

Standards, Guidance, and Practices for Improved Data Quality

PSAM-10

Martin B. Sattison

June 2010

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint should not be cited or reproduced without permission of the author. This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the United States Government or the sponsoring agency.

Standards, Guidance and Practices for Improved Data Quality

Martin B. Sattison
Idaho National Laboratory
Martin.Sattison@inl.gov

ABSTRACT

Since the mid 1980s, the Idaho National Laboratory (INL) has been involved with numerous projects to collect operational reactor data, develop data tools, and trend and analyze that data for the U.S. Nuclear Regulatory Commission (NRC). Because the data are used in subsequent risk analyses, the NRC must have confidence in the quality of the data delivered. The NRC tasked the INL to establish a comprehensive data quality assurance program.

A key element of that effort was to establish a set of requirements based on industry and government standards, guidance and good practices. This paper discusses where these were found, the process of extracting the requirements and recommendations from the source materials, and the consolidation of them into a concise set of 22 standards, guidance and practices for technical data quality. While this set was created specifically for the data programs within the NRC's Office of Research, they have broad application to many high-technology industries.

1 DATA QUALITY REQUIREMENTS

The first step taken in establishing a comprehensive data quality assurance program was to establish a set of standards, guidelines, and practices to which the program should conform. The primary activity in this effort was a literature search and review of the Office of Management and Budget (OMB) and NRC guidelines and other guidelines for maintaining data quality. The results of this work follow.

Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658) directed OMB to issue government-wide guidelines that provide policy and procedural guidance to federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by federal agencies. OMB issued its final guidelines on September 28, 2001 (Reference 1). Subsequent guidance was issued by OMB on February 22, 2002 (Reference 2). These guidelines required agencies subject to the Paperwork Reduction Act (which includes the NRC) to publish their own information quality guidelines within one year.

The OMB guidelines serve as the foundation for the NRC Information Quality Guidelines and were reviewed to better understand the underlying principles that all government agencies were to embrace. The literature review next explored other NRC documents for possible guidance in the area of data and information quality. These included NUREG reports, office instructions, regulatory guides, Commission (SECY) papers, and Advisory Committee on Reactor Safeguards (ACRS) reports. Finally, industry standards were examined for other pertinent practices. The results of this review are compiled in Section 2.

1.1 OMB Quality Guidelines

The OMB guidelines are general in nature and were written so "federal agencies can apply them in a common sense, workable, and appropriately tailored manner." The four guidelines are

1. Overall, agencies shall adopt a basic standard of quality (including objectivity, utility, and integrity) as a performance goal and should take appropriate steps to incorporate information quality criteria into agency information dissemination practices. Quality is to be ensured and established at levels appropriate to the nature and timeliness of the information to be disseminated. Agencies shall adopt specific standards of quality that are appropriate for the various categories of information they disseminate.
2. As a matter of good and effective agency information resources management, agencies shall develop a process for reviewing the quality (including the objectivity, utility, and integrity) of information before it is disseminated. Agencies shall treat information quality as integral to every step of an agency's development of information, including creation, collection, maintenance, and dissemination. This process shall enable the agency to substantiate the quality of the information it has disseminated through documentation or other means appropriate to the information.
3. To facilitate citizen review, agencies shall establish administrative mechanisms allowing affected persons to seek and obtain, where appropriate, timely correction of information maintained and disseminated by the agency that does not comply with OMB or agency guidelines. These administrative mechanisms shall be flexible, appropriate to the nature and timeliness of the disseminated information, and incorporated into agency information resources management and administrative practices.
4. The agency's pre-dissemination review, under paragraph III.2, shall apply to information that the agency first disseminates on or after October 1, 2002. The agency's administrative mechanisms, under paragraph III.3, shall apply to information that the agency disseminates on or after October 1, 2002, regardless of when the agency first disseminated the information.

