement concerning

how data quality can be investigated or even what the term *data* means. Fox et al.(1994) discussed five definitions for data in the literature and then described four factors associated with the quality of data values: accuracy, currentness, completeness, and consistency.

A variety of approaches have been undertaken in the study of data quality and the criteria that can be used to measure the quality of a database. The problem is twofold: first, to determine the characteristics a database should possess in order to be considered of high quality and second, given that the desired characteristics have been agreed upon for a given set or category of databases, to evaluate how well a database meets those standards.

Some studies have approached the first part of this problem empirically by starting with the database users themselves to determine what measures of quality they consider most significant. Wang & Strong (1996) conducted two surveys of data consumers. Their first survey, in which they asked consumers to list the attributes that come to mind when they think about data quality, generated a list of potential data quality attributes. Using the results from this survey and a pretest, they conducted a second survey to determine the importance consumers attached to each of 118 different quality attributes. By analyzing the data collected, they determined four major data quality categories:

1. Intrinsic data quality
2. Contextual data quality
3. Representational data quality
4. Accessibility data quality

Several studies, including (Tenopir, 1990), (Granick, 1991), (Medawar, 1995), and (Wilson, 1998), have referred to a list of quality criteria identified by the Southern California Online Users Group at a retreat held in 1990. Medawar (1995) described these quality criteria for online databases as the following:

1. Consistency
2. Coverage/scope
3. Timeliness
4. Error rate/accuracy
5. Accessibility/ease of use
6. Integration
7. Output
8. Documentation
9. Customer support and training
10. Value-to-cost ratio

Wilson (1998) surveyed online and CD-ROM database users and asked them to rank these ten database quality criteria according to degree of importance. In another study (Kuhn, Deplanque, & Fluck, 1994), the authors concentrated on scientific databases and considered the criteria of relevance, comprehensiveness, and reliability.

As was mentioned above, given that a set of database quality criteria has been identified, the problem remains of determining how well a database meets these criteria. Some of these types of efforts have focused on the quality of the user interface to the database, while others have focused on the content. In one study (Park & Chee, 1999), a methodology was developed for evaluating user interface designs according to usability criteria. The Chemical Abstracts Service has developed a system for evaluating the content quality of the CA File (CAS, 1999) on the basis of accuracy, consistency, and comprehensiveness of access points (Adams, 1997). A team of staff members knowledgeable in chemistry carries out the audit process as described by Adams (1997).

Others have developed methodologies to test for specific types of errors in database content. Cahn (1994) compared different databases based on the results of keyword searches for a list of ten misspelled words. Jacsó (1993a) and Jacsó (1993b) performed a two-part study of online and CD-ROM databases. The first part focused on what Jacsó called errors of omission, e.g., data elements with incomplete or missing information. The second part focused on inaccuracies and inconsistencies.

A different approach to database quality that has been taken is to assess quality-related issues at each of the different steps in the production of a database. For example, in the case of online bibliographic databases, research (Rittberger & Rittberger, 1997) was done to determine whether database quality could be evaluated by using attributes and indicators for different steps in the production process, such as document acquisition and document analysis. Mintz (1990) also covered various issues involved in database production, including the importance of documentation as a quality factor.

Another factor that many users consider when determining how much credibility to attach to data is the source of the data. For example, users may attach more credibility to data from one type of publication than another (ASTM, 1992). The data source is especially important to many users of Web-based information (Wathen & Burkell, 2002). For additional discussion of issues relating to the database quality, see (Hoxmeier, 1998), (Hoxmeier & Monarchi, 1996), and (Arnold, 1992). Also, additional work not relevant to the current study has been done by (Castelli & Locuratolo, 1996) and (Castelli & Locuratolo, 1995) on database design and quality.

In this paper, we apply the database quality concepts summarized above to the database production process. We emphasize not what criteria or attributes can be used to measure database quality, but rather how a process can be implemented to ensure that certain quality attributes or criteria are present in finished database products. That is, we discuss not so much *what* needs to be achieved, but rather *how* to achieve it.

