RAC Report No. 4-RACER LANL-2003-DRAFT

Reference

DRAFT REPORT

Risk Analysis, Communication, Evaluation, and Reduction at LANL

Task 4: Data Identification and Management

August 4, 2003

Submitted to Colorado State University
in partial fulfillment of contract No. CSU G-3959-1

"Setting the standard in environmental health" RAC

Risk As
41
Phone

9972

DRAFT REPORT

Risk Analysis, Communication, Evaluation, and Reduction at LANL

Task 4: Data Identification and Management

August 4, 2003

Authors

H. Justin Mohler
Jill Weber Aanenson
Kathleen R. Meyer
Helen A. Grogan

Principal Investigator

John E. Till, Ph.D., *Risk Assessment Corporation*

Executive Summary

The intent of Task 4 (Data Identification and Management) is to guide the approach of identifying available data, selecting relevant data, and managing those data throughout this project in a comprehensive fashion. This draft report describes the specific pieces of data needed for a defensible risk assessment and also identifies the sources of available data. The importance of providing public access to relevant data related to environmental monitoring is stressed, as is the need for a clear understanding of the review and validation procedures used by each data-collecting organization. Considerations related to selection of data appropriate for use in this project and to the application of those data are identified and *RAC*'s proposed approach is discussed to solicit input and direction from the perspective of all stakeholders. *RAC*'s overall approach for managing and utilizing data in an integrated and consistent manner to enable comprehensive use of all relevant data is also described, and a number of priority tasks are identified.

Several specific requests for stakeholder review and recommendations for data stewards are made throughout the report. Where these recommendations or requests occur in the report, they are identified as described in the section preceding the introduction, called "About This Report." Careful consideration of these recommendations and requests for review by all reviewers of this report is requested.

DRAFT

Contents

Executive Summary iii

Tables vii

About This Report viii

Introduction 1

Task 4 Objectives 2

Data Needs, Availability, Considerations, and Application 4

Data Needed for Risk Assessment 4

Data Available for Risk Assessment 11

Data Access and Format 12

Data Usability Considerations 13
Data Review Process 14
Changes to Data 15
Qualified Data 15
Detection Limits 16
Quality Assurance 17
Identifying Biases in Data 17

Data Application Considerations 18
Documenting the State of Knowledge 18
Establishing Adequacy of Monitoring 19
Identifying Materials of Concern 19
Deriving Representative Background Values 20
Establishing an Appropriate Level of Detail for Risk Calculations 21
Propagating Uncertainty: Quantitative or Qualitative 21
Request for Stakeholder Input 21

Data Management 21

Purpose and Utility of Consistent and Integrated Data Compilation 22

Process for Achieving an Integrated Data Management System 24

Priority Tasks 26
Analytes 26
Locations 28
Standards, Limits, and Background Values 29

Conclusions 30

References 31

DRAFT

Tables

Table 1. Specific Information Required for Contemporary Risk Assessment 6
Table 2. Sources of Data and Status of *RAC* Data Collection 12
Table 3. Examples of Qualifiers Applied to Analytical Data 16

About This Report

This draft report is issued as a work in progress. It provides a discussion of the *RAC* Team's current state of knowledge and approach on Task 4: *Data Identification and Management for the project, Risk Analysis, Communication, Evaluation, and Reduction (RACER) at Los Alamos National Laboratory*. It is being distributed to provoke thought on the topic and solicit specific comments from stakeholders on the information discussed and direction being considered. It also is intended to help identify sources of additional information or other issues that have not been incorporated in this current report.

Readers should take careful note of sections titled "Request for Stakeholder Input," which ask for careful review of and input to specified material. There are also sections titled "Recommendations" that describe suggestions *RAC* is making for future actions or immediate requests for information to help complete this project.

Comments or suggestions concerning this document should be submitted in writing to Justin Mohler with *RAC* by e-mail to jmohler@avicom.net or by mail to 125 Jackpot Lane, Belgrade, MT 59714. All written comments will be considered in subsequent versions of the report.