1.2 NRC Information Quality Guidelines

The most pertinent guidelines for the purposes of this project were those developed by the NRC itself, the NRC Information Quality Program (Reference 3). The NRC Guidelines are designed to ensure that the Agency establishes and maintains an appropriate level of quality commensurate with the nature of the information. Thus, the most influential scientific, financial, and statistical data are subjected to the most rigorous quality standards. The NRC Guidelines also set the scope of information subject to the guidelines. Of particular interest to this project are two areas covered: (1) technical reports, and (2) information that other parties provide to the NRC upon which the NRC relies or which the NRC disseminates.

Consistent with the OMB guidelines, the NRC Information Quality Guidelines break down "quality" into three major components: utility, integrity, and objectivity. Utility is defined as "the usefulness of the information to its intended users." To ensure adequate utility, the NRC Guidelines discuss how to make information available to the intended users, how to obtain public feedback on the dissemination processes and the quality of the information itself, and how to assist the public in locating the information they seek.

Integrity is defined in the NRC Guidelines as "the security of information from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification." To promote data integrity, the NRC Guidelines address the issues of personnel security, computer security, information security, and records management.

The NRC Guidelines divide objectivity into the elements of presentation and substance. Information must be presented in a manner that is accurate, clear, complete, and unbiased. The substance of the information must be accurate, reliable, and unbiased. The NRC Guidelines recommend several forms of review, ranging from internal project reviews to formal independent peer reviews to independent advisory committee reviews, such as those performed by the Advisory Committee on Reactor Safeguards (ACRS).

To address the proper presentation of information, the NRC Guidelines refer to the Quality Management Control standards. Additional guidance is provided in the areas of transparency and clarity.

Reproducibility is also addressed. The Guidelines state that reproducibility of original and supporting data for “influential scientific, financial, or statistical information will be consistent with commonly accepted scientific, financial, or statistical standards.”

1.3 NRC Information Quality Handbook

The NRC Information Quality Handbook (Reference 4) is designed to ensure that the Agency establishes and maintains an appropriate level of quality commensurate with the nature of the information. The Handbook is presented in three parts. Part I is a reproduction of the NRC Information Quality Guidelines. Part II is an overview of the quality of NRC products and how quality is achieved as defined in the NRC Information Quality Guidelines. Part III is a discussion of the administrative process for requests for correction to information disseminated by the NRC.

Part II re-iterates the use of a graduated approach to ensure information quality and provides examples of NRC QC practices. These include (1) the appropriate level of management review and approval in the concurrence process, (2) internal review groups like the Committee for Review of Generic Requirements and the Probabilistic Risk Assessment Steering Committee, (3) public comment on NRC policy before it is finalized, (4) public comment on rules and other documents, (5) participation of the public and affected parties in meetings with both the staff and the Commission, (6) early and substantial feedback from the Agreement States, (7) independent peer review of research products, (8) independent review by the ACRS and other Advisory Committees, and (9) review by the five-member NRC Commission.

The Handbook recognizes that “just because the NRC does not apply their guidelines to a particular NRC information product does not mean that the NRC is any less committed to the quality of its information, whether ‘disseminated’ or not. Indeed, the NRC will ensure the level of quality appropriate to each kind of information it generates. Therefore, in effect, the primary difference is that information subject to the guidelines will also be subject to correction through the special administrative mechanism called for by OMB’s guidelines and the NRC’s conforming guidelines, whereas information not subject to the guidelines and/or information already disseminated through a comprehensive public review process, will not be subject to correction through this special administrative mechanism.”

The Handbook provides two tables that help determine the types of documents and information that are subject to the NRC Guidelines. NUREG/CR reports and web page content (the most common forms of dissemination for the information covered by this project) are specifically named as being subject to the NRC Guidelines.

1.4 NUREG Reports

NUREG/BR-0058, *Regulatory Analysis Guidelines of the US Nuclear Regulatory Commission* (Reference 5), and NUREG/BR-0184, *Regulatory Analysis Technical Evaluation Handbook* (Reference 6), were reviewed. While the scope of this project does not cover any specific regulatory analysis work, the data collected and assessed by the NRC have broad general application and could be used to support a variety of regulatory analyses. Thus, it was deemed prudent to review these documents for guidance pertaining to the quality of information.