### 3 BACKGROUND

Because of its emphasis on critically evaluated data, NIST SRDP needs to be certain that when customers use NIST databases, the data they find are precisely the data that the data evaluators intended. Three quality criteria are of particular importance:

1. Accuracy - The data must be accurately stored and be the data as finalized by the data evaluators.
2. Correctness - The retrieved data must be the correct data for the property and substance selected.
3. Reliability - The database must be reliable, that is, it must work every time for every function.

As the trend toward making data available through databases increased, NIST SRDP found that databases prepared for public release were of uneven quality. While some were excellent and required little additional work before release, others had deficiencies. Many databases did not meet the above criteria and required considerable testing and correcting or modifying before they could be released. In addition, some databases, while meeting the above criteria, were still not of sufficiently high quality to meet the needs of a user community accustomed to commercial software with sophisticated interfaces. The goal is that all NIST database products be intuitive to use and follow commonly accepted standards of interface design so that a scientist in the field addressed by the database is able to use the database with a minimum of online help or written documentation.

Through the process of developing and releasing for sale to the public over 70 databases, NIST personnel have gained valuable experience. The importance of emphasizing planning and quality from the beginning of database projects has become very apparent. Understandingly enough, database builders feel that when they complete their product and submit it for release, their part is done. They are reluctant to make changes at that late stage. Also, the very process of making changes to a *completed* product can be very time-consuming, especially if the changes require a fundamental modification to the database design.

## 4 THE DATABASE PROCESS

Because of this need to build in quality, a NIST database quality process has been instituted, which emphasizes planning and communication well before any database work is begun. An internal manual, *The Database Process* (NIST, 1997), has been prepared that provides the basis for the quality aspects of the process, which has been divided into four phases:

1. Planning and design
2. Development and implementation
3. Product review and release
4. Post-release procedures

### 4.1 Planning and design

Careful planning is crucial to the success of a database quality process. A variety of decisions need to be made before any work can be done on the design or implementation of a scientific database. Some of these decisions are dependent on the type of database to be developed, for example, whether it contains bibliographic information, numeric data, or both. However, regardless of the type of database, certain questions need to be considered at the very beginning of the project.

Of paramount concern is finding the best way to meet the requirements of the projected user community. Some of the issues that need to be resolved include identifying the potential users and determining how they can be reached:

1. Is the database relevant only to users in one particular scientific field, or is there more general application for the data?
2. Are the potential users concentrated in universities, governmental agencies, commercial companies, or elsewhere?
3. Do the users have easy access to the Web, or are there compelling reasons why they need to have their own individual copies of the database or have the database installed on their own Intranet?
4. Is this a product that is integrated into computerized engineering systems or tools?
5. Is the database to be installed on any scientific instrument or equipment?

After the projected user community is identified, it is possible to begin to address the features and options that are to be made available with the database:

1. What are the different options that the users want to have available for searching the database?
2. What display features are to be provided?
3. Is there a demand for graphing capability or the display of spectra (or other x-y data) or chemical structures?
4. Is the database used primarily to search and display stored data or is there a need for routines for performing calculations?
5. Is there a need to download or transfer data from the database to the users' own programs or systems?
6. What type of documentation is required?
7. Is there a separate users' guide, or is all help information available online?

Many different ways can be used to determine user requirements, such as attending meetings and conferences of scientists in related fields and holding workshops and focus groups to bring together members of the user community. It is important to review any comments received from users of related databases or earlier versions, if any, of the database product.

Once the above questions and issues relating to the user community are resolved, questions relating to the design and implementation of the database can be addressed:

1. What are the system requirements?
2. Is a commercial database management system (DBMS) going to be used?
3. How large is the database?
4. Is extensive indexing required?
5. If the database is intended to be an Internet or Intranet system, how many users accessing the database concurrently are anticipated? Can the DBMS satisfy their demands without too much performance degradation?
6. If the database is to be a network or Web product, what kind of server is to be used? For example, does the server have a UNIX, Linux, or Windows operating system?
7. What are the skills of the programmers that are going to do the database development?
8. What programming language is to be used? For instance, does the interface development require using a language with a great deal of power and flexibility or can advantage be taken of rapid application development software? Is it necessary or highly desirable to program in a language that provides platform independence?

