

# **Data Quality Assurance Program Plan for NRC Division of Risk Analysis Programs at the INL**

Martin B. Sattison  
Thomas E. Wierman  
Kurt Vedros  
Shawn W. St. Germain  
Steven A. Eide  
Robert L. Sant

July 2009



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

# **Data Quality Assurance Program Plan for the NRC Division of Risk Analysis Programs at the INL**

**Martin B. Sattison  
Thomas E. Wierman  
Kurt Vedros  
Shawn W. St. Germain  
Steven A. Eide  
Robert L. Sant**

**July 2009**

**Idaho National Laboratory  
Risk, Reliability, and NRC Programs Department  
Idaho Falls, Idaho 83415**

**<http://www.inl.gov>**

**Prepared for the  
Data Quality Improvement Project  
Division of Risk Analysis  
Office of Nuclear Regulatory Research  
U.S. Nuclear Regulatory Commission  
Washington D.C. 20555  
Job Code N6145**

#### NOTICE

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, express 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 herein, or represents that its use by such third party would not infringe privately owned rights. The views expressed herein are not necessarily those of the U.S. Nuclear Regulatory Commission.

## **Abstract**

This report documents the activities and results of the Data Quality Improvement Project, Job Code Number N6145, conducted for the NRC's Division of Risk Analysis (DRA) by the Idaho National Laboratory (INL).

The INL has collected, coded, and analyzed commercial reactor operational experience data for nearly two decades. Those efforts have evolved over time, including the associated data quality assurance and control activities. The purpose of this project is to identify the actions needed to ensure that the data that feeds the various programs under the direction of DRA follow the applicable quality assurance guidelines and standards.

The activities addressed by this report include

- Review of data quality guidelines and standards and compilation of a consensus set of standards, guidelines and good practices
- Review of current data programs, products, and data quality control activities
- Evaluation of the quality assurance programs applied by the Institute for Nuclear Power Operations to the Equipment Performance and Information Exchange (EPIX) and characterization of EPIX quality
- Evaluation of current quality assurance activities at the INL against these requirements
- Creation of a Data Quality Program Plan to address activities to enhance the current data quality assurance processes.



# Contents

Abstract .....	iii
Executive Summary .....	ix
Data Quality Requirements .....	ix
Current Data Quality Activities .....	x
EPIX Data Quality .....	xiii
Data Quality Program Plan .....	xiv
Preface .....	xv
Abbreviations, Acronyms, and Initialisms .....	xvii
1. Data Quality Requirements .....	3
1.1 OMB Quality Guidelines .....	3
1.2 NRC Information Quality Guidelines .....	4
1.3 NRC Information Quality Handbook .....	4
1.4 NUREG Reports .....	5
1.5 RES Office Instructions .....	5
1.6 Regulatory Guides .....	5
1.7 ISO-7385 .....	6
1.8 American Institute of Chemical Engineers .....	6
1.9 National Institute of Statistical Sciences .....	6
1.10 Compiled Standards, Guidance and Practices .....	7
2. Current INL Data Quality Control Processes .....	13
2.1 Data Collection and Coding .....	13
2.1.1 Data Sources .....	13
2.1.2 Quality Controls .....	14
2.1.3 Products .....	16
2.2 Data Computations and Trending Process .....	17
2.2.1 Data Sources .....	17
2.2.2 Quality Controls .....	17
2.2.3 Products .....	19
3. EPIX Quality Measurement .....	23
3.1 EPIX Data Background .....	23
3.2 EPIX Data Quality Measurement Program .....	23

3.2.1	Tier 1 EPIX Data Quality Measurement.....	23
3.2.2	Tier 2 INL Data Quality Measurement.....	24
3.2.3	Product Dependency .....	25
3.2.4	Statistical Evaluation of EPIX Data Uncertainty .....	25
4.	Comparison of Current Practices against Standards.....	27
4.1	Graded Approach to Data Quality .....	27
4.2	Dissemination to the Public.....	27
4.3	Computerized Database Systems.....	28
4.4	Formal Review and Concurrence .....	28
4.5	Peer Review of NRC Research Products.....	29
4.6	Independent Advisory Committee Reviews .....	29
4.7	Feedback from Agreement States.....	30
4.8	Quality Review Meetings .....	30
4.9	Sound Statistical and Research Methods.....	30
4.10	Descriptions of Data and Methods .....	31
4.11	Editorial Review.....	31
4.12	Written Procedures and Instructions .....	31
4.13	Checking Procedures and Data Validity .....	32
4.14	Incorporating Data Quality from the Start .....	32
4.15	Data Accessibility .....	33
4.16	Data Users' Forum .....	34
4.17	External Audits.....	34
4.18	Database and Record Standardization.....	34
4.19	Data Quality Metrics .....	35
4.20	Anomaly Detection .....	35
4.21	Original versus Derived Data.....	35
4.22	Cross-Walk Plans for Changes in Database Structure or Scope .....	36
5.	Data QA/QC Plan.....	37
5.1	Improve Data Protection from Loss .....	37
5.2	Peer Reviews .....	37
5.3	Independent Advisory Committee Reviews .....	37
5.4	Quality Review Meetings .....	38
5.5	Editorial Reviews .....	38
5.6	Checking Procedures and Data Validity.....	38

5.7	Incorporating Data Quality from the Start.....	39
5.7.1	Data Quality Metrics.....	39
5.7.2	User Feedback and Corrective Actions.....	39
5.8	External Audits.....	40
6.	References .....	41

## Tables

1.	Data quality standards, guidance, and practices.....	8
2.	Consolidated data quality standards, guidance, and practices. ....	10
3.	NRC product dependency matrix.....	25

## Figures

1.	Overview of current data processing. ....	13
2.	Overview of data collection and processing. ....	15
3.	Overview of computations and trending.....	18





## Executive Summary

The Division of Risk Analysis (DRA), Office of Nuclear Regulatory Research (RES), must ensure that the quality of the data that feed into its programs follow Office of Management and Budget (OMB) and U.S. Nuclear Regulatory Commission (NRC) guidelines and possibly other standards and guidelines used in nuclear power plant risk analyses. This report documents the steps taken in DRA's Data Quality Improvement project (Job Control Number N6145) to develop a Data Quality Assurance Program Plan.

These steps were

1. Conduct a review of data quality requirements
2. Review current data programs, products, and data quality control activities
3. Review the Institute of Nuclear Power Operation (INPO) Equipment Performance and Information Exchange (EPIX) data quality programs and characterize the EPIX data quality and uncertainty
4. Compare these programs, products, and activities against the requirements
5. Develop a program plan that provides assurance that data quality is being maintained.

It is expected that the Data Quality Assurance Program Plan will be routinely implemented in all aspects of future data collection and processing efforts and that specific portions will be executed annually to provide assurance that data quality is being maintained.

### Data Quality Requirements

The primary activity in this effort was a literature search and review of the OMB and NRC guidelines and any other guidelines that may have applicability for maintaining data quality. The documents reviewed were

- The government-wide guidelines that were issued by the OMB in response to Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001
- The NRC Information Quality Program
- The NRC Information Quality Handbook
- NUREG/BR-0058, *Regulatory Analysis Guidelines of the US Nuclear Regulatory Commission* and NUREG/BR-0184, *Regulatory Analysis Technical Evaluation Handbook*
- ADM-007, "RES Information Quality Procedures"
- Several regulatory guides (Reg Guides) for power reactors:
  - Reg Guide 1.174, *An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis*
  - Reg Guide 1.200, *An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities*
  - Reg Guide 1.176, *An Approach for Plant-Specific, Risk-Informed Decision-Making: Graded Quality Assurance*
- International Standard, ISO-7385, "Nuclear Power Plants – Guidelines to Ensure Quality of Collected Data on Reliability"
- The American Institute of Chemical Engineers' Center for Chemical Process Safety's *Guidelines for Improving Plant Reliability through Data Collection and Analysis*
- The National Institute of Statistical Sciences' technical report, *Data Quality: A Statistical Perspective*

The final product of this review was the following compilation of standards, guidelines, and recommended practices:

1. A graded approach to ensure information quality
2. Policy on the dissemination of information to the public
3. Computerized database systems
4. Formal review and concurrence
5. Peer review of NRC research products
6. Independent advisory committee reviews
7. Feedback from the Agreement States
8. Quality review meetings
9. Sound statistical and research methods
10. Descriptions of the data and methods
11. Editorial reviews
12. Written procedures and instructions for data collection and processing
13. Checking procedures and data validity
14. Incorporation of data quality from the start
15. Data accessibility
16. Data users' forums
17. External audits
18. Database and records standardization
19. Data quality metrics
20. Anomaly detection
21. Original versus derived data
22. Cross-walk plans for changes in database structure or scope.

### **Current Data Quality Activities**

A review of the current data quality activities was performed to establish a baseline for evaluation. The major steps in the data collection and coding process are depicted in Figure ES-1. Blocks containing steps involving some form of quality assurance/quality control activity are numbered in the upper right corner and are outlined in blue.

The quality controls in the data collection and coding process are

- Automated quality checks (Blocks 1, 2, 3 and 7)
- Manual quality checks (Blocks 4, 5, 6, and 8)
- INL manager review (Block 9).

The major steps in the data computation and trending process are depicted in Figure ES-2. Blocks containing steps involving some form of QA/QC activity are numbered and outlined in blue.