Both of these documents encourage and follow prudent QA practices, such as using systematic, disciplined, open, and transparent processes, and providing clear and well-documented explanations of decisions and actions. Specifically, NUREG/BR-0058 called for the use of the “best available peer-

reviewed studies and data collected by accepted or best available methods.” However, neither report provided any new specific standards, guidance, or practices not already identified in the NRC Guidelines.

1.5 RES Office Instructions

The list of Office of Nuclear Regulatory Research (RES) Office Instructions was reviewed for any possible useful information. ADM-007, “RES Information Quality Procedures” (Reference 7) was identified as a document that should be reviewed. This instruction provides guidance on the administrative process for handling information correction requests and appeals. It does not provide any insights into quality standards, guidance, or practices outside the information correction process.

1.6 Regulatory Guides

The list of “Regulatory Guides (Reg Guides) for Power Reactors” provided on the NRC web site was reviewed for any Reg Guides that addressed information quality. There are a number of Reg Guides that address QA and QC programs during the design, construction, and operations phases of nuclear power plants, but these do not address the types of information associated with operational event experience and risk and reliability analyses.

Reg Guide 1.174, *An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis* (Reference 8), and Reg Guide 1.200, *An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities* (Reference 9), deal with the use of probabilistic risk assessment (PRA) in various risk-informed activities. Because the information of concern for this project supports a number of PRA-related programs, these documents were reviewed. Both of these Reg Guides address quality issues at a high level and do not provide specific guidance pertaining to operational or risk and reliability data. They both advocate using appropriate QA and QC methods in the development of a PRA, including peer reviews, industry PRA certification programs, and PRA cross-comparison studies. In the area of data and quantification, both Reg Guides stress the importance of the proper treatment of uncertainty. Reg Guide 1.174 goes further to suggest the need for documentation control methods, independent reviews, verification processes, and calculation checks as means to ensure quality.

Reg Guide 1.176, *An Approach for Plant-Specific, Risk-Informed Decision-Making: Graded Quality Assurance* (Reference 10), provides guidance on the use of PRA to establish a graded approach to the QA treatment of nuclear power plant structures, systems, and components. It recommends the following practice applicable to this effort:

“PRA evaluations in support of regulatory decisions should be as realistic as practicable, and appropriate supporting data should be publicly available for review.” (emphasis added)

1.7 ISO-7385

International Organization for Standardization (ISO)-7385, “Nuclear Power Plants – Guidelines to Ensure Quality of Collected Data on Reliability” (Reference 11), provides explicit guidance on the quality aspects of collecting reliability data from nuclear power plants. It was written from the perspective of establishing a new data collection system at a nuclear power plant; therefore, a significant portion of the standard addresses activities by the plant staff in preparing data reports, activities that are carefully prescribed to licensees by regulation. However, the underlying principles are apropos.

According to ISO-7385, a data collection system must:

- Clearly define the purpose of the data collection effort, the means of data collection and treatment, and the data needed
- Have written procedures and instructions
- Use a computerized system
- Use checking procedures to ensure validity of the information inserted in the database
- Define data treatments in procedures with special attention to (1) the type of treatment, (2) frequency of treatment, (3) output information, and (4) distribution
- Define in a procedure how to obtain ad hoc treatments to meet special needs for which a processing method has not been set up.

1.8 American Institute of Chemical Engineers

The American Institute of Chemical Engineers' (AIChE's) Center for Chemical Process Safety published a book titled *Guidelines for Improving Plant Reliability through Data Collection and Analysis* (Reference 12). Chapter 6 of that book, "Quality Assurance of Data," provides excellent guidance on data quality. It discusses the characteristics of high-quality data: completeness, compliance with set definitions, accurate input, and proper data handling. Next it discusses some basic principles of quality as applied to equipment reliability data. This is followed by a discussion of data verification. Lastly, there are guidelines for quality plans for reliability databases.