Finally, it is important to clarify the overall strategy for the type of database to be developed. Frequently, a choice has to be made between two different approaches. One approach is to concentrate on developing and releasing a product with minimal functionality in a relatively short period of time. A second approach is to produce a relatively complex product even though it takes longer to develop. To decide between these alternatives, consider the following questions. Is it better to get a product out quickly and evaluate user reaction in preparation for a follow-up version with more features? Or is there already such high user expectation for the database that a product with minimal functionality will be unacceptable to the user community? While rapid prototyping can be a very useful step toward producing a first version, often people are reluctant to abandon the investment in a prototype even if a new approach is needed.

One of the most valuable tools in the NIST database quality process is the organizational meeting. At this meeting, which includes everyone involved in the database project, the above questions and issues are reviewed in detail. Typically, as soon as a new database project is proposed and approved, the organizational meeting is held with the project participants, including managers, scientists or data evaluators, and database developers, to review the complete database process. *The Database Process* manual is used as a reference during this meeting, which often lasts up to two hours.

One of the main topics discussed at this meeting is the potential user community for the database and its needs. The details of what the database review covers are thoroughly described so that everyone who is involved in the project knows what to expect and can plan accordingly. In addition, issues of schedules and timing are discussed. Although it is not possible to predict the exact release date at this early stage in the process, it is important that everyone understands the issues involved so that no commitments or promises are made until a preliminary time line can be worked out and agreed upon.

During the initial planning process, and certainly immediately following the organizational meeting, a written description of the database project needs to be prepared and approved by everyone involved.

In addition to the organizational meeting, other smaller informal meetings are held as needed during the planning and design stage. For example, the scientists and database developers may meet to outline the basic design of the interface. Frequently, proposed screen layouts are prepared, then reviewed and evaluated by appropriate groups. If feasible, potential users are surveyed for their input. The database developers may meet with other database experts to go over the database schema and software to be used. In cases where the scientists and the database developers work in different groups, issues of how the data are to be transferred are resolved. During these database planning meetings, communication is repeatedly stressed. For the database process to work smoothly, it is not enough to have one group develop the database and another group review it; there must be a team effort throughout every stage of the process.

## 4.2 Development and implementation

Next is the database development and implementation phase, which may take anywhere from several months to a couple of years depending on the complexity of the project and the database development staff available to work on the project. The projects that tend to be short-term include data updates or conversions of current releases to different platforms, e.g., converting a PC database to a Web-based system. The longer-term projects tend to be new databases, which usually require a minimum of a year or more for development.

The development and implementation phase typically can be thought of as having two different components, one relating to the data and the other relating to the user interface. The data need to be loaded into a DBMS or otherwise converted into a format for easy access by the database program. An essential ingredient in building quality into the final product is the application of good database design procedures in designing the database schema. This cannot be accomplished without a thorough knowledge of the characteristics of the data and metadata to be included. For example, decisions need to be made about how the data are to be formatted and what standards or conventions are to be followed for scientific terms such as chemical names.

In addition, data validation procedures are performed as applicable:

1. Spell check text fields.
2. Check numeric fields to ensure that values fall within expected ranges. Verify the validity of any outliers.
3. Make sure abbreviations and acronyms are used consistently, that is, the same abbreviation or acronym is always used for the same item.
4. Ensure that no information is lost in the data loading process. For example, to preserve the number of significant figures in numeric data, it may be necessary to store those data in text fields or to add additional fields to keep of track of the numbers of significant figures in each numeric field.

Similarly, when developing the user interface, it is very important to follow good programming practices such as documenting the source code and using a consistent naming scheme for modules and variables. Each module or section of code is tested separately after it is developed.