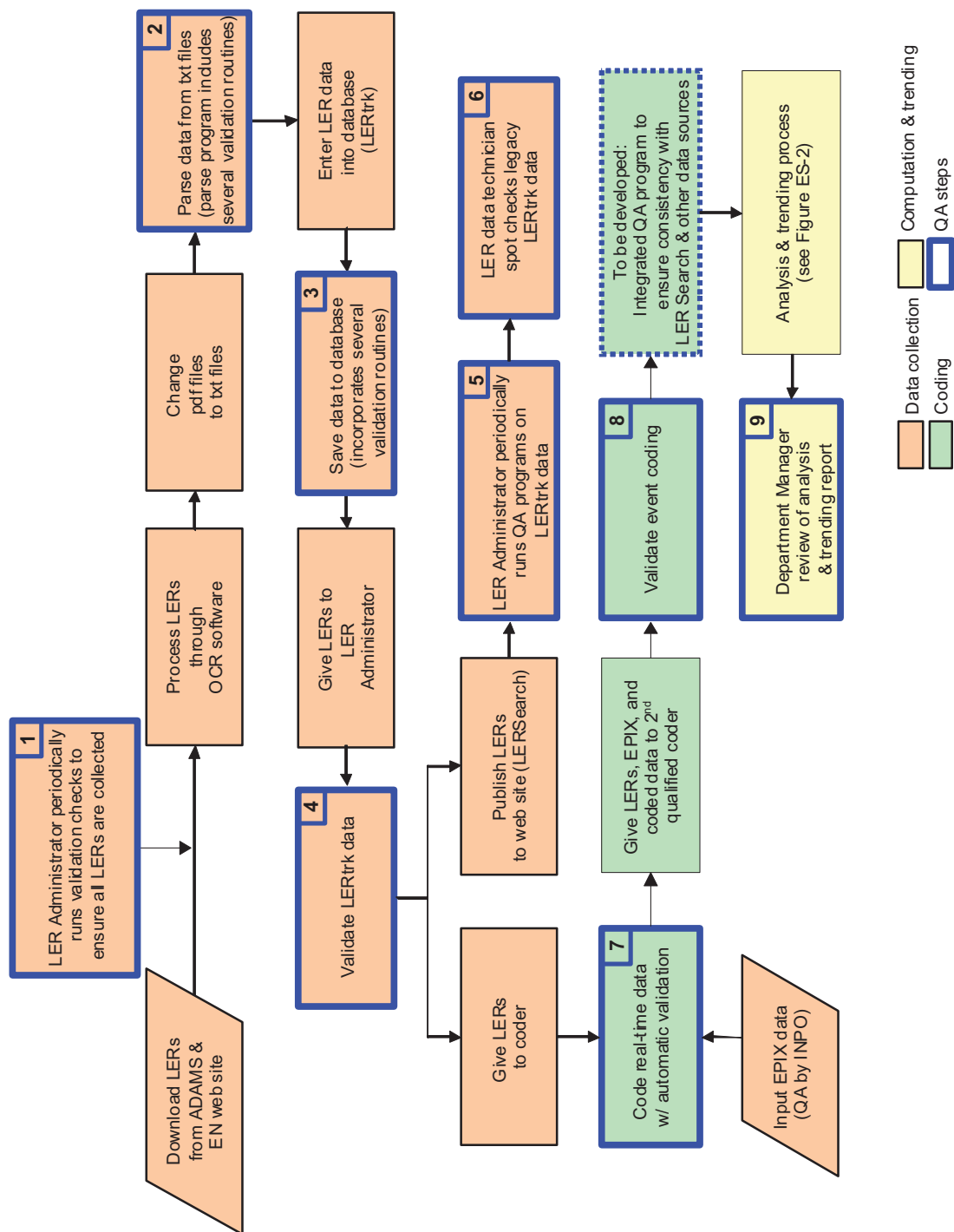


Figure ES- 1. INL data collection and coding process.

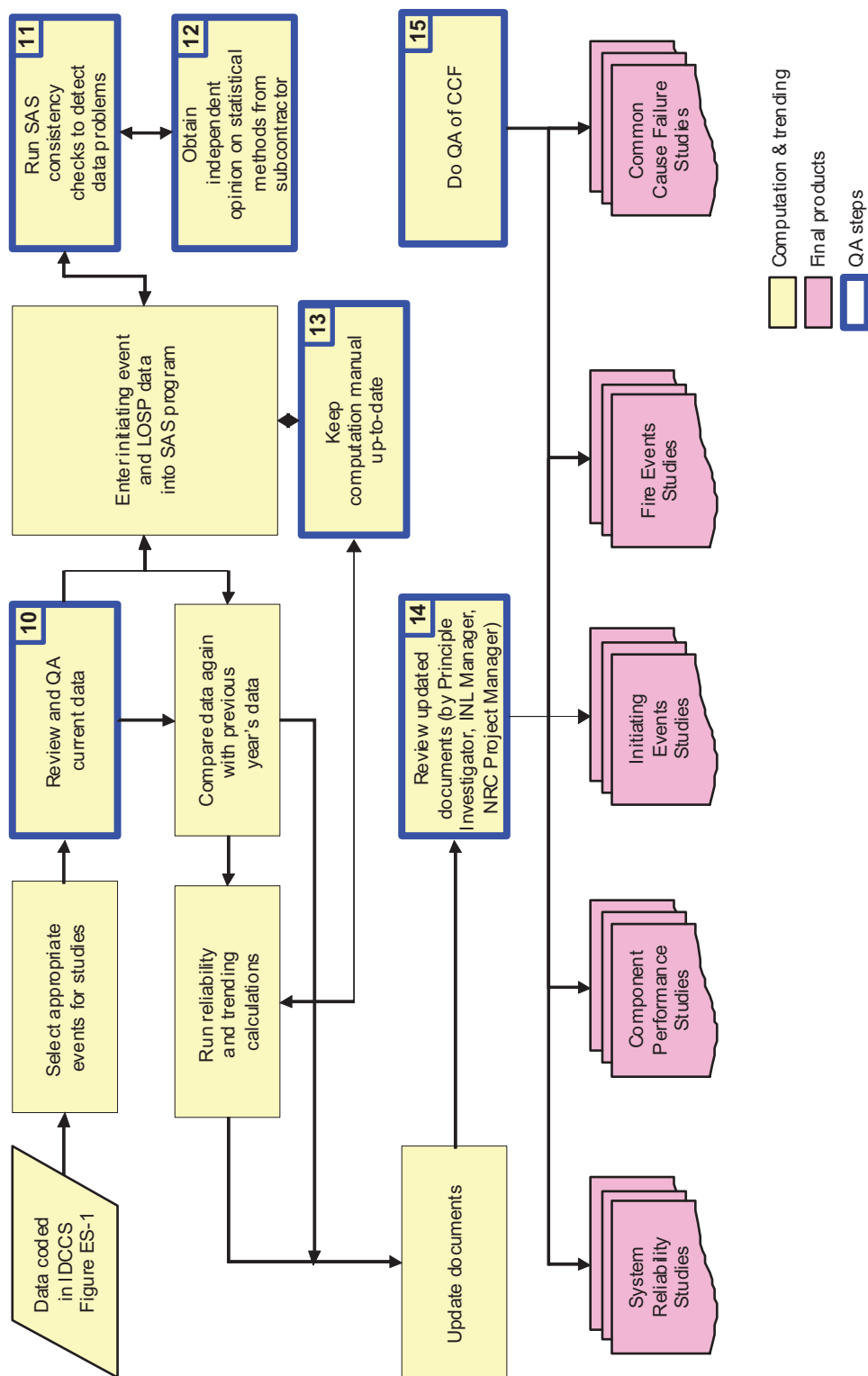


Figure ES- 2. INL data computations and trending process.

The quality controls in the data computation and trending process are

- QA of current data (Block 10)
- SAS program consistency checks (Block 11)
- Independent statistical assistance (Block 12)
- Computation manual (Block 13)
- Update reviews (Block 14)
- QA of common cause failure (CCF) records (Block 15).

## **EPIX Data Quality**

The EPIX database was developed as an improvement over the retired nuclear plant reliability database system database and was implemented in 1997 by INPO. The mission of EPIX is to promote excellence in nuclear power plant equipment performance through the exchange of component-level operating experience. With the advent of the Mitigating Systems Performance Index (MSPI) Program, the EPIX database was restructured to better capture the data needs of that program. Thus, in EPIX one sees a database structure, elements, fields, and nomenclature shaped by MSPI.

The EPIX database is provided to the INL through the NRC sponsor to enable the analysis and reporting of U.S. Commercial nuclear power plant equipment operating performance. The studies and reports created using EPIX data include

1. CCF event data
2. Reliability and Availability Data System (RADS)
3. Component Performance
4. System Reliability
5. Fire
6. *The Industry Average Performance for Components and Initiating Events at U.S. Commercial Nuclear Power Plants.*

The quality of the data used to provide these reports is dependent on the quality of the EPIX database as provided to the INL and the quality of the INL processes and analysis of that EPIX data prior to finalizing the reports.

INPO provides QA of the EPIX data through the use of controlling documents, an automated data entry coaching system, and post entry processing of submitted data. INPO's Consolidated Data Entry (CDE) web site contains pages that guide the licensee in the data entry process for all CDE data elements using the EPIX Coach. The CDE provides INPO with metrics on the completeness and accuracy of the data provided. The EPIX coach is also used, post data entry, to request edits and completeness of the data by the licensee.

The INL closely scrutinizes the EPIX data during the processing for use in NRC applications. A number of issues and concerns are routinely uncovered in the process. While some of the data problems stem from the fact that satisfying all the various NRC data needs was never part of the objectives of the EPIX database, most fall into the categories of completeness and accuracy. Two metrics, the EPIX reporting error rate over time and the reporting uncertainty factor, will become part of the Data Quality Assurance Program Plan.

## **Data Quality Program Plan**

All of the project elements discussed here come together in the development of the Data Quality Program Plan. The current data quality activities were compared to the requirements established earlier. The following activities are the major resulting activities called out in the Data Quality Program Plan:

1. Perform a detailed review of the computer security aspects of the data coding programs and databases to better understand weaknesses and then address weaknesses.
2. Establish guidelines within the Program Plan for when the INL should recommend to the NRC that a peer review be conducted on a specific data product.
3. Establish guidelines within the Program Plan for when the INL should recommend to the NRC that an independent advisory committee review be conducted on a specific data product or data process.
4. Establish an internal Data Quality Committee, with its own charter and schedule for periodic review meetings.
5. Establish guidelines within the Program Plan on the use of editorial reviews, emphasizing the use of plain language in reports and publications (including web sites), and including checks for plain language usage in management reviews.
6. Perform a detailed review of the data checks currently made automatically and manually, catalog them into a checklist and identify additional checks that should be made. Establish the acceptance criteria for these checks.
7. Establish a set of data quality metrics that are measurable and meaningful and permit the evaluation of the general health of the data collection and processing efforts at the INL.
8. Develop a continuous improvement system to formally manage feedback and corrective actions.
9. In close coordination with the NRC, establish an audit program, including INL internal audits and reviews, NRC audits and reviews and external audits by an outside organization.
10. Perform a top-down review of the current anomaly detection methods, evaluate their effectiveness, and address any weaknesses.
11. Establish a data quality assurance/control process for EPIX data based on the research conducted for this project and provide periodic updated failure rate information for use as deemed necessary by the various PRA application programs and projects.

## **Preface**

This report documents the activities performed as part of the NRC's Office of Nuclear Regulatory Research under Job Code Number N6145, Data Quality Assurance. This project was created to

- Identify the standards, guidelines, and good practices applicable to the nuclear operating data collected, coded, and processed for the NRC by the Idaho National Laboratory
- Review and document the data quality assurance activities currently being practiced at the INL
- Compare the current practices against the standards and identify any gaps that may exist
- Recommend additions and changes to the current data quality assurance activities to address these gaps
- Create a Program Plan for Maintaining Data Quality in DRA Programs that can be executed annually to provide assurance that data quality is being maintained.





## Abbreviations, Acronyms, and Initialisms

ACRS	Advisory Committee on Reactor Safeguards
ADAMS	Agency-wide Document Access and Management System
AIChE	American Institute of Chemical Engineers
AOV	air-operated valve
BRIIE	Baseline Risk Index for Initiating Events
CCF	Common Cause Failure
CRGR	Committee to Review Generic Requirements
DRA	Division of Risk Analysis
EDG	emergency diesel generator
EPIX	Equipment Performance and Information Exchange
FTLR	fail-to-load-and-run-for-one-hour
FTR>1H	fail-to-run-beyond-one-hour
IDCCS	Integrated Data Collection and Coding System
INL	Idaho National Laboratory
INPO	Institute of Nuclear Power Operation
ISO	International Organization for Standardization
JCN	Job Code Number
LER	licensee event report
LOSP	loss of offsite power
MDP	motor-driven pump
MOR	monthly operating report
MOV	motor-operated valve
MSPI	Mitigating System Performance Indicator
NISS	National Institute of Statistical Sciences
NPRDS	Nuclear Plant Reliability Data System
NRC	U.S. Nuclear Regulatory Commission
OMB	Office of Management and Budget
PI	performance indicator
PRA	probabilistic risk assessment
QA	quality assurance
QC	quality control
RADS	Reliability and Availability Database System
Reg Guide	Regulatory Guide
RES	Office of Nuclear Regulatory Research

SAPHIRE	Systems Analysis Programs for Hands-on Integrated Reliability Evaluations
SPAR	Standardized Plant Analysis Risk
TDP	turbine-driven pump

# **Program Plan for Maintaining Data Quality in Performance and Reliability Branch Programs**

Since the mid 1980s, the Idaho National Laboratory (INL) has been involved with numerous projects to collect data, develop data tools, and trend and analyze 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. To start, projects relied heavily on independent second person reviews. As work progressed and an occasional data quality problem was identified, the quality assurance (QA) and quality control (QC) processes were reviewed and modified as deemed appropriate. As the databases and computer systems have advanced, many improvements in the data coding and calculation tools have been enhanced by automatically checking for invalid data entries and providing pick lists to standardize data field entries.