DRAFT

Introduction

The Los Alamos National Laboratory (LANL) began as Project Y of the Manhattan Project during World War II. Its primary mission in 1943 was to develop the world's first nuclear weapon. This undertaking drove the Laboratory's operations from its inception in 1943 to August of 1945 when that goal was fulfilled. During these early years, LANL was collocated with the living quarters of the scientists involved in the project and their families. After successful completion of its original mission, the Laboratory moved toward peacetime research and development of nuclear weapons. During this time in LANL history, operations gradually were moved to the adjacent mesas, allowing a more distinct separation between the technical areas of the Laboratory and the residential portions of Los Alamos. The Laboratory was renamed Los Alamos Scientific Laboratory in 1947, which in turn became Los Alamos National Laboratory in 1981. The primary mission of nuclear weapons development has been expanded to include ensuring the safety and reliability of the nation's nuclear deterrent; reducing the global threat of weapons of mass destruction; and addressing national energy, infrastructure, and health security. As research continued at the Laboratory, concerns about the impact of facility operations on the environment increased.

The Laboratory encompasses about 27,500 acres (110 km^2) in northern New Mexico and is situated on the Pajarito Plateau, described as a series of finger-like mesas separated by deep east-to-west oriented canyons cut by intermittent stream flow. The mesas range in elevation from approximately 7800 ft (2377 m) on the eastern flanks of the Jemez Mountains to about 6200 ft (1890 m) above the Rio Grande Canyon. During its 60-year history, radionuclides and chemicals have been released to the air, soil, and surface water, and indirectly, to groundwater as a result of routine operations, accidents, and waste disposal practices. Although the Laboratory has substantially reduced radionuclide and chemical releases, the impact and risk associated with current releases and disposal of chemical and radioactive waste continues to be a concern to individuals in nearby communities. The cleanup, removal, and treatment of historical releases of radionuclides and chemicals, referred to as legacy waste, are also concerns.

In 2003, Colorado State University (CSU) undertook an independent and comprehensive risk assessment for public health and the environment for chemicals and radionuclides associated with LANL operations. *Risk Assessment Corporation* (*RAC*) of Neeses, S.C., was selected to conduct the technical work as the primary, independent sub-contractor. CSU is being supported by the University of California (UC) to conduct this work and to manage the project as an independent entity. CSU also will oversee the technical peer review process and convene a nationally recognized technical peer review panel to evaluate the approach and major deliverables of this project.

The primary objectives of this project are to develop (1) a methodology to estimate contemporary (current) human health risks and ecological impacts from LANL using available data on chemicals and radionuclides measured in environmental media; (2) a methodology to implement a comprehensive risk-informed decision analysis framework, including a prospective risk and ecological impact assessment and other quantitative and qualitative criteria, to guide long-term management of risks and ecological impacts at LANL; (3) a consistent approach for efficiently compiling, using, and updating data to support the risk assessment and decision-making processes; and (4) a process for extensive stakeholder involvement in the risk assessment and decision-making processes for LANL.

A major element of this project is developing a quantitative understanding of environmental conditions in the vicinity of LANL, which will be done using analytical measurement data associated with relevant environmental sampling. There are a number of reasons to monitor conditions in the environment, and the purpose of monitoring can impact how useful it may be for risk assessment purposes. Quite often, the primary reason for collecting environmental samples is to satisfy regulatory compliance or federal and state permit requirements. In other cases, environmental sampling is used to characterize specific areas of known contamination. A sound environmental monitoring program also assists with understanding the presence, distribution, and both current and historical trends of concentrations of chemicals and radionuclides in environmental media that may be associated with the operation of a given facility.

Consistent monitoring over time at specific locations can provide a baseline against which changes in the distribution and concentration of radionuclides and chemicals can be measured. Utilizing environmental monitoring data is one of the best methods to assess the impact that operations at a given facility have on its surroundings. As a result, environmental monitoring data are an important component of presenting a defensible risk assessment related to past, present, and future operations. To do this in a comprehensive way and maximize the value of existing data, it is important to identify, compile, and evaluate the available data in a consistent and efficient manner. This report outlines the approach *RAC* is taking to facilitate data identification, acquisition, compilation, evaluation, and management for the current CSU risk assessment project.