1.9 National Institute of Statistical Sciences

The National Institute of Statistical Sciences (NISS) published a technical report in March 2005 titled *Data Quality: A Statistical Perspective* (Reference 13) that took a close look at the problem of data quality from the technical perspective of a statistician. The primary principle is that "good" or "improved" data quality is often not an end in itself because it is the quality of the decisions, not the data that ultimately matters. Data quality is the capability of data to be used effectively, economically, and rapidly to inform and evaluate decisions. Data quality should be present from the start, rather than created by re-work.

This report covers the topics of what is data quality, measurement, and assessment of data quality, characteristics of quality data (objectivity, utility, integrity, accuracy, completeness, consistency, validity, accessibility, identifiability, interpretability, integrability, relevance, and timeliness), components of the data quality process, and provides case studies as illustration of these topics.

2 COMPILED STANDARDS, GUIDANCE, AND PRACTICES

After reviewing the documents discussed above, the pertinent standards, guidance, and recommended practices were compiled into Table 1. Note that not all reviewed documents provided a contribution.

A number of the references had very similar standards and recommendations so the similar items across references were consolidated. Table 2 shows the resulting 22 consolidated standards, guidance, and practices that should be considered in establishing an effective data quality assurance program. The third column in Table 2 gives the cross reference(s) to the items in Table 1. The fourth column indicates a QA phase. Three QA phases have been defined. The first phase is the process phase where the day-to-

day data collection and processing takes place. The second phase is the periodic review phase where routine management QA activities are performed either on a set schedule or prior to delivery of a data product. The third phase is the programmatic phase where external, independent organizations conduct high-level reviews of the overall quality program. This would include ACRS reviews, independent peer review committees, public comment opportunities, and audits.

3 REFERENCES

1. Office of Management and Budget, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies," Federal Register, Vol. 66, No. 189, p. 49718, September 28, 2001.
2. Office of Management and Budget, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication," Federal Register, Vol. 67, No. 36, p. 8452, February 22, 2002.
3. U.S. Nuclear Regulatory Commission, "NRC Information Quality Program," Management Directive 3.17, April 9, 2009.
4. U.S. Nuclear Regulatory Commission, "Information Quality Handbook," NRC Internal Web Site, not available to the public.
5. U.S. Nuclear Regulatory Commission, *Regulatory Analysis Guidelines of the U.S. Nuclear Regulatory Commission*, NUREG/BR-0058, Revision 4, September 2004.
6. U.S. Nuclear Regulatory Commission, *Regulatory Analysis Technical Evaluation Handbook*, NUREG/BR-0184, January 1997.
7. U.S. Nuclear Regulatory Commission, "RES Information Quality Procedures," ADM-007, Revision 0, July 29, 2005.
8. U.S. Nuclear Regulatory Commission, *An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis*, Regulatory Guide 1.174, Revision 1, November 2002.
9. U.S. Nuclear Regulatory Commission, *An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities*, Regulatory Guide 1.200, Revision 1, January 2007.
10. U.S. Nuclear Regulatory Commission, *An Approach for Plant-Specific, Risk-Informed Decision-Making: Graded Quality Assurance*, Regulatory Guide 1.176, August 1998.
11. International Organization for Standardization, "Nuclear Power Plants – Guidelines to Ensure Quality of Collected Data on Reliability," ISO-7385, First Edition, 1983.
12. American Institute of Chemical Engineers, Center for Chemical Process Safety, *Guidelines for Improving Plant Reliability through Data Collection and Analysis*, New York: New York, 1998.
13. National Institute of Statistical Sciences, *Data Quality: A Statistical Perspective*, Technical Report Number 151, A. F. Karr, A. P. Sanil, and D. L. Banks, March 2005.

COPYRIGHT STATEMENT

This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Table 1. Data quality standards, guidance, and practices.