Many of the data projects contained small tasks to address data quality concerns as they arose, but none of them have had a significant scope to assess the QA and QC activities and processes. Job Control Number (JCN) N6145, "Data Quality Improvement," is the first NRC project at the INL dedicated to establishing a planned, comprehensive data QA program. This report documents the steps taken in the project, culminating in a Program Plan for Maintaining Data Quality. The steps were

1. Conduct a review of data quality requirements
2. Review current data programs, products, and data QC activities
3. Review the INPO EPIX data quality programs and characterize the EPIX data quality and uncertainty
4. Compare these programs, products, and activities against the requirements
5. Develop a program plan that provides assurance that data quality is being maintained.

The sections that follow describe each of these steps.

It is expected that the Program Plan will be routinely implemented in all aspects of future data collection and processing efforts and that specific portions will be executed annually to provide assurance that data quality is being maintained.



# **1. Data Quality Requirements**

The first step in creating a Program Plan for Maintaining Data Quality in Division of Risk Analysis (DRA) programs was to establish a set of standards, guidelines, and practices to which the Program Plan 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 1 year.

The OMB guidelines serve as the foundation for the NRC Information Quality Guidelines and were reviewed as part of this project 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 1.10.

## **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 DRA has 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 involved in the projects sponsored by DRA.

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 generated by DRA projects 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 the types of data and information generated by



DRA-sponsored projects. 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.

## **1.10 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 will shape the Program Plan. 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.

The next step is to determine the applicability and practicality of these recommendations to the DRA data collection and analysis efforts. Once this set of data quality requirements has been established, the required and recommended practices can be compared to the data quality practices and processes currently in place. Gaps between the requirements and current practices can then be addressed in a comprehensive data QA program.

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.
	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.

Table 1. (continued).

Reference Document	Standard, Guidance, or Practice
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
		1a, 2a, 2h		
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.			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.	1i, 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 into 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

Table 2. (continued).

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





## 2. Current INL Data Quality Control Processes

This section describes the data collection and QA/QC activities that support DRA programs at the INL. The data handling work naturally falls into two distinct categories: (1) data collection and coding, and (2) data computations and trending. The sections that follow describe the processes that occur in each of the two categories.

Figure 1 provides a pictorial overview of the current data processing. The data are collected and coded according to established rules. From there, the data are passed on for computation and trending and the results are published.

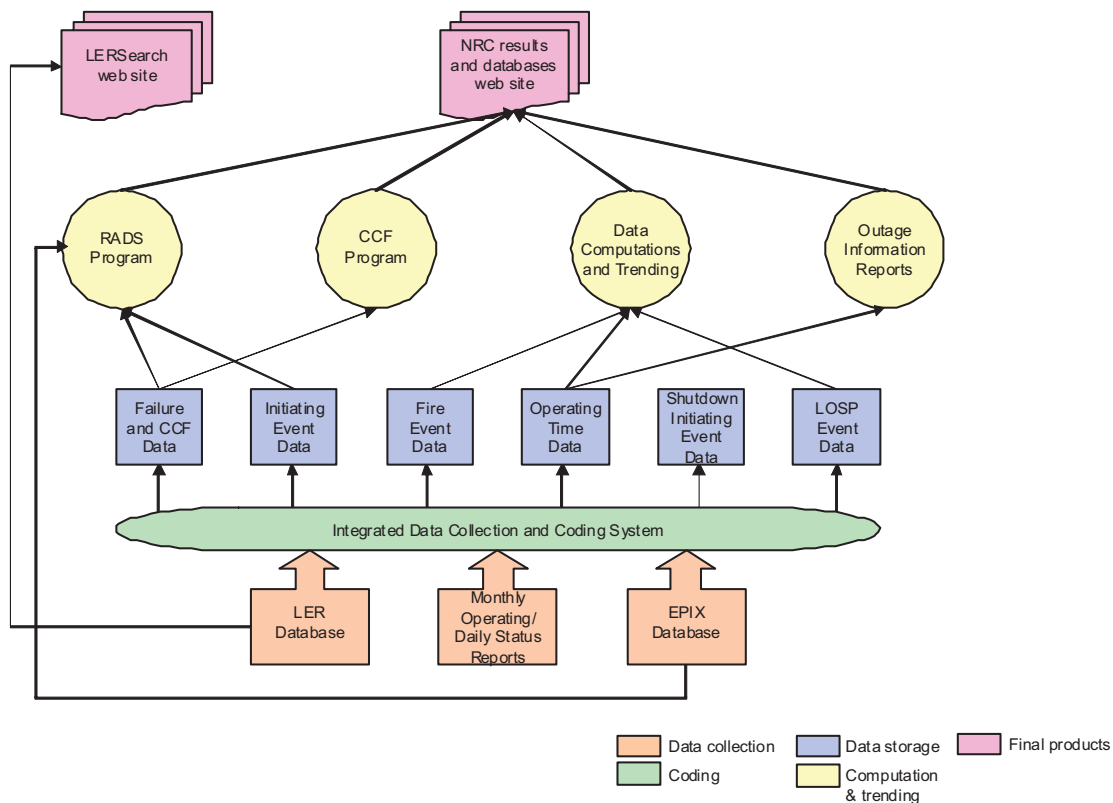


Figure 1. Overview of current data processing.

### 2.1 Data Collection and Coding

#### 2.1.1 Data Sources

As shown in Figure 1, the entire data collection and coding process starts by receiving raw input data. The sources of these data are

- Licensee event reports (LERs)—LERs are downloaded from the NRC's Agency-wide Document Access and Management System (ADAMS) database weekly.
- Equipment Performance and Information Exchange (EPIX)—EPIX data are provided to the NRC by the Institute of Nuclear Power Operations (INPO). The data are forwarded to the INL quarterly.



- Monthly operating reports (MORs) and daily status reports—The MORs are sent to the NRC by each utility and are forwarded to the INL quarterly. The reports contain information about each plant's online hours, critical hours, and shutdown information. The NRC's Emergency Response Center also calls each commercial nuclear power plant daily and obtains their operating status. These daily reports are forwarded to the INL on a monthly basis through the NRC project manager.

### 2.1.2 Quality Controls

The INL LER technician downloads new LERs from ADAMS and inputs them into the LER database weekly. The technician uses LERtrk, a software program created at the INL, for this purpose. A users' guide has been written for LERtrk (Reference 14). The major steps in the data collection and coding process are depicted in Figure 2. Blocks containing steps involving some form of QA/QC activity are numbered in the upper right corner and are outlined in blue.

The quality controls in the data collection and coding process are

- Block 1, LER Completeness Check. The INL staff ensures that all LERs, including LERs submitted by phone under the 50.73 criteria, have been added to the LER tracking database. This ensures that there are no missing data.
- Block 2, Parsed Data Checks. Validation routines in two independent data parsing programs cull data from the LER file downloaded from ADAMS and the file produced by the OmniPage optical character recognition program. The parsing program makes sure the parsed values are valid (for example, the docket must be a 3-digit number and must appear on the list of valid dockets), and compares the values from the two files.
- Block 3, Automated Quality Checks in LERtrk. LERtrk has several automated QA checks to ensure LER data are valid. Here, the program looks for anomalies like invalid docket numbers, missing revisions, invalid dates, incorrect operating mode or power level, and blank fields. Any event reported under 10CFR73.71 is marked safeguards information until verified otherwise.
- Block 4, LER Data/Text Validation. The LER Administrator verifies that the docket number, event date, operating mode, power level, and CFR on the Form 366 match the information given in the narrative section of the LER, and the report date matches (at least approximately) the date of the LER transmittal letter. The Administrator then verifies that information on the form was correctly entered into LERtrk. LERtrk maintains a log of when each batch of LERs was quality checked and by whom.
- Block 5, Periodic Quality Checks in LERtrk. Several computerized QA routines are periodically run on the LERtrk data. The QA programs check for things such as LERs in the database but not on the web page (and vice versa), LERs on the web page with a missing link, empty fields, missing revision or sequence numbers, invalid file names, and invalid LER years.
- Block 6, LER Spot Checks. As time allows, the LER technician pulls up legacy LERs and verifies the data in the database against the data on the LER Form (Form 366).
- Block 7, Data Coding and Automated Validation Checks. LERs and EPIX data records are read and new records are entered in the appropriate databases. The coding guidance and automatic validation checks built into the Integrated Data Collecting and Coding System (IDCCS) provide QA of this initial data entry step. Quality checks for the Common Cause Failure (CCF), Initiating Events, Fire, Operating Time, Shutdown Initiating Events, and Loss of Offsite Power (LOSP) databases are performed similarly. Each of these records is initially coded by a qualified coding engineer using the IDCCS. The IDCCS program uses numerous lookup tables to accomplish the first quality check. The lookup tables relate specific systems and components to each plant. When filling in a new record, these lookup tables will only allow the coder to choose from options that are actually applicable to the plant of interest. If a conflict exists between what the coder believes is the correct entry and what is available in the IDCCS, the coder will conduct additional research on the plant