If a team of database builders is doing the development, regular meetings are important to ensure that all team members are clear what their roles and responsibilities are as well as to provide opportunities for them to advise and assist each other. If there are any misunderstandings about how the data are to be handled by the program, the team members need to resolve them. The data expert needs to check the product periodically during development to ensure that the proper terminology is being used, the data items are labeled correctly, all numeric fields are displayed in the correct units, and units' conversions are done correctly.

The database authors review the database periodically as it is being developed. Sometimes it is very useful to give copies or prototype versions to colleagues for technical review or beta testing. In the case of new databases or databases with significant changes from prior versions, interim reviews may be indicated at certain stages in the development process.

In some cases, the data experts do not have the occasion to work closely with the database developers on a day-to-day basis. Sometimes databases are developed by outside contractors. At NIST, the data experts are located in data centers in one of the NIST Measurement and Standards Laboratories, whereas SRDP is located in a different operating unit. Because the database authors' offices and the SRDP office are separated both organizationally and physically, a special effort needs to be made to maintain frequent contact during all phases of the database process.

Whether the database development is being done by an SRDP developer, by a developer in one of the NIST data centers, or by an outside contractor, the need for communication continues throughout the development process. If the development is being done by separated parties, it is especially important that frequent meetings be held between the developer and the scientist to ensure that the menu options and the data displays are correct and appropriate for the science involved and meet the needs of the users. Some issues that may need to be clarified include the number of decimal places to be displayed for the numeric data and the units to be used. For example, NIST databases always use SI units as the default, but the users in some scientific fields need to have additional choices available.

One of the main advantages to maintaining contact among all interested parties during this phase is that it enables any problems that arise to be worked out in a timely fashion. Occasionally, for example, the data may not be ready as soon as was anticipated at the beginning of the project or certain database features may prove unexpectedly difficult to implement. Whenever necessary, adjustments are made to the proposed schedule for the review and release of the product. The primary goal is a high quality product, not necessarily one that is released quickly.

### 4.3 Product review and release

The next phase in a database quality process involves a formal review of the complete database product by the authors, i.e., the data evaluators and developers, and by an independent party. The review by the data evaluators is essential, because they are in the best position to determine whether the data in the database are accurate and being displayed as they intended. No database is released without final approval by the authors. The review by an independent party is also essential. Because of the natural tendency of database authors to miss noticing problems simply because they are so familiar with the database, there is no substitute for an independent review by someone who has not been involved in the design or development and has no preconceived ideas of how the database should work. What seems obvious to a developer may not be at all obvious to a user.

After installing the database for PC products or simply connecting to the URL for Web-based databases, the first step in the review process is to check the basic database identifying information. For Windows products, verify that the name of the database, the version number, the names of the authors, and any pertinent disclaimers are displayed on the opening screen or in the *About* dialog box. For Web-based products, verify that this information is displayed, or linked to, on the opening page. If an application is going to be made for a copyright, verify that the proper copyright notice is displayed and that the application is prepared and submitted.

Next the reviewer goes through the database systematically and tests all user options and, to the extent feasible, every possible path through the program. The following procedures are followed:

1. If the data were originally made available in a publication, perform random checks by comparing the values in the database with the corresponding values in the publication.
2. Check that reference citations are provided for all data.
3. If drop-down lists are provided for querying the database, test the different choices to see if they retrieve the correct data and if they are comprehensive.
4. For fields where numeric input is required, input both valid and invalid numbers to make sure the program handles both appropriately.
5. Check to see if the correct units are displayed with all numeric values.
6. If the program provides options to change units, test representative samples of the conversions to determine if they are working correctly.
7. If source code is included with the package, compile the code to verify that it compiles without any errors.
8. Test any special features, such as file input or output, data downloading, printing, or plotting.
9. Check that software vendors are credited if required.

The reviewer also evaluates the usability of the database by considering the following:

1. Are the menu options or user choices clearly presented so the user can understand what to do without any special instructions?
2. Are all data items well defined or described, e.g., are the units and the corresponding uncertainty values, if any, provided?
3. Is context sensitive help available when needed?