Task 4 Objectives

Task 4 of this project involves (1) identifying the data needed to perform the calculations in the Task 5 risk assessment and (2) designing a system for managing data to facilitate efficient and repeatable risk assessment calculations. The first activity is intended to define the data needed to evaluate public health risks and ecological impacts and associated uncertainties as part of this project. After the data needed to perform the calculations for this project have been defined, *RAC* will determine, to the extent possible, how well the currently available data meet those needs.

Most data from LANL are available primarily as a result of three separate sampling and analysis objectives: (1) environmental monitoring and surveillance, (2) regulatory compliance, and (3) site characterization. Additional information collected as part of special sampling or for research purposes at LANL and surrounding areas also may be useful for understanding and documenting current environmental conditions.

Through interactions with the data stewards (those involved in data collection and management) and data users from various organizations, *RAC* is examining current data collection activities to develop a complete understanding of the types of data collected, how and where they are collected, why and how often they are collected, and how data are used. Existing data are being assessed with regard to known needs and current uses, including estimating risk, meeting regulatory requirements, monitoring environmental conditions, understanding the system and physical processes that control movement of materials, evaluating spatial and temporal trends, comparing measured concentrations with regional background levels, and validating measurements. To complete this task, *RAC*, working closely with stakeholders, is learning how data are currently collected and managed by various organizations like LANL, the New Mexico Environment Department (NMED), the Environmental Protection Agency (EPA), and other

DRAFT

stakeholders. With this knowledge, *RAC* and the various data stewards may be able to identify how current data collection and management activities could be modified to improve the ability to evaluate risk and measure changes in risk over time, in an open and coordinated process.

The second objective of this task is to define a system (e.g., relational database) for managing the data that will be compiled for this project. This system will facilitate an efficient and readily reproducible calculation of risk and measurement of change in risk, may address other data and reporting needs for LANL, and can serve as the foundation for identifying the fundamental data elements (e.g., consistent chemical and radionuclide identifiers and terminology) that should be shared among data-collecting organizations. In designing this system, *RAC* and the data stewards are considering the uses and needs for data identified as part of the first component of this task and are building on existing structures as much as possible. The aim of this undertaking is to facilitate data interpretation, comparison, accessibility, dissemination, storage, and routine updates of newly acquired data among all data users. The long-term goal of this task is to develop a system that maximizes efficient use of available data and meets stakeholder expectations for data availability and access.

Other objectives of this task include understanding and communicating issues that impact data interpretation, adopting methods for easily selecting and combining data for specific needs, and developing effective techniques for interpreting and presenting these data. The data analysis procedures currently used by various organizations are being evaluated and will be supplemented if necessary and where appropriate. The overall goals of this task are consistent with the objectives described by Canepa (1999) for an integrated information management plan, which include developing a LANL-wide environmental database management system, assessing the accuracy of database contents, documenting changes to data, performing data gap analyses, and initiating a process to share data with stakeholders.

During the first year, *RAC* is focusing on collecting and organizing data from LANL, NMED, and other stakeholders to complete the contemporary risk assessment, which will be based primarily on the most recent comprehensive data set for which final review and quality assurance have been completed. Certain data from previous and subsequent monitoring efforts also may be used to the extent they have been validated and can quantitatively supplement the state of knowledge about contemporary environmental conditions. The goal of this effort is to develop a consistent mechanism to update information about environmental conditions at LANL (e.g., on an annual basis). This effort is striving to achieve a state of knowledge that represents actual current conditions to the extent possible considering limitations on the time needed to finalize and compile the data sets. First year activities for Task 4 include:

- Identifying data required for defensible and comprehensive risk assessment
- Summarizing environmental monitoring and other key data collected by LANL and other agencies that are applicable to, and can be incorporated into, the contemporary risk assessment
- Developing a database containing as much information as possible to support the contemporary risk assessment utilizing *RAC*'s initial proposed structure for comprehensive data management
- Proposing approaches others can use to focus future data collection and to modify ongoing data management systems and practices, when necessary, to ensure consistency and overall efficiency.

During the second and third years, *RAC* will devote more attention to the LANL site characterization data and other data needed to support an assessment of potential future risks, although evaluating these data also may be an important part of the first-year activities. Specific approaches for focusing future data collection and modifying data management activities will be proposed on an ongoing basis, as needed, so that implementation can be as timely as possible.