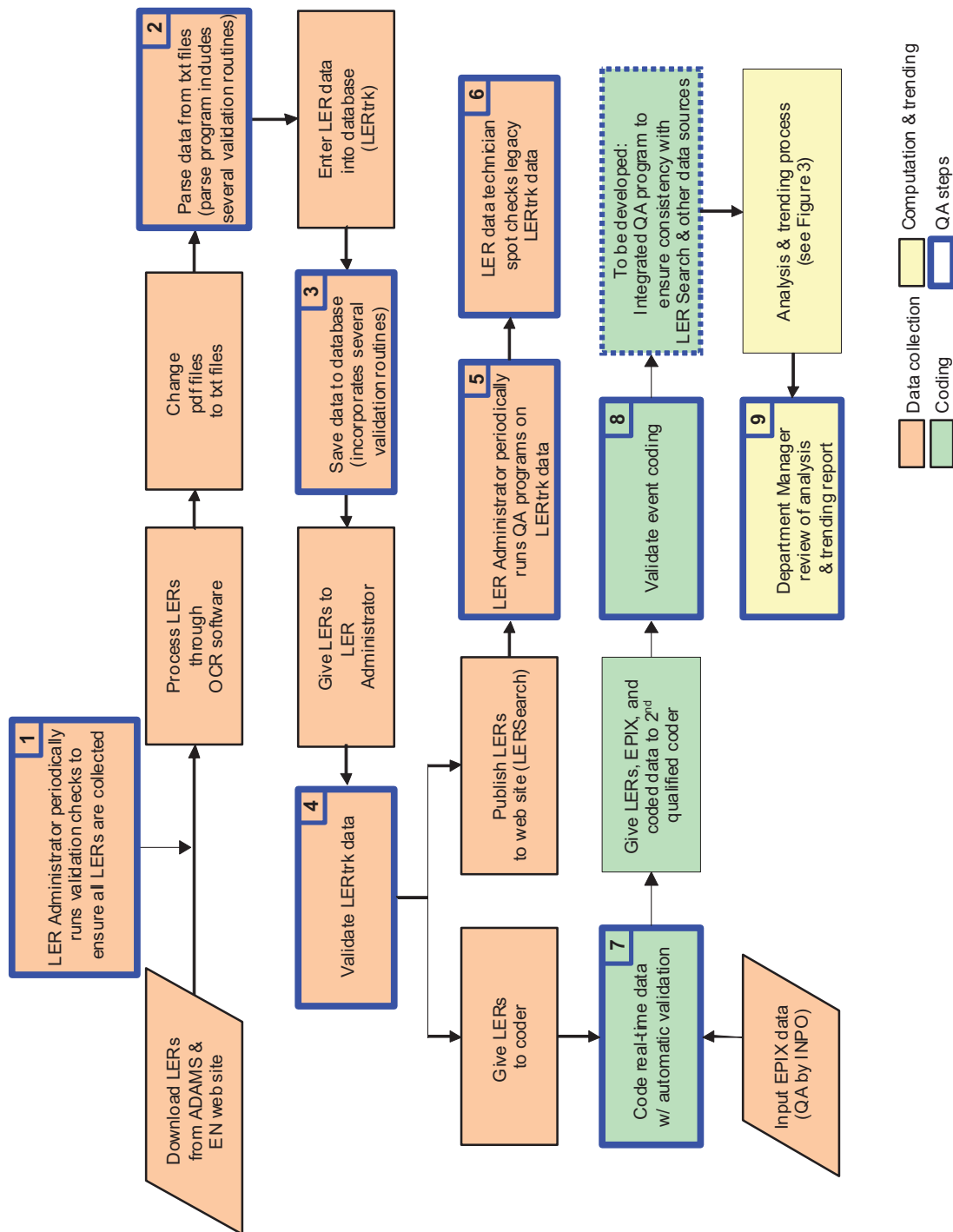


Figure 2. Overview of data collection and processing.

configuration, get a peer check on the analysis, and modify the lookup table with the correct data. Several real-time quality checks are currently built into the IDCCS program to ensure all required data fields are filled in; more computerized QA checks are planned for development.

- Block 8, Coding Second Checks. The most important quality check performed on entries in the IDCCS program is performed by a second qualified coding engineer. The IDCCS program tracks which records require a review and allow the reviewer to filter those records. The second checker reads the entire source record and verifies all applicable studies have been coded and that the coding is accurate. Good communication between the initial coder and the second checker ensures that more difficult events are properly coded. Either the initial coder or the second checker will seek additional guidance if they encounter an issue which is outside their expertise. The data are further reviewed when evaluated for trends by the programs that use the data.
- Block 9, INL Manager Review. The INL manager provides a third level of review on both the data collection and coding and the data computation and trending. This review selects a subset of events and computations and verifies correctness. This review is documented in a transmittal letter to the NRC Project Manager.
- Additional Miscellaneous Checks.
  - The Plant Outage Information (OUTINFO) database is validated during the quarterly performance indicator (PI) reporting process. The programs that create the PI report have several built-in quality checks. These checks include comparing OUTINFO data to LER data for operating mode and power level, comparing OUTINFO data to MOR data, and several other similar checks. When discrepancies are found between the various data sources, they are researched further and the utility may even be contacted to ensure the correct information is recorded.
  - The MOR database is quality checked with an automated program. This program checks for such things as reactor critical hours > generator online hours, online hours + shutdown hours = number of hours in the month, and missing records checks. This automated QA program is run quarterly.
  - The MOR files are also manually quality checked and compared against the OUTINFO reports and both are compared to event notifications and LERs for forced outage start times and reactor criticality times.

### **2.1.3 Products**

The data collection and coding process produces the following products:

- IDCCS—This system contains detailed event information on plant and system performance (Reference 15). This information consists mainly of system demand and failure information; the level of detail goes down to the component level. Events that contain data relevant to RES risk studies (approximately 40% of all LERs) are coded.
- LERSearch—The INL maintains a web page called LERSearch for the NRC that allows access to the LER database (Reference 16). LERSearch facilitates searches on the LER header fields, abstract, corresponding narrative, and coding information from the IDCCS. The database currently contains LERs from 1980 to the present.
- Failure and CCF database—The Failure database contains failure records for components of interest. The database is Microsoft Access-based and maintained at the INL. The data sources for this database are LERs and EPIX data, which are coded and QA checked by INL personnel. The CCF database is created using data coded into the Failure database and contains records for groups of failures for which a CCF mode was identified. These data are coded using the IDCCS.

- Initiating Events database—The Initiating Events database is an Access database that contains records of industry initiating events (events which result in an automatic or manual reactor scram). This database is used to determine historical frequencies for the occurrence of initiating events. These records are coded and quality checked by INL coding engineers from LER records using the IDCCS.
- Fire Events database—The Fire Events database is an Access database of events involving fires. LERs and EPIX data are the primary sources of data, with some data coming from National Electrical Insurers Limited. These data are coded using the IDCCS.
- Operating Time Data consists of two databases:
  - Plant Outage Information database—The INL maintains a database (OUTINFO) of commercial nuclear power plant daily operating status. The database captures all scheduled and forced plant outages based on reactor criticality. The data source for this database is the daily status reports collected by the NRC’s Emergency Response Center and transmitted to the INL via the NRC project manager.
  - MOR database—The INL maintains an access database of all operating and shutdown periods for each plant. Information in this database includes reactor critical hours, generator on-line hours, forced outage hours, and capacity factors. This information is used in availability and reliability calculations for other projects at the INL. The information is entered into the database from data files submitted by each plant monthly.
- Shutdown Initiating Events database—These data are collected and stored in a database to support Low Power and Shutdown Standardized Plant Analysis Risk (SPAR) model development at the INL. The data source for the shutdown initiating events is primarily LERs. Shutdown initiating events are defined as events in which a plant was in a shutdown mode and the normal mode of decay heat removal was lost. The data are coded using the IDCCS.
- LOSP database—The LOSP database maintains records of loss of offsite power events. The records are derived from LER records and are coded using the IDCCS.

## **2.2 Data Computations and Trending Process**

### **2.2.1 Data Sources**

The primary source of data used in the computations and trending work is the output from the IDCCS as described above. The second input into the process is the SPAR models. The SPAR models provide the framework for determining the system unreliabilities and unavailabilities trended in the System Studies. The SPAR models undergo their own QA reviews and checks (Reference 17) and are accepted by this process as quality products. Additionally, the SAPHIRE risk assessment program is used to perform the system unreliability and unavailability computations. SAPHIRE is also put through a separate QA program (References 18 and 19) and is accepted by this process.

### **2.2.2 Quality Controls**

Figure 3 shows the flow chart that describes the data computation and trending process. As in Figure 2, the steps involving some form of QA/QC activity are numbered in the upper right corner and are outlined in blue.

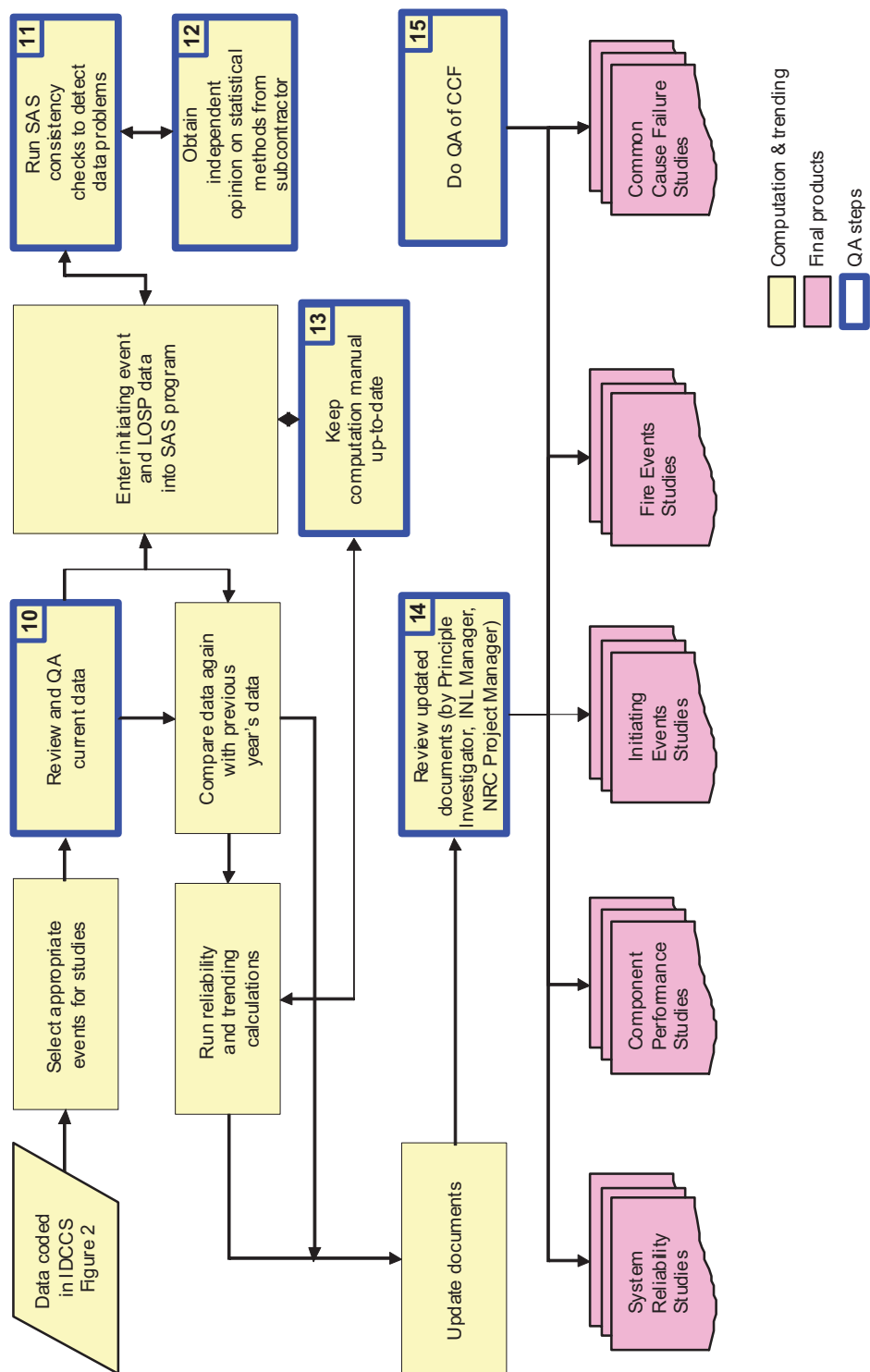


Figure 3. Overview of computations and trending.