Both the online help and the printed users' guide are carefully proofread. At NIST, the database reviewer checks the users' guide for clarity and completeness. The standard format for NIST Standard Reference Database users' guides is documented in *The Database Process* manual (NIST, 1997).

In addition, certain procedures and specialized tests are performed depending on the database platform. For Windows products, the following apply:

1. Test the program using the mouse and also using keystrokes, that is, pressing ALT-letter, if that feature is available.
2. Check to ensure that all entry boxes are cleared when new searches are initiated.
3. If windows can be resized or moved, test those options.
4. Test running the database from a network drive.
5. If needed, prepare an installation program.
6. Install and test the database on different PCs with different versions of Windows as appropriate.

For Web-based products, different types of tests are required:

1. Check all hyperlinks see if they are clearly described and link to the correct location.
2. Check for return hyperlinks on every page.
3. Verify that a contact is provided for comments or questions.
4. Verify all email addresses.
5. Access the version history and check for completeness.
6. Check to see if pages display date created and last modified, if appropriate.
7. Test the site using different Web browsers.
8. Check to see if the site complies with the requirements for accessibility to people with disabilities.

Frequently, a question arises concerning what version of a browser can reasonably be required. This involves a tradeoff between using the latest, most sophisticated features that require users to have the most current version of a browser, versus using features that, while not the latest, are available to the great majority of users. Now that browser updates can be easily downloaded from the Web, NIST SRDP assumes its users' browsers are fairly up-to-date while, at the same time, avoiding features that work only for the very latest version.

During the review phase, the database developers and the reviewer communicate with each other frequently. Typically, there are several exchanges where the reviewer prepares a report for the developers of any problems that were found plus suggestions for changes or improvements to the interface or help information. The developers make the changes and corrections and return the revised product to the reviewer for retesting. To facilitate this process, it is important that the reviewer provide as complete a description of each problem as possible. If the developer knows not only what did not work correctly, but also what sequence of steps led to the problem, it is usually easier to trace and correct.

A systematic approach for testing data products has been established and refined during the years that SRDP has been reviewing NIST databases. Database Review Sheets, which provide checklists of recommended procedures and tests, have been developed for PC and Web products. These review sheets are updated periodically as requirements and technologies change.

For example, when NIST was releasing its early DOS databases, the users' guide was very important. Many users needed step-by-step instructions on how to install and use the database. A sample session was routinely included which showed how a typical user might use the product. This sample session provided copies of screens with descriptive text interleaved explaining exactly what text to enter and which keys to press. Today the typical NIST database user has extensive computer experience, so there is little need for detailed users' guides. Consequently, the requirements for users' guides have changed. A relatively brief printed manual is provided for PC products. In addition, a condensed Hypertext Markup Language (HTML) version of the manual is prepared for those products that can be downloaded as part of an online purchase.

No database is released to the public until the review process has been completed and all concerned parties, including the database authors, reviewers, and appropriate managers, approve the final versions of the database and the users' guide. Some organizations may have additional requirements to fulfill before release. For example, NIST has an editorial review board that must give its approval.

#### **4.4 Post-release procedures**

As soon as a NIST database product is released to the public, a description of the database is added to the NIST Standard Reference Data Program Web site (<http://www.nist.gov/srd>). For PC products, ordering information is provided along with the option of filling out an online order form. Most of the PC products are available for downloading after credit card information has been entered. For Internet products, a link is provided from the SRDP Web site to the database Web site. In addition, indexing information on the database is collected for inclusion in the next release of the NIST Data Gateway (<http://srdata.nist.gov/gateway>), a Web portal that provides easy access to many NIST databases.

After a database is released, the steps that are followed to build the database should be documented as completely as possible before the details of the development process have been forgotten. Depending on the type of database, this information might include a list of any special requirements, such as DLLs or other system files, compiler instructions, or documentation on how the data are indexed. At NIST, a copy of the final product plus data and source code files, if available, is placed in an archive for safekeeping. Related files, including the electronic version of the users' guide, are also archived for possible future reference.