Data Needs, Availability, Considerations, and Application

This section discusses the approach and steps being taken to identify the data needed for this project, to understand the data that are currently available, and to address some of the issues that will both direct and impact our use of data.

Data Needed for Risk Assessment

A number of basic pieces of information are required to understand spatial (location) and temporal (time) trends and to assess the potential for public health and environmental impact. One of the primary objectives of Task 4 is to identify, or define, the data necessary to perform a defensible and comprehensive assessment of impacts related to LANL operations.

Concentrations of radionuclides and chemicals in the following environmental media are required to estimate both human health risk and ecological impact:

- Air
- Sediment
- Soil
- Surface water
- Storm water
- Groundwater
- Foodstuffs (produce, animals, applicable animal products)
- Vegetation.

In addition to concentration data for these sampled media, other types of data are important for evaluating the concentration in a given sample. These other types of data can be divided into four general categories:

- Chemical/Analytical: Information related to the analytical result obtained for specific environmental samples, the methodologies used to collect/analyze the sample, or the sample itself
- Location: Data specific to the location at which an environmental sample was collected
- Analyte information: Information related to the analyte (chemical or radionuclide) for which the environmental sample was analyzed
- Comparisons: Values that can be compared to analytical results to provide perspective on the measured concentrations or to identify potential materials of concern (e.g., background concentrations, human health standards, or regulatory limits).

DRAFT

Table 1 lists and describes the specific pieces of information in each of these categories that are considered important for assessing and measuring change in risk at LANL. It is anticipated that the majority of the information listed in Table 1 will be available from data stewards. Gray highlighting has been used to indicate information not expected to be currently maintained by the data stewards, such as:

- Environmental exposure pathways (Pathways): The exposure pathways for which the data would be used in the risk assessment
- Analyte-specific information needed to calculate risk (Calculators): Information required for risk assessment that is analyte specific, such as bioconcentration factors, slope factors, risk coefficients, etc.

In general, the information in Table 1 applies broadly to all sampled media; however, it may be most efficient to maintain information related to certain media (e.g., air) in separate tables. A data dictionary, which defines the contents, structure, and relationships of the database tables needed to compile the information for this project (see Data Management section), will be created based on the information in Table 1. It is this data dictionary and the thought given to data required to complete this project that form the fundamental basis of the database management design discussed later in this document. Reviewers of the information in Table 1 should be aware that some pieces of information are listed more than once (e.g., Analyte ID and Location ID) because they are related to more than one category of data. This is useful to note during the planning stage, but it does not imply that *RAC* intends to collect redundant pieces of information.

The majority of the information in Table 1 is based on the fundamental data needed to:

- Examine temporal and spatial trends of chemical and radionuclide concentrations in environmental media
- Characterize contemporary environmental conditions in the LANL vicinity
- Identify materials of concern
- Estimate potential human health risks and ecological impacts associated with LANL operations.

However, some of the information is related to our current understanding of the format and content of existing data sets. Based on previous experiences at LANL (Rood et al. 2002, Rocco et al. 2002, Aanenson et al. 2002), *RAC* is aware of certain procedural or operational data that will be relevant for this project (e.g., geomorphic, canyon, and reach designations; potential release site [PRS] aggregate identifiers). As a result, some of the information in Table 1 may apply only to LANL and may not be applicable at other facilities, particularly those in significantly different geographic or cultural settings.

Request for Stakeholder Input. An important step in this process is to provide an opportunity for stakeholder input regarding the specific pieces of information that should be considered relevant for conducting a risk assessment. Careful review of Table 1 for accuracy and completeness is requested.