The quality controls in the data computation and trending process are

- Block 10, Review and QA of Current Data. Each year, before running data computations, the INL principal investigator reviews a sample of the events that were coded since the last update. This review verifies that the coding guidance was properly used during both the initial coding and the second verification in the data collection process. Any discrepancies are discussed with the coding team until agreement is achieved. This also serves to train the coding staff. Any needed enhancements to the coding guidance are identified and implemented.
- Block 11, SAS Program Consistency Checks. The initiating event and LOSP databases are read into the commercially available SAS program each year for the computation of frequencies and trends. As the data are read into SAS, several data consistency checks are performed:
  - Each input data set is checked for consistency with the particular plants under study and the operating periods associated with the plants (e.g., the low power license and decommission dates)
  - Plants lacking data are noted
  - Unexpected content, if any, in the fields of the records applicable to the unreliability calculations is flagged for further review
  - The SAS logs are checked for instances of missing values or other abnormalities.
- Block 12, Independent Statistical Assistance. The INL maintains a contract with an independent statistician to provide assistance for the INL staff on statistical methods and procedures.
- Block 13, Computation Manual. The Computation Manual (Reference 20) is maintained by the INL staff as a living document. Any changes to the methodology or programs are documented as the changes are made. In addition, the manual and SAS code are reviewed for editorial changes. This forces the staff to review the methods and flow for inconsistencies and to fix or change as necessary.
- Block 14, Current Updates are Reviewed. The draft update documents are reviewed by the INL principal investigator, the INL manager, and by the NRC program manager. The INL principal investigator reviews the documents for accuracy with the SAS output files, format, and results. The INL manager reviews the documents for consistency with previous updates, format, and results and spot checks the data collection and statistical results. The NRC program manager checks the documents for consistency with previous updates, format, and results.
- Block 15, QA of CCF Records.
  - Quarterly, the INL staff compiles the CCF events and creates the CCF records in accordance with the coding guidelines. The CCF data are reviewed for adherence to the CCF coding guidelines as a third INL check. The CCF data consist of the independent records and the CCF event records. All CCF records are reviewed and verified. The independent records are spot checked.
  - NRC CCF Review. The NRC technical monitor in charge of CCF reviews all of the CCF records for compliance with the coding guidelines and forwards any comments to the INL principal investigator for resolution.
- External CCF Review. Currently, the Westinghouse Owners Group is providing a 100% review of CCF events. Any comments are resolved through the NRC technical monitor by the INL principal investigator.

### **2.2.3 Products**

The data computations and trending process produces the following products:

- *Industry-Average Performance for Components and Initiating Events at U.S. Commercial Nuclear Power Plants* data report (Reference 21). This report documented work performed to develop



industry average component and initiating event basic event parameters that represent current industry performance in support of these risk-informed regulatory activities. The report documents the parameter estimation process to update the basic event failure probabilities and failure rates, maintenance unavailabilities, and initiating event frequencies for the Level 1 SPAR models. Initiating event data ending in 2002 (with various start dates, depending upon the initiator) were used to generate frequency distributions for 18 of 24 initiating event categories found in SPAR (excluding interfacing system loss-of-coolant initiators). LOSP and most loss-of-coolant accident frequencies were developed under separate studies (References 22 and 23). Results from those studies are included in this report. Results for 49 component types and 153 component type and failure mode combinations are also documented. Additionally, component unavailability data, representing test and maintenance basic events, were collected. The component unreliability and unavailability data generally span a 5-year period from 1998 through 2002. Comparisons of the results with other current sources and historical estimates were also conducted. Data previously used in these models typically reflected performance at U.S. commercial nuclear power plants during the early 1990s, the 1980s, and in some cases, even earlier. The improved SPAR models, and their enhanced use for support of risk-informed regulatory activities, require up-to-date data that are more representative of current industry performance.

- **Baseline Risk Index for Initiating Events (BRIIE).** This report (Reference 24) summarizes an effort to enhance the NRC Industry Trends Program in the area of initiating events. The BRIIE includes two tiers. Tier 1 involves the monitoring of individual initiating events at the industry level against performance-based prediction limits, while Tier 2 involves an integrated, risk-informed indicator at the industry level that combines the risks from individual initiating events. Technical bases for decisions made in the development of the BRIIE are also presented, as well as historical results at both the Tier 1 and Tier 2 levels.
- **System Studies.** System performance evaluations (References 25–31) have been made for the following safety systems: High Pressure Coolant Injection, High Pressure Core Spray, Isolation Condenser, Reactor Core Isolation Cooling, Auxiliary Feedwater, High Pressure Safety Injection, Emergency Diesel Generator Power, and Reactor Protection. Each evaluation was based on reactor operating experience contained primarily in LERs. These studies have undergone a transition from the reliability data based on LERs to reliability data based on EPIX. In addition, the system studies are now based on the fault trees in SPAR and the system reliability calculations use plant-specific models rather than the generic models previously used. These evaluations look at system unreliability and unavailability and trend these attributes over time. Annual updates of these studies are reported on an NRC web page (Reference 32).
- **Component Performance Studies.** The original analysis of the performance of safety-related components used in both pressurized water and boiling water reactors in risk-important systems is documented in References 33–36. Annual updates of these studies using only EPIX data are reported on an NRC web page (Reference 32).
- **CCFs.** Equipment failures that contribute to CCF events are identified during searches of LERs, Nuclear Plant Reliability Data System (NPRDS) failure reports, and EPIX failure reports. The quantitative results are presented as both alpha factors and multiple Greek letter parameter estimations. The alpha factor distributions are also presented. Insights studies were performed on the set of CCFs of emergency diesel generators, motor-operated valves, motor-driven pumps, and circuit breakers. The data studied were derived from the NRC CCF database, which is based on U.S. commercial nuclear reactor power plant data. These reports are the result of in-depth reviews of the CCF data for the subject components. The objective of these reports are to look beyond the CCF parameter estimates to gain further understanding of why CCF events occur and what measures may be taken to prevent, or at least mitigate, the effect of these CCF events. The insight studies are documented in NUREG/CR-6819 (Reference 37).

- As an extension of the CCF work, the NRC (and INL) supports the International Common-Cause Data Exchange Project. Its purpose is to allow multiple countries to collaborate and exchange CCF data to enhance the quality of risk analyses that include CCF modeling. Because CCF events are typically rare events, most countries do not experience enough CCF events to perform meaningful analyses. Data combined from several countries, however, yields sufficient data for more rigorous analyses.
- Fire Events. The original fire study (Reference 38) uses operating experience to characterize the frequency and nature of fire event data from operating U.S. commercial nuclear power plants during 1987 to 1997. This report presents an analysis of fire event frequencies at U.S. nuclear power plants. The evaluation is based on the operating experience from fiscal year 1987 through 2005. The data sources for this report include
  - LERs, 1987 to 2005
  - NPRDS, 1987 to 1996
  - EPIX, 1997 to 2005
  - Electric Power Research Institute, 1986 to 1988
  - National Electric Insurers Limited, 1993 to 2003.Annual updates of these studies are reported on an NRC web page (Reference 32).





### 3. EPIX Quality Measurement

The EPIX database has been available to the INL since approximately 1997. The first unreviewed use of EPIX data was to generate the component failure rates presented in the journal article “Historical Perspective on Failure Rates for U.S. Commercial Reactor Components” (Reference 39). Those failure rates provide the component unreliability baselines used in the Mitigating Systems Performance Index (MSPI) (Reference 40). The second major use of unreviewed EPIX data was to generate the component failure rates documented in NUREG/CR-6928 (Reference 21). That document provides baseline component failure rates for use in NRC programs and industry risk models. Unreviewed EPIX data are currently loaded into the Reliability and Availability Database System (RADS) software. Results from RADS are being used to support the SPAR models and by the NRC for various reasons.

#### 3.1 EPIX Data Background

EPIX data have been considered as somewhat suspect at the individual plant and event level. If one desired to look at a specific component’s performance at a given plant, the results from EPIX/RADS were not considered completely reliable. This was the result of several factors:

- A significant fraction of EPIX events were assigned failure modes in that database that if reviewed in detail might more appropriately have been assigned a different failure mode. In addition, the RADS software used a mapping scheme to assign PRA failure modes to the EPIX failure descriptions. That mapping scheme introduced additional uncertainty in the failure modes.
- A very limited review of failure events in EPIX with known failures from LERs indicated that some failures events were missing from EPIX.
- Some EPIX entries for numbers of demands or run hours (either estimates for a several-year period or actual counts) appeared to be incorrect. An example is emergency diesel generators (EDGs) with more load and run demands than start demands.

In spite of these deficiencies, the EPIX/RADS results were considered to be appropriate for the generation of industry-average performance estimates. At the industry level, the limitations listed above tended to be masked. This was verified by reviewing failure event descriptions for selected components such as EDGs and turbine-driven pumps (TDPs). For example, a detailed review of EDG failures indicated up to 20 or 30% of the fail-to-start events should have been coded as fail-to-load-and-run-for-one-hour (FTLR), fail-to-run-beyond-one-hour (FTR>1H), or as a no failure event. However, some of the FTLR, FTR>1H, and no failure events should have been coded as fail-to-start.

#### 3.2 EPIX Data Quality Measurement Program

In the INL processing of EPIX data, the data are either (1) accepted and used as provided by INPO or (2) recoded by INL staff. A two-tiered EPIX quality metric is used. The first tier quantifies the uncertainty of unreviewed EPIX data using the INPO requirements as the standard and using the unreviewed EPIX coding as the initial state and the INL staff review as the current state. This metric assumes that the INL review and coding is the optimal correctness of the data, which is assessed in Tier 2.

The second tier is a quality check based on the sampling of a portion of those EPIX (and/or LERs where applicable) coded events that identify an error rate for the INL data coding personnel.

##### 3.2.1 Tier 1 EPIX Data Quality Measurement

Tier 1 characterizes the completeness of the raw EPIX data and the uncertainty between raw EPIX data and the interpretation of the EPIX coding guidelines by INL staff. To support current NRC programs, all MSPI devices are evaluated and will continue to be evaluated by the INL staff. This creates a 100% sample of MSPI device failure data. There are many non-MSPI devices that are reviewed 100%

by the INL staff primarily to support the CCF program. However, the remaining non-MSPI devices are never reviewed by the INL staff and there are no plans to start such an on-going review (e.g., motor-driven compressors, fans, and air-handling units).

The current INL activities establish corrected system, component, failure mode, and detection parameters through the CCF, component performance, and MSPI support programs for selected MSPI and non-MSPI devices.

#### **3.2.1.1 EPIX Failure Data Completeness and Accuracy**

The completeness data element of the IDCCS characterizes the completeness of each EPIX failure data record, reflecting the additional effort required to determine the appropriate system, component, failure mode, and detection parameters. This data element is collected for each EPIX failure record that is reviewed by the INL staff at the same time as the data collection takes place.

The completeness data check was implemented in January 2009 and is included in the MSPI data check project (which includes the failures of MSPI devices since the beginning of 2003). In addition, any data collection occurring as a part of the on-going data collection program is evaluated to the completeness criteria. In the future, a sample of non-MSPI devices that are not currently scheduled to be reviewed by the INL staff will be reviewed using the same criteria to provide insights to the completeness of MSPI versus non-MSPI EPIX data entries.

The missing data issue is not measurable using a data collection effort. Currently, the qualitative analysis arguments are the only ones we have. Plants that under- or over-report failures should be removed from the pools of data analyzed for industry trends.

#### **3.2.1.2 EPIX Data Review**

The EPIX data review element focuses on the failure mode assigned to the reported key component. This is the area of the highest uncertainty and lends itself to comparisons between the raw data and the interpreted INL data. The failure mode that is assigned to the component determines whether the event is counted at all (if it is categorized as Unavailable) or whether it is counted in its appropriate failure mode.