Reference Document	Standard, Guidance, or Practice
1. NRC Information Quality Guidelines	a. Use a graded approach. The most influential data should be subjected to the most rigorous quality standards.
	b. Adhere to NRC policy on the dissemination of information to the public, which clearly specifies what is to be made available to the public and when.
	c. Assist the public in quickly and conveniently locating the information they seek through the NRC's Public Document Room or its web site.
	d. Computer systems must be designed and tested to prevent inadvertent or deliberate alteration and assure appropriate access controls.
	e. Employees and contractors with access to electronic information and associated computer systems are screened for trustworthiness and assigned the appropriate level of access.
	f. Agency records must be properly maintained and protected. The ADAMS system was cited as a properly protected system.
	g. Provide formal review of and concurrence with all information disseminated (including technical reports).
	h. Encourage peer review of NRC research products.
	i. Adhere to Quality Management Control standards prior to disseminating information at the NRC's public web site.
	j. Apply sound statistical and research methods to generate data and analytical results for scientific and statistical information.
	k. Use peer reviews of agency-sponsored research that is relied upon.
	l. Use reviews of agency information by independent advisory committees, as appropriate, including the ACRS.
	m. Provide opportunities for the public and States to comment on rulemakings, Commission policy statements, regulatory guides, and other information products, as appropriate.
	n. Hold public meetings to seek public views and solicit public comments through the NRC's web site and Federal Register notices, as appropriate.
	o. Include in relevant agency information products descriptions of the data and methods used to develop the information product in a way that would make it possible for an independent, qualified individual or organization to reproduce the results.
	p. Adhere to the NRC policy and guidance overseeing the performance of regulatory analyses as provided in NUREG/BR-0058 and NUREG/BR-0184. <i>Note: While this project does not cover regulatory analyses, the data provided by the INL may be used by those performing such analyses, and therefore, must be able to support the assessment of uncertainty required by these documents.</i>
	q. Adhere to the agency's Plain Language Program in written and electronic products.
	r. Ensure that all disseminated information receives appropriate editorial review.
2. NRC Information Quality Handbook	a. Use a graduated approach to ensure information quality.
	b. Use appropriate level of management review and approval in concurrence process.
	c. Use internal review groups, like Committee to Review Generic Requirements (CRGR), as appropriate.
	d. Provide opportunities for public comment, as appropriate.
	e. Obtain early and substantial feedback from the Agreement States.
	f. Use independent peer review of research products.
	g. Use independent review by ACRS.
	h. Ensure the level of quality appropriate for all information generated, whether subject to the NRC Guidelines or not.
3. Reg Guide 1.176	a. PRA supporting data should be available for public review, as appropriate.
4. ISO-7385	a. Clearly define the purpose of the data collection effort, the means of data collection and treatment, and the data needed.
	b. Have written procedures and instructions for data collection and processing.
	c. Use a computerized system for the database.
	d. Use checking procedures to ensure validity of the information inserted in the database.
	e. Define data treatments in procedures with special attention to (1) the type of treatment, (2) frequency of treatment, (3) output information, and (4) distribution.