Table 1. Specific Information Required for Contemporary Risk Assessment

Category of Information	Specific Information	Description
Chemical/Analytical	Sample ID	Sample number assigned by sample collecting organization
	Organization ID	Unique number that identifies the group or organization responsible for collecting and analyzing the sample
	Medium	Specifies the medium (e.g., soil, sediment, groundwater, produce) sampled
	Location ID	Unique number or character string used to identify location.
	Sample Type	Identifies a sample as original, duplicate, triplicate, split, collocated, blank, spiked, etc.
	Composite Type	Identifies a sample result as representing a composite over time, multiple locations, multiple depths, or a single sample collected at a single location
	Composite Number	Identifies the number of individual samples comprising a composite sample
	Composite Time Period	Specifies the time period covered by the composite sample
	Composite Area	Specifies the surface area encompassed by the composite sample
	Composite Depth	Specifies depth represented by the composite sample
	Analyte ID	Unique identifier used to refer to a specific analyte
	Result	Provides a numeric value of the analytical result
	Uncertainty	Provides a numeric value describing the analytical uncertainty associated with the reported result
	Units	Indicates the units in which results are reported
	Collection Date	Indicates the date on which the sample was collected from the field
	Start Time	Indicates the time at which the individual sampling event began (may not be recorded for all samples, e.g., soil). Necessary if multiple samples are collected at different times from the same location during a single day.
	Stop Time	Indicates the time at which the individual sampling event ended
	Sampling Duration	Indicates the duration of sampling
	Begin Depth	Indicates the start depth at which a sample was collected (e.g., for soil, sediment, or groundwater)
	End Depth	Indicates the end depth at which a sample was collected (e.g., for soil, sediment, or groundwater)
	Wet Weight	Indicates the sample wet/fresh weight
	Dry Weight	Indicates the sample dry weight
	Air Flow Rate	Indicates the flow rate at which an air sample was collected
	Water Flow Rate	Indicates the flow rate measured at the time a water sample was collected
	Analysis Date	Indicates the date on which the sample was analyzed
	Sampling Purpose	Specifies the purpose for which the sample was collected (e.g., monitoring, compliance, characterization, risk assessment)
	Laboratory Qualifier	Qualifier recorded by analytical laboratory
	Secondary Validation	Qualifier recorded by sample collecting organization
	Less Than	Indicates the true value to be something less than the reported result

DRAFT

Table 1. Specific Information Required for Contemporary Risk Assessment

Category of Information	Specific Information	Description
	Method Detection Limit	Specifies the minimum concentration capable of being detected by a specific analytical method (MDL)
	Instrument Detection Limit	Specifies the minimum concentration (or activity level) capable of being detected by the instrument used for the analysis (IDL)
	Minimum Detectable Activity	Specifies the minimum activity (for radionuclides) detectable (MDA)
	Quantitation Limit	Specifies the quantitation limit for a sample
	Sampling Method	Describes the method used to collect the sample, which may be important for comparing results from different organizations using various collection methods
	Analytical Technique	Documents the basic technique used to analyze the sample, which may be important for comparing results from different organizations using various analytical techniques (e.g., gamma spectrometry, alpha spectroscopy, or liquid scintillation)
	Analytical Method	References the specific, documented analytical procedure (e.g., EPA), if applicable
	Method Bias	Identifies any known mechanisms by which the collection or analytical technique/method used may create biases (negative or positive) in the analytical results or otherwise hinder interpretation of the results
	Data Status	Indicates whether the data should be considered final or preliminary. A final data set is one whose information is not expected to change, whereas a preliminary data set may change as a result of validation efforts or additional entries. A data set should not be considered final until all necessary quality assurance has been completed. It may be necessary to establish multiple data status levels to correspond to the entire process used to create a "usable" data set. Reference to the completed review, or validation, procedures would be through the recorded data status level.
	Change to preliminary data	Indicates a change or difference from preliminary data set
	Reason for change	Specifies the reason for all changes to preliminary data
	Comments	Provides any additional information pertinent to the record
Location	Location ID	Unique number or character string used to identify location
	Location Name	Provides a text name or description of the location (should be useful for ID of location in the field)
	Location Alias/Synonym	Some locations have several names or aliases associated with them, so this information may be best compiled in a separate table
	Location Group	Identifies locations for characterizing a general area where an exact location may not be sampled annually, or where multiple locations may best describe a given area
	Exposure Scenario	Identifies specific exposure scenario(s) for which results from this location are appropriate to characterize (this may serve the same purpose as Location Group, above)
	Open Date	Date when sampling first began at this location
	Close Date	Date when sampling ended at this location, if appropriate
	Sampling Frequency	Indicates how frequently location is sampled (e.g., biweekly, annually, variably, etc.)