### **3.2.2 Tier 2 INL Data Quality Measurement**

Each data record coded by the INL is marked with the names of the staff members that performed the initial coding and second check. Any discrepancies noted by the second checker are confirmed by the initial coder and an agreement is reached. This process is carried out for 100% of the data coded by INL personnel. To measure the error rate after this two-person agreement, a third review step is implemented for a 10% sample of all data coded by INL personnel (including LER-based data). The primary purpose of this third QA check is to create the measurement. The error rate is assessed and appropriate statistics developed. In addition, errors are discussed by staff, corrected, and avoided in future efforts.

Error rates are trended; any undesirable deviation from the baseline error rate is evaluated and steps are taken to correct the problem(s). At the onset of the Tier 2 QA check, a 10% error rate is assumed to be the baseline. Any error rate greater than 10% requires larger samples until problems are resolved and observed error rates decrease to less than 10%.

The metric chosen to assess the quality of the INL coding is two-fold: (1) the results of the quantitative assessment, which includes any part of the analysis that is counted to produce the estimate, are rated on a scale of 1 to 5, and (2) the results of the qualitative assessment, which includes any part of the analysis that contributes to qualitative analyses, are also rated on a scale of 1 to 5.

In the reporting for the Tier 2 measurement, any results not evaluating to a perfect 5 and 5 need to have comments added to describe the shortcoming(s). In addition, the deficient record is corrected and the staff is trained on the type of error found.

### 3.2.3 Product Dependency

The programs and studies provided to the NRC have different dependencies to the two tiers of data quality. Table 3 shows these dependencies.

*Table 3. NRC product dependency matrix.*

Program	Tier	Description
CCF	2	All CCF records are coded by INL staff; the data quality measurement is the QA of coded records.
Component Performance (EDGs and TDPs)	2	All EDG and TDP component performance records are coded by INL staff; the data quality measurement is the QA of coded records and is dependent on the Tier 2 data quality measurements.
Component Performance (MDPs, MOVs, and AOVs)	1 and 2	<p>The MSPI designated component performance records for motor-driven pumps (MDPs), motor-operated valves (MOVs), and air-operated valves (AOVs) are coded by INL staff; the data quality measurement is the QA of coded records and is dependent on the Tier 2 data quality measurements.</p> <p>The non-MSPI designated component performance records for MDPs, MOVs, and AOVs are dependent on the Tier 1 data quality measurements.</p>
System Reliability	1 and 2	The system reliability analyses use both MSPI and non-MSPI component performance data and data from Reference 21.
Fire	2	EPIX was never intended to provide data on fire events. The extraction of fire data from EPIX is fully the function of the INL data collection effort.
RADS	1 and 2	<p>All EDG and TDP component performance records are coded by INL staff; the data quality measurement is the QA of coded records and is dependent on the Tier 2 data quality measurements.</p> <p>The MSPI designated component records for MDP, MOV, and AOV are coded by INL staff; the data quality measurement is the QA of coded records and is dependent on the Tier 2 data quality measurements.</p> <p>The non-MSPI designated component for MDP, MOV, and AOV performance records are dependent on the Tier 1 data quality measurements.</p>

### 3.2.4 Statistical Evaluation of EPIX Data Uncertainty

Any PRA parameter estimate produced by the INL for use in NRC programs is accompanied by an uncertainty distribution based on the distribution of the data as counted and summed from EPIX. The

uncertainty distribution can be determined by Bayesian methods if sufficient data exist or is presented as the result of the constrained non-informative prior distribution. Incorporating the observed uncertainty in the coding of the data requires simulation techniques to arrive at the combined uncertainty due to both data scatter and data uncertainty.

## **4. Comparison of Current Practices against Standards**

After establishing a set of data quality standards, guidelines, and recommended practices and determining the current status of data QA and QC activities, the next step in developing the Program Plan for Maintaining Data Quality was to compare the current practices against the standards to identify areas of strength and areas for improvement. This section documents that comparison for each of the 22 consolidated standards, guidance, and practices listed in Table 2.

### **4.1 Graded Approach to Data Quality**

Standard 1 from Table 2 reads:

“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.”

The data collected and processed for DRA serves a wide range of uses. Data used in PRA applications, such as component and failure mode failure rates, initiating event frequencies, and CCF parameters, may ultimately be used as part of a regulatory analysis; therefore, the data should be subjected to the most rigorous standards, including independent peer reviews and public comment. Other data, provided for information purposes and background, are subjected to less rigorous treatment.

The INL data QA/QC processes make distinctions among the various data types regarding the types of QA/QC activities performed and the level of detail involved. Low-level checks and automated checks are employed for less critical data, such as plant power level, plant operating modes, and other general information associated with LER data. More important data are subjected to second coder checks, periodic INL and NRC management reviews, and final product reviews. The most influential data, the end-user projects, typically receive the greatest level of scrutiny. For example, initiating event frequency and component failure rate data used in the SPAR risk models undergo engineer reviews by the SPAR project team before going into the models. Additional reviews are provided by the licensees in the plant visits made before the models are finalized and approved for use. The SPAR project has undergone a number of ACRS reviews to ensure conformance with the general best practices and standards of the PRA community. Lastly, each application of the SPAR models to a specific issue subjects all aspects of the models and data to close scrutiny.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

### **4.2 Dissemination to the Public**

Standard 2 reads:

“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.”

The INL data products for DRA are disseminated primarily in three ways: (1) letter reports, (2) NUREG/CR reports, and (3) NRC web sites. Only the NUREG/CR reports are intended to be part of the public dissemination process. The creation and submittal of these reports follow the standard NRC processes and formats. The determination of the degree of public review and comment is made by the appropriate NRC staff. The letter reports and internal NRC web site postings are intended for NRC staff use only. If the NRC determined that information contained in these media should be externally released, the NRC would either direct the INL to provide the information through one of the standard public dissemination processes or would do so themselves.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

### **4.3 Computerized Database Systems**

Standard 3 reads:

“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.”

All INL data collection and coding activities use computerized systems. These systems are part of the INL information technology (IT) infrastructure and are governed by the same high standards required by the U.S. Department of Energy for all other computer systems at the INL. Access to these systems and privileges granted once in the systems are controlled by the INL IT System Administrator and the INL Project Manager and authorized by the Department Manager. Password protection of files and databases is employed at various levels. The data coding software requires a user login and stamps each record with the individual that created or modified the record. While access to the areas of the servers where the database files reside is limited to only a small group of INL staff working on NRC projects, further security measures to protect individual database files from inadvertent erasure or modification should be investigated.

Every INL employee is screened during the hiring process. Every individual assigned to an NRC project is approved by the INL Department Manager and the NRC Project Manager. The servers on which the databases reside are properly protected from unauthorized access. These servers are backed up on a daily basis.

The INL does not possess or handle Agency records. The INL maintains its own project records. These records are handled in accordance with the INL Records Management system under the supervision of the Department Records Coordinator.

In general, the current INL data QA/QC processes and practices are evaluated as adequately meeting this standard. The only exception noted is the potential for tighter security of individual database files.

### **4.4 Formal Review and Concurrence**

Standard 4 reads:

“Provide formal review of and concurrence with all information disseminated (including technical reports). Use appropriate level of management review and approval in concurrence process.”



All major data deliverables from the INL undergo INL Project Manager and Department Manager reviews and NRC Technical Monitor reviews. The INL Department Manager reviews are documented in official letters from the reviewing manager to the NRC Branch Chief and the appropriate NRC Project Manager and Technical Monitor. The INL Department Manager is the direct supervisor over the project managers, principal investigators, and technical staff performing the work. He also serves as the INL funds manager, directly responsible to the local U.S. Department of Energy Representative for NRC projects for the fulfillment of all project obligations.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.5 Peer Review of NRC Research Products**

Standard 5 reads:

“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.”

The extent of the use of peer reviews is at the discretion of the NRC Project Manager and Technical Monitor and is included either as a task in our assigned work scope or is handled under separate contract managed by the NRC. Historically, specific peer reviews of data have been performed by nuclear risk industry consultants in the areas of CCFs and statistical methods. The degree of independence of these peer reviewers has varied from subject to subject and as time passes. Currently, the Westinghouse Owners Group is acting as an independent peer review for the area of CCFs. Dr. Cory Atwood fills a review function in the area of statistical methods; however, he has been involved in actual production work as an author of several NUREG/CR reports so his level of independence is questionable. Several other projects and applications using the INL data products also employ peer reviews where the associated data and data analysis methods would be within the scope of the review.

The current INL data QA/QC processes and practices will be enhanced to include an action to evaluate data products for the potential benefits of having a peer review. Recommendations will be made to the NRC project leadership, as appropriate.

## **4.6 Independent Advisory Committee Reviews**

Standard 6 reads:

“Use reviews of agency information by independent advisory committees, as appropriate, including the ACRS. Use internal review groups, like CRGR, as appropriate.”

As with peer reviews, the extent of the use of independent advisory committee reviews is at the discretion of the NRC. The INL has been tasked on a number of occasions to support presentations to the ACRS on various aspects of the work performed at the INL; however, it has been many years since any of the aspects of the data collection, coding, computations, and trending work has been reviewed by an independent advisory committee. Projects and programs that are end-users of the data, such as SPAR, are routinely subjected to ACRS review.

The current INL data QA/QC processes and practices will be enhanced to include an action to evaluate data products for the potential benefits of having an independent advisory committee review. Recommendations will be made to the NRC project leadership, as appropriate.



## **4.7 Feedback from Agreement States**

Standard 7 reads:

“Obtain early and substantial feedback from the Agreement States.”

Typically the Agreement States do not have any specific roles in the data collection, coding, computations, and trending processes performed for DRA. All data are provided by the licensees directly to the NRC or to the NRC via INPO (EPIX data). The Agreement States may be users of the information that is publicly disseminated and have access to the NRC processes for providing public feedback.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.8 Quality Review Meetings**

Standard 8 reads:

“Quality review meetings should occur on a periodic basis.”

The INL data projects do not schedule routine periodic quality review meetings. There is daily interaction among the data coders and those providing the data computations and trends. Ad hoc meetings are called whenever an issue arises that needs to be addressed. With the implementation of a Program Plan for Maintaining Data Quality in DRA Programs, an annual quality review meeting would be appropriate. The meeting could be the conduit to executing the annual requirements of the Program Plan and a forum for discussing the findings and promoting continuous improvement.

The current INL QA/QC processes and practices will be enhanced to include an annual quality review meeting to meet this standard.

## **4.9 Sound Statistical and Research Methods**

Standard 9 reads:

“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.)”

The statistical methods employed in the computations and trending activities have been developed for more than 15 years while such work has been conducted at the INL for the NRC. Early in the projects, experts such as Dr. Cory Atwood, Dr. Ali Mosleh, Mr. Dennis Bley, Mr. Karl Fleming, and Dr. Dale Rasmuson provided direction for the statistical methods used in the creation of the various data products. Some of these same experts are still being consulted as the databases and end-user products evolve.

Data products that could be used to support regulatory analysis (especially initiating event frequencies, component failure rates, and CCF parameters) conform to the guidelines provided in NUREG/BR-0058 and NUREG/BR-0184, in that not only are point estimate values determined, but uncertainty parameters are calculated and published whenever possible. The regulatory analysis guidelines also recommend that the sources of uncertainty be discussed.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.10 Descriptions of Data and Methods**

Standard 10 reads:

“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.”