Reference Document	Standard, Guidance, or Practice
	f. Define in a procedure how to obtain ad hoc treatments to meet special needs for which a processing method has not been set up.
5. AIChE	a. High-quality data is (1) complete with respect to the information needed for the intended analysis, (2) gathered in compliance with well-documented definitions, and (3) accurately input and handled. b. Seven major components of a quality data system: (1) processes have a documented mission, defining the database's purpose and expectations, (2) the basis for the collection work process design, expected output, and use of output are clearly specified, (3) the scope of the database is clearly defined, specifying inputs/outputs, resource requirements and limitations, (4) a database administrator should be appointed, (5) work processes should specify communications of quality system results – to whom, how often, by what means, (6) a verification system, including a measurement process to assess whether or not objectives are being met, is required, (7) a continuous improvement mechanism should be in place to evaluate and implement changes. c. Only validated data should be accepted into the database. d. Data should be recorded according to standard approved and implemented procedures. e. A method to ensure data security must be in place. f. A forum for promoting a common interpretation and understanding among data users should be provided. g. All deviations should be formally documented and submitted in a timely manner. h. Quality review meetings should occur on a periodic basis. i. Historical data must adhere to the same standard of quality that is expected of “contemporary” data. There should be more concern about the effects of populating the database with large amounts of uncontrolled data than the loss of such data. j. There should be an internal verification of data quality containing (1) a list of items to check, and (2) acceptance criteria for the checklist. k. Data should be verified before acceptance into the database. l. There should be an external audit of the quality assurance work process in place. m. There should be a quality plan that includes (1) definition of the database mission, (2) description of the scope of the database, (3) basic policies for ensuring the quality of the data, (4) procedures for handling the data, including confidentiality, (5) specification of data requirements, (6) standardized reports for data analysis, (7) procedure for verifying data quality prior to acceptance into the database, (8) procedures for reporting and handling deviations, (9) procedures for handling corrective actions.
6. NISS Report	a. Incorporate data checks in the Extract, Transform, and Load process. b. Identify data elements that are invalid because they violate physical, logical, or metadata-based constraints that can be specified independently of the actual data observed. c. Use standardization and record matching to identify and remove duplicates. d. Incorporate data quality from the start, rather than create by re-work. e. Data quality metrics should (1) not depend on the size of the database, (2) should reflect users' need for the data, and (3) the number of metrics should not be too large. f. Checks should exist for completeness, consistency, and validity. g. Each record must have a unique identifier (primary key). h. The data must be accessible (dissemination systems, modern file formats, current software current platforms). i. The data from multiple databases should be easily integrated (consistent attribute definitions, same primary keys, etc.). j. There should be a prescribed means of anomaly detection (rule-based corrections, statistical methods, visualizations). k. Each piece of data should be documented as either original data or derived data. l. Whenever there is a change to the database structure or the scope of the data being collected, there should be a cross-walk plan to collect data by both the old and new methods for a period of time.

Table 2. Consolidated data quality standards, guidance, and practices.

No.	Consolidated Standard, Guidance or Practice	Table 1 References	QA Phase
1	Use a graded approach to ensure information quality. The most influential data should be subjected to the most rigorous quality standards. Ensure the level of quality is appropriate for all information generated, whether subject to the NRC Guidelines or not.	1a, 2a, 2h	1
2	Adhere to NRC policy on the dissemination of information to the public, which clearly specifies what is to be made available to the public and when. Adhere to Quality Management Control standards prior to disseminating information at the NRC's public web site. Assist the public in quickly and conveniently locating the information they seek through the NRC's Public Document Room or its web site. Provide opportunities for the public and States to comment on rulemakings, Commission policy statements, regulatory guides, and other information products, as appropriate. PRA supporting data should be available for public review, as appropriate. Hold public meetings to seek public views and solicit public comments through the NRC's web site and Federal Register notices, as appropriate.	1b, 1c, 1i, 1m, 1n, 2d, 3a	1, 3
3	Use a computerized system for the database. A method to ensure data security must be in place. Computer systems must be designed and tested to prevent inadvertent or deliberate alteration and assure appropriate access controls. Employees and contractors with access to electronic information and associated computer systems are screened for trustworthiness and assigned the appropriate level of access. Agency records must be properly maintained and protected. The ADAMS system was cited as a properly protected system.	1d, 1e, 1f, 4c, 5e	1
4	Provide formal review of and concurrence with all information disseminated (including technical reports). Use appropriate level of management review and approval in concurrence process.	1g, 2b	2, 3
5	Encourage peer review of NRC research products. Use peer reviews of agency-sponsored research that is relied upon. Use independent peer review of research products.	1h, 1k, 2f	3
6	Use reviews of agency information by independent advisory committees, as appropriate, including the ACRS. Use internal review groups, like CRGR, as appropriate.	1l, 2c, 2g	3
7	Obtain early and substantial feedback from the Agreement States.	2e	3
8	Quality review meetings should occur on a periodic basis.	5h	2, 3
9	Apply sound statistical and research methods to generate data and analytical results for scientific and statistical information. Adhere to the NRC policy and guidance overseeing the performance of regulatory analyses as provided in NUREG/BR-0058 and NUREG/BR-0184. (While this project does not cover regulatory analyses, the data provided by the INL may be used by those performing such analyses, and therefore, must be able to support the assessment of uncertainty required by these documents.)	1j, 1p	1, 2, 3
10	Include in relevant agency information products descriptions of the data and methods used to develop the information product in a way that would make it possible for an independent, qualified individual or organization to reproduce the results.	1o	1, 2
11	Ensure that all disseminated information receives appropriate editorial review. Adhere to the agency's Plain Language Program in written and electronic products.	1q, 1r	1, 2
12	Clearly define the purpose of the data collection effort, the means of data collection and treatment, and the data needed. Data should be recorded according to standard approved and implemented procedures. Have written procedures and instructions for data collection and processing. Define data treatments in procedures with special attention to (1) the type of treatment, (2) frequency of treatment, (3) output information, and (4) distribution. Define in a procedure how to obtain ad hoc treatments to meet special needs for which a processing method has not been set up. All deviations should be formally documented and submitted in a timely manner.	4a, 4b, 4e, 4f, 5d, 5g	1, 2, 3
13	Use checking procedures to ensure validity of the information inserted in the database. Only validated data should be accepted into the database. Data should be verified before acceptance into the database. Incorporate data checks in the Extract, Transform, and Load process. Identify data elements that are invalid because they violate physical, logical, or metadata-based constraints that can be specified independently of the actual data observed. Checks should exist for completeness, consistency and validity. There should be an internal verification of data quality containing (1) a list of items to check, and (2) acceptance criteria for the checklist.	4d, 5c, 5j, 5k, 6a, 6b, 6f	1