On some occasions, a post-release meeting of all project participants, including the senior managers, scientists or data evaluators, database developers, and the reviewer, provides an excellent opportunity to go over what happened during the various phases of the database process. Valuable lessons can be learned both in terms of what went well and what problems were encountered. Frequently, useful suggestions for improvements that can be implemented in the next version are proposed at this meeting.

Because the ultimate measure of success of any database project is how well the database meets the users' needs, suggestions and comments are solicited from users in a variety of ways. Email addresses are provided with Web products. With PC products, contact information such as names, phone numbers, and email addresses, are provided with the database program and the users' guide. In NIST SRDP, a user response form is mailed with each database purchased. Copies of all relevant comments received from users are sent to the database authors and also maintained by NIST SRDP for its records. Thus, these comments are available for review when planning for the next version of the database.

It is also important to provide ongoing user support to handle any questions or problems that users have concerning the database. In particular, policies and procedures need to be in place for handling a situation in which a problem cannot be solved unless corrections or modifications are made to the database product. For example, a user may discover a problem with the database program.

NIST provides support for all the NIST Standard Reference Databases and responds to phone and email requests. Many of the problems or questions can be handled easily. Perhaps a CD is bad and needs to be

replaced, or a customer requires additional help with the installation process. Sometimes technical questions arise that need to be answered by the database author. However, occasionally a problem with the database program or an error in the data is discovered after the product has been released. In this case, the authors and SRDP evaluate the severity of the problem. If it is decided that the database needs to be corrected as soon as possible, that is, without waiting for the release of the next scheduled update, the developers and SRDP make producing a corrected version one of their highest priorities. Then, after the database has been corrected and tested, it is made available to all customers. For PC products, an errata letter is sent to all current users explaining the situation and letting them know how to obtain the corrected version. For Web products, the corrected version is made available online.

## 5 CONCLUSION

Ensuring that quality is built in from the very beginning of every database project is critical to the development of a high quality database. A process with well-defined procedures that are understood by all involved parties, including managers, data experts, database developers and reviewers, needs to be instituted. Careful planning at the start of every project and communication among all participants throughout the duration of the project are especially important parts of this process.

NIST SRDP has implemented a database quality process based on these precepts. An internal manual, *The Database Process* (NIST, 1997), has been prepared to address the quality aspects of the process of building a database. This manual serves as a guide for the various phases of the database process, including planning and design, development and implementation, product review and release, and post-release procedures. This process has been applied in over 20 database projects, and the results have been rewarding. Project participants have a better understanding of their roles and responsibilities. There are fewer misunderstandings resulting from unrealistic expectations. Scheduling conflicts have been reduced, and project milestones are more closely met.

More than 40 databases, new products and updates, have gone through the review phase of the database process. While the independent review continues to be an extremely important part of the process, the resources required for the review have been greatly reduced. The quality of the databases being submitted for review has substantially improved, and consequently the review cycle time has decreased. This is especially true for the Internet products, where the NIST SRDP reviewers typically review the database and return comments to the developer within a few days.

In summary, NIST SRDP strives to provide not only the highest quality scientific data, but also high quality databases to make those data readily available to customers. To achieve this, quality must be a consideration from the very first planning steps of a project, and a formal database quality process with well-defined procedures must be in place. Finally, although the procedures followed in the database process need to be continuously updated to keep pace with the rapid changes in information technology, planning and communication remain essential to the success of the database quality process.

## 6 ACKNOWLEDGEMENTS

The authors would like to thank Geraldine R. Dalton, L. Diane Decker, Phoebe Fagan, Cheryl Williams Levey, Joan C. Sauerwein, and Han Thai for their contributions to the database quality process at NIST SRDP.

## 7 REFERENCES

Adams C.L. (1997) Database quality: Standards, sampling and measurement, *Proceedings of the 1997 Chemical Information Conference* (pp. 7-15). Nimes, France.