The data collection and coding guidance is developed for each of the end uses as they are approved and added to the project work scope. The coding guidance is listed in the appropriate reports (e.g., NUREG/CR-6268 for CCF). An independent reviewer should be able to duplicate the coding results using the coding guidance. The same holds true for the computations, calculations, and trending analyses performed using the data. The statistical methods are documented in the product reports with the intent of providing sufficient detail to allow reproduction by others.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.11 Editorial Review**

Standard 11 reads:

“Ensure that all disseminated information receives appropriate editorial review. Adhere to the agency’s Plain Language Program in written and electronic products.”

The current INL data QA/QC processes and practices do not address the topics of editorial reviews and use of plain language in documents and products. Editorial reviews are conducted as time and funding permit. The use of plain language has not been emphasized. The INL data QA/QC processes and practices will be enhanced to provide guidance on when to employ editorial reviews, emphasize the use of plain language in reports and publications of all sorts, and check for plain language usage during management reviews.

## **4.12 Written Procedures and Instructions**

Standard 12 reads:

“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. All deviations should be formally documented and submitted in a timely manner. 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.”

The purpose of each data collection effort, the means of data collection, the data needed, and the treatment of that data are clearly defined in the NRC Form 189 written for each project. Typically, the first task in each project is to write a project work plan and obtain NRC approval of the plan before proceeding further. The data collection and coding efforts have written procedures that explain the goals and objectives of the data collection activity, define the data fields in the database structure, provide in detail the methods and steps to be taken, and provide guidance on interpretation of raw data for coding purposes. Deviations from the written guidance are annotated in the records of concern.

The computations and trending activities are documented in a computation manual. This manual describes the goals and objectives of the computations and trending activities, defines the inputs and outputs of the computations, and documents the statistical methods and algorithms used. This manual is updated whenever changes are made to the existing computations or new treatments (including ad hoc) are developed.

The written materials just described are internal INL documents intended for the use of INL project staff, but they also serve as resource materials for the descriptions of the databases and calculation methods provided in the data products submitted to the NRC.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

### **4.13 Checking Procedures and Data Validity**

Standard 13 reads:

“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.”

Section 2 discusses some of the checks used to ensure validity. All data entered into the IDCCS is second checked by another engineer. Only data that have been through the second checks are permanently made part of the databases. The databases themselves have some validation built into them. For example, the use of lookup tables and pick lists limits the coding to predefined values and ensures required fields are filled in before allowing the record to be saved. A number of checks are made to ensure the data is not in violation of physical laws or logic. There are checks performed to ensure that all data in a series (e.g., LER numbers are continuous without gaps) are present or annotated as to why gaps exist.

This is an area that is constantly undergoing improvement. The new IDCCS program will have additional checks in place. The current INL data QA/QC processes and practices do not have a formal list of items to check, nor acceptance criteria for an internal verification of data quality.

### **4.14 Incorporating Data Quality from the Start**

Standard 14 reads:

“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.”

The INL has been involved with the collection, coding, analysis, and trending of nuclear industry operational data for over 20 years. In that time, the INL has consistently delivered high quality data products. Over 16,000 LERs and EPIX records have been collected and coded, resulting in hundreds of thousands of individual data field entries. Naturally, with that volume of data, there have been a number of coding errors that have made it through the QA/QC processes and were discovered in the final data products. However, that number is very small, on the order of one or two a year.

The NRC processes for establishing a data collection and coding project help to define the mission of the effort, the task’s purpose and expectations, the expected outputs and deliverables, and the breadth and depth of the scope of the database. The project work plan developed as the first step is establishing a new data collection and coding task clearly defines the data structure, identifies the data inputs and outputs, stipulates the data coding rules and procedures, and addresses the data quality activities and acceptance criteria.

The Principal Investigator for the Reactor Operating Experience for Risk Applications project has the responsibility of being the database administrator, although that role has not been clearly stated in the past. Responsibility for some of the smaller databases may be assigned to assistants for the day-to-day administration and upkeep; however, the project Principal Investigator has overall responsibility for the databases within the project.

There are several areas in this standard where improvements may be made. First, our work processes do not address data quality metrics and communication of quality system results. Second, our QA/QC processes do not include an evaluation of how well the data collection objectives were met. And lastly, there is no explicit continuous improvement process in place to formally handle corrective actions and provide feedback into our data efforts. These shortcomings will be addressed later in this Program Plan.

## **4.15 Data Accessibility**

Standard 15 reads:

“The data must be accessible (dissemination systems, modern file formats, current software, current platforms).”

The systems and databases used and developed as part of the data collection, coding, computations, and trending activities at the INL conform to standard industry practices. Databases are in Access or Sequel file formats and reside on Unix- or Windows-based servers and work stations. All software tools and operating system platforms are either commercially supported or maintained and kept current by INL staff.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.16 Data Users' Forum**

Standard 16 reads:

“A forum for promoting a common interpretation and understanding among data users should be provided.”

The user community for the tools and data provided by the data collection, coding, computations, and trending activities performed for DRA is small, narrowly-focused, and highly-skilled in the areas of interest. The interaction between the end-users and the INL data project team is frequent and typically informal. Access to the INL staff is quite open. Questions concerning interpretation and understanding of the data products are readily handled on a case-by-case basis. To date there has been no indication of a need for a more formal forum.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.17 External Audits**

Standard 17 reads:

“There should be an external audit of the QA work process in place.”

There is currently no routine external audit of the QA/QC processes and activities. The Program Plan for Maintaining Data Quality in DRA Programs will establish periodic internal reviews and audits and recommend various NRC reviews as well. It is recommended that the NRC consider sponsoring a periodic review/audit of the INL QA/QC processes and activities by an external organization.

## **4.18 Database and Record Standardization**

Standard 18 reads:

“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.).”

Currently, the data collection and coding efforts populate a number of separate but related databases. Each record is assigned a unique identifier. It has been recognized that the current structure of the various databases is not conducive to database integration and standardization. JCN Y6468, Reactor Operating Experience Data for Risk Applications, has a task to develop a more fully integrated data collection and coding system that will reduce the need to maintain and merge multiple databases, promote standardization, and reduce the likelihood of duplicate records. The beta version of this new system will be made available for testing in 2009.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## 4.19 Data Quality Metrics

Standard 19 reads:

“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.”

Currently, the INL has no data quality metrics in place. It is recommended that data quality metrics be established and that periodically these metrics be assessed, reported, and used in a continuous improvement program.

## 4.20 Anomaly Detection

Standard 20 reads:

“There should be a prescribed means of anomaly detection (rule-based corrections, statistical methods, visualizations).”

Currently, INL uses a set of anomaly detection measures that has grown as needs developed. The tools include

- QA measures built into the data entry, such as validation of the options selected for each event.
- Review of each database event description by a second nuclear engineer.
- Statistical analyses of the data that begin with additional data processing and cross-checks between fields in the records. For example, comparisons between the “Power Level” field and the “Unit Mode” field may yield questions for investigation.
- Data processing checks oriented toward selecting the proper subset of the data for analysis. That subset needs to be complete and applicable, both for failures and for demands or operating times. Between-record cross-checks include issues such as whether unplanned demand records exist corresponding to failures that are described as occurring on such demands.
- Statistical analyses themselves that may reveal anomalies in the data. These methods augment the rule-based corrections. For example, cross-tabulations of records based on various attributes may reveal unusual combinations. The events in the associated categories are reviewed to ensure consistent interpretations. In some cases, the data may reveal that the unusual combination of attributes is valid because an anomalous event has occurred.
- Other statistical analyses and the associated graphs. Particularly for trends, the graphs may reveal areas for further review of the data.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## 4.21 Original versus Derived Data

Standard 21 reads:

“Each piece of data should be documented as either original data or derived data.”

For original data, the associated data record contains the specific LER or EPIX source information. Derived information will most likely have multiple data inputs, either original or processed (derived). These inputs will be properly documented for traceability, thus, there may not be a specific data field indicating original versus derived data, but such designations are readily discerned through the database structure and documentation.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.

## **4.22 Cross-Walk Plans for Changes in Database Structure or Scope**

Standard 22 reads:

“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.”

Because of the large volume of data collected and coded over the life of the work performed at the INL for DRA, a change in the database structure or scope of the data collected is given serious consideration. There have been times when such changes were dictated by changes in the LER reporting requirements or program needs. Every time a change is made, a plan is developed on how previous data should be treated and how future data should be incorporated.

The upcoming change in the IDCCS is a good example of such a case. The INL continued to code using the old program and coded into the new program concurrently for a period of time, making comparisons of the two data sets, evaluating the need for changes to the new program, and completing the conversion over to the new program when its performance and reliability were proven.

The current INL data QA/QC processes and practices are evaluated as adequately meeting this standard.



## **5. Data QA/QC Plan**

The reactor operating data collection, coding, and computations and trending work at the INL for DRA have a long history and have evolved into a comprehensive, integrated program. Along with the changing technical work, the QA/QC efforts have evolved as well. In the endless pursuit of excellence, the data collection, coding, computation, and trending projects have put in place many QA/QC activities and processes that have helped produce excellent data products for the NRC. However, because of the evolutionary nature of the changes, there is currently less structure and formality to the QA/QC program than would be expected from a more planned and comprehensive program. The QA program that comes out of the work under JCN N6145 will provide the additional structure and formality desired.

Based on the comparison and analysis of the data standards against the current INL QA/QC processes and practices, the following actions are recommended to achieve full compliance and establish a baseline data QA/QC program.

### **5.1 Improve Data Protection from Loss**

Standard 3 addresses computerized databases. One aspect of this standard is the protection of data to prevent inadvertent alteration or erasure. Current practices limit access to authorized users, a key feature in this area. Additionally, the databases reside on servers that are backed up on a daily basis, limiting the amount of data that would be lost in the case of an inadvertent erasure. However, there are no provisions to completely prevent the possibility of loss of data. Additional research into potential software features and capabilities that may strengthen the current status is recommended. The INL IT group has tremendous expertise that can be brought to bear in this task.

### **5.2 Peer Reviews**

Standard 5 encourages the use of peer reviews of NRC research products. Current practices have included peer reviews by members of the commercial nuclear industry for the common cause data activities and for statistical methods. Peer reviews are performed at the discretion of the various NRC projects for applications and research projects that use the data coming from the data collection and analysis processes. These peer reviews include the data, when appropriate. The data collection and analysis processes themselves are rarely subjected to peer review.

To demonstrate compliance with Standard 5, it is recommended that a peer review of our data collection and processing activities be conducted as an initial outside, independent assessment. Further peer reviews will be recommended to the NRC on an as-needed basis. Actions that would warrant the need for a new peer review include

- Major revision of the database structure
- Major change in the scope and content of the data collection efforts
- Major shift in the end uses of the data
- Major shift in the methods for processing and trending the data.

In the absence of one of these triggers, a peer review should be conducted at an interval not to exceed 10 years. This is to ensure that the data being processed for NRC research stays relevant with a high degree of utility and will be able to serve the NRC needs into the foreseeable future.

### **5.3 Independent Advisory Committee Reviews**

Standard 6 encourages the use of independent advisory committees to review NRC information. Over the years, various aspects of the data products have been reviewed by the ACRS as the programs and



projects using the data have come before the committee. The entire set of data collection and processing activities has not been the focus of an ACRS review for many years, if at all.