No.	Consolidated Standard, Guidance or Practice	Table 1 References	QA Phase
14	Incorporate data quality from the start, rather than create by re-work. High-quality data is (1) complete with respect to the information needed for the intended analysis, (2) gathered in compliance with well-documented definitions, and (3) accurately input and handled. Seven major components of a quality data system: (1) processes have a documented mission, defining the database's purpose and expectations, (2) the basis for the collection work process design, expected output, and use of output are clearly specified, (3) the scope of the database is clearly defined, specifying inputs/outputs, resource requirements and limitations, (4) a database administrator should be appointed, (5) work processes should specify communications of quality system results – to whom, how often, by what means, (6) a verification system, including a measurement process to assess whether or not objectives are being met, is required, (7) a continuous improvement mechanism should be in place to evaluate and implement changes. There should be a quality plan that includes (1) definition of the database mission, (2) description of the scope of the database, (3) basic policies for ensuring the quality of the data, (4) procedures for handling the data, including confidentiality, (5) specification of data requirements, (6) standardized reports for data analysis, (7) procedure for verifying data quality prior to acceptance into the database, (8) procedures for reporting and handling deviations, (9) procedures for handling corrective actions.	5a, 5b, 5m, 6d	1
15	The data must be accessible (dissemination systems, modern file formats, current software, current platforms).	6h	1
16	A forum for promoting a common interpretation and understanding among data users should be provided.	5f	2, 3
17	There should be an external audit of the QA work process in place.	5l	3
18	Each record must have a unique identifier (primary key). Use standardization and record matching to identify and remove duplicates. The data from multiple databases should be easily integrated (consistent attribute definitions, same primary keys, etc.).	6c, 6g, 6i	1, 2
19	Data quality metrics should (1) not depend on the size of the database, (2) should reflect users' need for the data, and (3) the number of metrics should not be too large.	6e	1, 2, 3
20	There should be a prescribed means of anomaly detection (rule-based corrections, statistical methods, visualizations).	6j	1, 2
21	Each piece of data should be documented as either original data or derived data.	6k	1
22	Whenever there is a change to the database structure or the scope of the data being collected, there should be a cross-walk plan to collect data by both the old and new methods for a period of time. Historical data must adhere to the same standard of quality that is expected of "contemporary" data. There should be more concern about the effects of populating the database with large amounts of uncontrolled data than the loss of such data.	5i, 6l	1