Table 1. Specific Information Required for Contemporary Risk Assessment

Category of Information	Specific Information	Description
	Sampling Purpose	Identifies the purpose of the sample collection. May be captured by following 3 fields.
	Compliance	Is location sampled for compliance purposes?
	Surveillance	Is location sampled for the purpose of monitoring environmental conditions (i.e., environmental surveillance)?
	Characterization	Is location sampled for the purpose of characterizing conditions at a contaminated site?
	Background	Is location suited for establishing regional background concentrations?
	Collocated	Is location sampled by another organization?
	X-UTM	UTM x-coordinate (NAD 83, Zone 13, meters)
	Y-UTM	UTM y-coordinate (NAD 83, Zone 13, meters)
	Accuracy(x-y)	Documented accuracy of x-y coordinates.
	X-NMSP	NMSP x-coordinate (NAD 83, New Mexico Central, feet)
	Y-NMSP	NMSP y-coordinate (NAD 83, New Mexico Central, feet)
	Latitude	Latitude measure of location
	Longitude	Longitude measure of location
	Elevation UTM	Elevation above sea level for sample location in UTM (meters)
	Elevation NMSP	Elevation above sea level for sample location in NMSP (feet)
	Watershed	Ties location to specific watershed
	GeoUnit	Ties location to specific geomorphic unit. It may also be important to capture information related to stratigraphic or hydrostratigraphic (aquifer) unit.
	Reach	Ties location to specific reach
	Canyon	Ties location to specific canyon
	Aggregate	Ties location to specific watershed aggregate
	Work Element Set	Ties location to specific work element set
Analyte	Analyte ID	Unique number assigned to each analyte
	Analyte String	Unique character string for each analyte, if needed
	CAS Number	Chemical Abstract Service (CAS) number, if available
	Analyte Code	Code or abbreviation used to refer to analyte
	Analyte Name	Full name of analyte
	MAQ Analyte	Name (or ID) used by Meteorology and Air Quality Group at LANL
	WQH Analyte	Name (or ID) used by Water Quality and Hydrology Group at LANL
	ECO Analyte	Name (or ID) used by Ecology Group at LANL
	ER Analyte	Name (or ID) used by Environmental Restoration Group at LANL
	NMED Analyte	Name (or ID) used by New Mexico Environment Department
	Analyte Group	General grouping for each analyte (e.g., Radionuclide, Organic, Inorganic, Physical Property, Other). This field would be populated primarily for the purpose of sorting analytes into similar groups of materials. It is similar to the Analyte Suite field except that each analyte would fall into only one group.

DRAFT

Table 1. Specific Information Required for Contemporary Risk Assessment

Category of Information	Specific Information	Description
	Analyte Subgroup	More specific grouping for each analyte (e.g., metal, anion, VOC, SVOC, HEXP, HERB, PEST, Hydrocarbon, PAH, etc.). This field would be populated for the purpose of sorting analytes into similar groups of materials, and possibly for identifying the most appropriate analytical suite result to use for evaluating the meaning of the data.
	Material of Concern	Is analyte a material of concern?
	Background	Has a background value been established for the analyte?
	Analyte Suite	General grouping for each analyte based on the analytical suite used to obtain results for that analyte (e.g., metals, rads, SVOA, HEXP, PEST/PCB, Geninorg, VOA, Dioxin/Furan). This information differs from the Analyte Group/Subgroup information in that some analytes will be tied to multiple suites, so this information may be best compiled in a separate table.
	Analyte Class	Class of analyte for risk assessment purposes (radionuclide, chemical carcinogen, chemical noncarcinogen)
	Class Category	Indicates the classification for chemicals with respect to their carcinogenicity
	Mechanism of Action	Primary mechanism of action for analyte with respect to risk endpoints
	Primary Tissue	Primary tissue impacted by analyte in terms of risk
	Analyte Properties	Separate table to compile various analyte-specific properties, including molecular weight, boiling point, melting point, half-life, density, octanol-water partition coefficient, equilibrium partition coefficient (sediment/water), water solubility, Henry's Law constant, bioaccumulation factor, soil-to-plant transfer factor, milk transfer factor, and meat transfer factor
	Analyte Synonyms	Separate table to identify other naming conventions for analytes, as needed or appropriate
	Analyte Source	Separate table to identify primary (or secondary, etc.) source(s) of analyte in the environment
	Important Degradation Products	Separate table to identify important degradation products that should be evaluated based on the presence of the original material. TCE is one where this could be an issue
Pathways	Exposure Pathway ID	Unique number assigned to each exposure pathway
	Exposure Pathway Description	Description of exposure pathway [Inhalation, Immersion (air), Immersion (water), Ingestion (water), Ingestion (dietary), External exposure (ground shine), Dermal contact]. It also will be necessary to provide a link between the sampled medium and the pathways through which exposure to that medium may occur.
Comparisons	Comparison ID	Identifies limit or standard to be used for comparison (e.g., EPA preliminary remediation goals, ER screening action levels, regional statistical reference level, DOE derived concentration guide, EPA drinking water standards, background values, etc.)
	Value	Value of limit or standard
	Units	Units for each limit or standard
	Source	References the source of the comparison value
	Basis/criterion	Indicates the basis for derivation of the comparison value (e.g., dose or risk criterion)
Calculators	Calculator ID	Identifies coefficient/factor used to estimate potential health impact (e.g., FGR risk coefficient, dose conversion factor, slope factor, reference dose, etc.)