It is recommended that the INL operational data collection and analysis activities be subjected to an ACRS review. This would give the ACRS a more complete picture of the risk and reliability activities being performed by the NRC and a better appreciation of the amount of data and the types of insights being obtained from the industry operating experience. The unique perspective of the ACRS in their review would help ensure that the efforts in this area are properly focused.

Like peer reviews, future ACRS reviews should be considered on the same as-needed basis and, of course, at the request of the ACRS.

## **5.4 Quality Review Meetings**

Standard 8 calls for periodic quality review meetings. The INL data collection and analysis processes do not include routine meetings of this type. Whenever an issue arises, the INL holds ad hoc meetings with the appropriate individuals to address the concern. These meetings are held to address an identified issue and do not have a structured format. They do not meet the purpose of a periodic quality review meeting.

It is recommended that the INL establish a standing NRC Data Quality Review Committee. This committee would meet, as a minimum, annually to ensure the proper execution of this Program Plan and to provide internal direction and guidance. The committee would be chaired by the funds manager (Department Manager) over the data collection and analysis projects. The Committee would be comprised of the Principal Investigators of the data collection and analysis projects, the key coding and analysis staff, and selected members of the project teams that represent the end-users of the data. The first act of this committee would be to write a charter to establish its goals and objectives and powers and authorities. The primary product coming out of the committee would be a report to the NRC summarizing the reviews conducted and the meetings held. This report would verify to the NRC that the Data Quality Assurance Program Plan had been properly executed over the previous fiscal year.

## **5.5 Editorial Reviews**

Standard 11 requires editorial reviews of all disseminated information and adherence to the NRC's Plain Language Program. Editorial reviews and checks for use of plain language are not routine parts of the current data QA activities.

It is recommended that these types of reviews be routinely incorporated into the work scope associated with all data activities that produce written materials for public consumption, including NUREG reports and web sites. Data Quality Review Committee activities will include technical editing reviews and plain language checks as part of their charter. Additional reviews and checks will be part of the management reviews of data products before they are submitted to the NRC.

## **5.6 Checking Procedures and Data Validity**

Standard 13 provides direction on the use of data checking procedures to help ensure data validity. The current INL QA/QC processes and database software capabilities implement a significant number of data validity checks. The continuous improvements to the database collection software, including the IDCCS, are adding more checks all the time.

The INL recommends that a thorough review of the data collection and analysis processes be conducted to establish a baseline list of data validity checks being performed. This list would then be maintained and kept current within the Data Quality Assurance Program Plan. This list should include a description of the data checked and the data errors that the check effectively prevents. This list should be

examined to determine if there are any areas with readily identifiable shortcomings and corrective actions needed. The Data Quality Assurance Program Plan should address these corrective actions. Items on the list that are not automated should be transformed into a checklist of checks to be performed as the data are developed. The Data Quality Assurance Program Plan should also require that software updates and data collection and analysis process changes include identification and development of appropriate data validity checks as part of the change.

## **5.7 Incorporating Data Quality from the Start**

Standard 14 provides substantial guidance for new data projects on how to incorporate data quality from the start. While the NRC data projects have been in place for many years, there are a number of characteristics in this standard that can be used to improve the current data collection and analysis processes.

### **5.7.1 Data Quality Metrics**

The first aspect of this standard where improvements can be made is the establishment of data quality metrics. Metrics are needed in three areas: data collection and coding, data analysis, and EPIX data processing. In the first area, the data metrics proposed consist of an annual count of the new data records created during the previous fiscal year, the number of coding errors uncovered by all the various reviews after the second checks, and a calculation of the number of coding errors per 1000 new records. These metrics would be reported in the annual report from the Data Quality Review Committee.

For the data analysis area, the metric would be an annual count of feedback reports (from the proposed new feedback and corrective action process; see below), and the number of corrections/changes made to the analyzed data and data analysis process from discovered problems. These metrics would be reported in the annual report from the Data Quality Review Committee.

In the final area, EPIX data, the metrics would consist of an annual count of the number of EPIX records processed during the fiscal year, an annual count of the number of records that underwent a 100% INL coding review, and an annual calculation of the reporting uncertainty factors (Reference 41) for those equipment types where the data are available with a trending of the reporting uncertainty factors over time. Based on these calculations, updated component unreliability and unavailability parameters will be determined and made available for use by the various application end-user programs, as they deem appropriate.

Because much of the quality of the data coding and analysis processes is dependent on the effectiveness of the two-person coding and checking system, a metric to evaluate the quality of that system is proposed. An additional third-party check will be performed on a 10% sample of all two-person coded data. This check will evaluate the original coding results in both quantitative and qualitative respects. Errors found by the third-party check will be categorized as contributing to quantitative analysis errors or qualitative analysis errors. Error rates will be calculated and trended. Both error types, quantitative and qualitative, will be rated on a scale of 1 through 5, with 5 being no problems found in the data record. Records not evaluating to a perfect 5 and 5 will be properly annotated and corrected. Any lessons learned will be provided as feedback and training to the data coders.

### **5.7.2 User Feedback and Corrective Actions**

Currently, there is no formal mechanism to provide user feedback and to properly document corrective actions. With the user community for this information being rather small and most working either directly on projects for the NRC or with ready access to the NRC, feedback and corrective actions have been handled on an informal basis. While this method is easy to implement and has some advantages, it does not provide the structure needed to gain insights through trends and patterns.

It is proposed that a formal user feedback system be established. This would consist of

- A data feedback form that would be made available to those with access to the data
- A means of getting the forms to the INL data team
- A process of determining a course of corrective action (with appropriate approval steps)
- Assignment of corrective actions with timelines to the appropriate individual(s)
- Documentation of the actions taken
- A mechanism of closeout with the submitter of the feedback form
- Development of any lessons learned to be shared with the data coding and processing staff
- A submittal requirement to the Data Quality Review Committee for inclusion in the annual report
- An archival process for completed forms.

## **5.8 External Audits**

Standard 17 states that there should be an external audit of the QA work process in place. It is recommended that the NRC perform a bi-annual audit of the data QA activities conducted on their work. The audit team should be appointed and assigned by DRA. It is recommended that the audits take place in either the first or second quarter of the fiscal year and cover the data or activities of the two previous fiscal years. The results and findings of that audit would then be addressed through the feedback and corrective actions process.

## 6. 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.
14. Idaho National Laboratory, *LERtrk Software User’s Manual*, March 2005.
15. Idaho National Laboratory, *Integrated Data Collection and Coding System’s Functional Requirements Specification*, August 2004.
16. U.S. Nuclear Regulatory Commission, “LERSearch” web site <https://nrc.nrel.gov/lersearch>.
17. U.S. Nuclear Regulatory Commission, *Standardized Plant Analysis Risk (SPAR) Model Quality Assurance Plan (draft)*, July 2006.

18. U.S. Nuclear Regulatory Commission, *Software Quality Assurance Program and Guidelines*, NUREG/BR-0167, February 1993.
19. Smith, C. L., R. Nims, K. J. Kvarfordt, C. Wharton, *Systems Analysis Programs for Hands-on Integrated Reliability Evaluations (SAPHIRE) Quality Assurance Manual*, INL/EXT-05-00655, NUREG/CR-6952, August 2008.
20. Idaho National Laboratory, *Computation Manual for the U.S. Nuclear Regulatory Commission Reactor Operational Experience Results and Databases Web Page Updates*, not available to the public, December 2006.
21. U.S. Nuclear Regulatory Commission, *Industry-Average Performance for Components and Initiating Events at U.S. Commercial Nuclear Power Plants*, NUREG/CR-6928, February 2007.
22. U.S. Nuclear Regulatory Commission, *Reevaluation of Station Blackout Risk at Nuclear Power Plants*, NUREG/CR-6890, Vols. 1–3, December 2005.
23. U.S. Nuclear Regulatory Commission, *Estimating Loss-of-Coolant Accident (LOCA) Frequencies through the Elicitation Process*, Draft Report for Comment, NUREG-1829, June 2005.
24. U.S. Nuclear Regulatory Commission, *Baseline Risk Index for Initiating Events (BRIIE)*, NUREG/CR-6932, June 2007.
25. U.S. Nuclear Regulatory Commission, *Reliability Study: Auxiliary/Emergency Feedwater System, 1987–1995*, NUREG/CR-5500, Vol. 1, August 1998.
26. U.S. Nuclear Regulatory Commission, *Reliability Study: High-Pressure Coolant Injection (HPCI) System, 1987–1993*, NUREG/CR-5500, Vol. 4, September 1999.
27. U.S. Nuclear Regulatory Commission, *Reliability Study: Emergency Diesel Generator Power System, 1987–1993*, NUREG/CR-5500, Vol. 5, September 1999.
28. U.S. Nuclear Regulatory Commission, *Reliability Study: Isolation Condenser System, 1987–1993*, NUREG/CR-5500, Vol. 6, September 1999.
29. U.S. Nuclear Regulatory Commission, *Reliability Study: Reactor Core Isolation Cooling System, 1987–1993*, NUREG/CR-5500, Vol. 7, September 1999.
30. U.S. Nuclear Regulatory Commission, *Reliability Study: High-Pressure Core Spray System, 1987–1993*, NUREG/CR-5500, Vol. 8, September 1999.
31. U.S. Nuclear Regulatory Commission, *Reliability Study: High-Pressure Safety Injection System, 1987–1997*, NUREG/CR-5500, Vol. 9, October 2000.
32. U.S. Nuclear Regulatory Commission, “Reactor Operational Experience Results and Databases” web page, <https://nrc.nrel.gov/results>.
33. U.S. Nuclear Regulatory Commission, *Component Performance Study – Turbine-Driven Pumps, 1987–1998, Commercial Power Reactors*, NUREG-1715, Vol. 1, April 2000.
34. U.S. Nuclear Regulatory Commission, “Component Performance Study – Motor-Driven Pumps, 1987–1998, Commercial Power Reactors”, NUREG-1715, Vol. 2, June 2000.
35. U.S. Nuclear Regulatory Commission, *Component Performance Study – Air-Operated Valves, 1987–1998, Commercial Power Reactors*, NUREG-1715, Vol. 3, July 2001.

36. U.S. Nuclear Regulatory Commission, Component Performance Study – Motor-Operated Valves, 1987–1998, Commercial Power Reactors, NUREG-1715, Vol. 4, September 2001.
37. U.S. Nuclear Regulatory Commission, Common-Cause Failure Event Insights, NUREG/CR-6819, Vols. 1–4, May 2003.
38. U.S. Nuclear Regulatory Commission, Special Study, Fire Events - Feedback of U.S. Operating Experience, AEOD/S97-03, June 1997.
39. Eide, S. A., “Historical Perspective on Failure Rates for U.S. Commercial Reactor Components,” Reliability Engineering and System Safety, Vol. 80, 2003, pp. 123–132.
40. U.S. Nuclear Regulatory Commission, “Independent Verification of the Mitigating Systems Performance Index (MSPI) Results for the Pilot Plants,” Final Report, NUREG-1816, February 2005.
41. Sant, R. L., Idaho National Laboratory, letter report to John C. Lane, U.S. Nuclear Regulatory Commission, “Characterization and Measurement of EPIX Data Quality,” June 29, 2009.