Arnold, S.E. (1992) Information Manufacturing: The Road to Database Quality. *Database 15* (5), 32-39.

ASTM Committee E-49 on Computerization of Material and Chemical Property Data, Subcommittee E49.05 on Data and Database Quality (1992) *Standard Guide for Formatting and Use of Material and Chemical Property Data and Database Quality Indicators* (pp. 545-547). Philadelphia, PA, USA: American Society for Testing and Materials.

Cahn, P. (1994) Testing Database Quality. *Database* 17 (1), 23-30.

CAS (1999) Lesson 1: STN Online Searching Basics. Retrieved from the Chemical Abstracts website: <http://www.cas.org/ACAD/lesson1.html>

Castelli D. & Locuratolo E. (1995) Enhancing Database System Quality through Formal Design, *proceedings of the 4<sup>th</sup> Software Quality Conference 2* (pp. 359-366). Dundee, UK.

Castelli D. & Locuratolo E. (1996) Database Design for Quality, *Proceedings of the Third International Conference on Achieving Quality in Software* (pp. 237-248). Florence, Italy.

Fox C., Levitin A., & Redman T. (1994) The Notion of Data and Its Quality Dimensions, *Information Processing & Management* 30 (1), pp. 9-19.

Granick, L. (1991) Assuring the quality of information dissemination: responsibilities of database producers, *Information Services & Use* 11 (3), 117-136.

Hoxmeier J.A. (1998) Typology of database quality factors, *Software Quality Journal* 7, 179-193.

Hoxmeier J.A. & Monarchi, D. (1996) An Assessment of Database Quality: Design it Right or The Right Design?, *Proceedings of the Association for Information Systems Annual Meeting 1* (1), 416-418, Phoenix, AZ, USA.

Jacsó P. (1993a) Searching for Skeletons in the Database Cupboard Part I: Errors of Omission, *Database* 16 (1), 38-49.

Jacsó P. (1993b) Searching for Skeletons in the Database Cupboard Part II: Errors of Commission, *Database* 16 (2), 30-36.

Kuhn P., Deplanque R., & Fluck E. (1994) Criteria of Quality Assessment for Scientific Databases. *Journal of Chemical Information and Computer Sciences* 34, 517-519.

Medawar K. (1995) Database quality: a literature review of the past and a plan for the future, *Program* 29 (3), 257-272.

Mintz, A. (1990) Quality Control and the Zen of Database Production. *Online* 15, 15-23.

Munro, R.G., & Chen, H. (1997) Data Evaluation Methodology for High-Temperature Superconductors, In Nishijima, S., & Iwata, S., (Eds.), *Computerization and Networking of Materials Databases: Fifth Volume, ASTM STP 1311*, (pp. 198-210) Philadelphia, PA: American Society for Testing and Materials.

NIST (1997) *The Database Process*. An internal manual available from The National Institute of Standards and Technology, Standard Reference Data Program, 100 Bureau Drive, Stop 2310, Gaithersburg, MD 20899-2310.

Park, K.S., & Chee, H.L. (1999) A structured methodology for comparative evaluation of user interface designs using usability criteria and measures. *International Journal of Industrial Ergonomics* 23, 379-389.

Rittberger, M., & Rittberger, W. (1997) Measuring quality in the production of databases. *Journal of Information Science* 23 (1), 25-37.

Rumble, J. Jr., & Sturrock, C. (1998) Reliable Materials Data: The Whys and Wherefores of Data Evaluation, *AMPTIAC Newsletter* 2 (1).

Tenopir, C. (1990) Database Quality Revisited. *Library Journal* (Jan 10), 64-67.

Wang, R.Y. & Strong, D.M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12 (4), 5-34.

Wathen, C.N. & Burkell, J. (2002) Believe It or Not: Factors Influencing Credibility on the Web. *Journal of the American Society for Information Science and Technology* 53 (2), 134-144.

Wilson, T.D. (1998) EQUIP: a European survey of quality criteria for the evaluation of databases. *Journal of Information Science* 24 (5), 345-357.