Table 1. Specific Information Required for Contemporary Risk Assessment

Category of Information	Specific Information	Description
	Value	Value of coefficient/factor
	Units	Units for each coefficient/factor
	Source	References the source of each calculator
	Degree of uncertainty	Provides an estimate of the degree of uncertainty associated with the calculator value. May be best accomplished qualitatively (e.g., 1 is least uncertain, and 3 is most uncertain).
	Conservatism factor	Provides an estimate of the degree of conservatism associated with the calculator value

DRAFT

Data Available for Risk Assessment

An essential aspect of this project is to locate and identify data available to characterize environmental conditions at and around LANL. These data will be the basis of both the contemporary and prospective risk evaluations. Obtaining a comprehensive understanding of the available data at the outset of the project is critical to help identify gaps or limitations in information, and to support the development of a data management structure that can be used to compile, organize, and track data. While data gaps may be present, the existing data are the least uncertain means of documenting current environmental conditions at LANL and also can assist with developing models to predict concentrations at locations where data were not collected. Information expected to be available includes data collected by LANL and by other stakeholders, including NMED, EPA, U.S. Geological Survey (USGS), and citizens' groups.

As indicated in the Task Objectives, three basic types of environmental monitoring data are available for use during this project: surveillance data, compliance data, and characterization data. Surveillance data are collected to assist in evaluating environmental impacts of the facility for purposes other than to demonstrate regulatory compliance. These data may be related to samples from both off-site and on-site locations and will likely be the type of data obtained from stakeholders other than LANL. Compliance data are collected specifically for the purpose of demonstrating conformity with state or federal regulations. Characterization data are collected with the goal of better understanding the nature and extent of existing areas with known elevated concentrations of radionuclides and/or chemicals that have resulted from LANL facility operations.

Documenting the nature and extent of areas with elevated concentrations of chemicals and radionuclides is a primary goal of ER Project sampling, and as a result much of the data collected by the ER Project is related to site characterization. In some cases, data may be considered characterization data because the sampling objective is aimed at tracking nature and extent, such as sampling of contaminated canyon sediments. However, that same data also may contribute to the evaluation of environmental impacts in the same way surveillance data would be used. As a result, the distinction between what might be considered characterization versus surveillance data can become less clear as the distance down the canyons from the original outfall or PRS increases.

Because the distinction between data types may not always be easily made, it will be important to avoid a strict adherence to categorizing data. Incorporating relevant information from each data type will help provide a comprehensive picture of the environment at and around LANL and may add to the state of knowledge about transport of materials from on-site locations to those living, working, and recreating offsite. The most important objective of this Task is to identify and compile specific data that are relevant for evaluating current and future environmental impacts related to LANL operations, regardless of the type or category that may best describe the data.

Based on initial contact with individual data stewards, much of the information necessary to complete the risk assessment is currently maintained in separate databases and spreadsheets by each group. Table 2 summarizes the data-collecting organizations that have been identified to date, the specific environmental media each organization samples, and the current status of *RAC*'s interaction with them.

Request for Stakeholder Input. An important step in this process is to solicit stakeholder input to ensure that all relevant data sources are considered, so careful review of this table is requested. In addition, contact information and direction to specific data are requested.

Table 2. Sources of Data and Status of *RAC* Data Collection

	Media Types								
Group	**Air**	**Sediment**	**Soil**	**Surface Water**	**Storm Water**	**Groundwater**	**Foodstuffs**	**Vegetation**	**Status of *RAC* Data Collection**
LANL RRES[a]-MAQ[b]	✔								Have obtained specific data
LANL RRES-WQH[c]		✔		✔	✔	✔			Have obtained specific data
LANL RRES-ECO[d]			✔				✔	✔	Have obtained data spreadsheets
LANL RRES-ER[e]		✔	✔	✔	✔	✔		✔	Working group being set up to refine data request
NMED[f] Air Quality	✔								Have met and identified data availability
NMED Surface Water Quality				✔					Have met and identified data availability
NMED Groundwater Quality						✔			Have met and identified data availability
NMED DOE Oversight	✔	✔	✔	✔	✔	✔	✔	✔	Have received initial data and have meeting scheduled for August
USGS[g]									No contact established
EPA[h]									No contact established
Citizen Groups									Have established contact for stakeholder involvement processes, but no data obtained to date
Universities (e.g., CSU, UNM, others?)									Need contact information

[a]Risk Reduction and Environmental Stewardship
[b]Meteorology and Air Quality
[c]Water Quality and Hydrology
[d]Ecology
[e]Environmental Restoration
[f]New Mexico Environment Department
[g]United States Geological Survey
[h]Environmental Protection Agency

Data Access and Format

Providing access to raw, or preliminary (i.e., as received from the analytical laboratory), data is important to establish and maintain credibility with the public. Stakeholders want access to data both in raw form and in finalized form after undergoing validation and verification. Currently, LANL does not have a policy for providing raw data to the public; the data that are made available have generally gone through one or multiple levels of review. This approach creates

DRAFT

mistrust by the public and suggests that LANL is hiding information. Developing a future protocol for data access and availability will help build credibility and trust in LANL among stakeholders.

One concern RAC related to providing access to raw data involves possible misuse of erroneous or otherwise inaccurate information. If raw datasets are accompanied with a disclaimer that the data have not undergone verification and validation, and the validated datasets are made available as soon as possible with documentation about the changes that were made to the data, *RAC* believes that stakeholders will appreciate and not exploit the preliminary data. The importance of documenting any changes that were made to the data and the rationale for making those changes cannot be understated and is discussed in greater detail below.

It may be most useful to develop distinct databases of identical structure to separately compile preliminary and final data. A determination will need to be made to identify the specific data elements that a final data set should contain if the review and validation process results in removal or consolidation of certain records. It also may be useful to create a third database to compile the data that are appropriate to use for risk assessment purposes. In any event, it is critical that the rationale and intent behind all differences between datasets be clearly documented.

To facilitate access to data by others, a common format for data compilation is critical. Compiling all of the relevant data in a consistent format, both for purposes of this project and to facilitate future access, is complicated by the fact that many data-collecting organizations maintain distinctly separate formats for managing data. Some initial steps have been taken by LANL to adopt a consistent format (see Data Management section), but the various formats currently used by other groups (both within and outside of LANL) will need to be carefully evaluated to determine the most efficient approach for merging all relevant data into a consistent format. These data-format issues are important to identify at the outset of the project, before data migration into the management structure developed for this project. Objectives of the database structure for this project are to enable efficient and comprehensive use of data relevant for risk assessment, make possible cross-comparisons between all datasets, allow common use of radionuclide- and chemical-specific information, and facilitate data access and availability.

Recommendation. That LANL and other data stewards adopt and implement a policy for providing access to both preliminary and final data, along with a clear description of and reason for all differences between the preliminary and final data sets. The process of identifying the data needed for this project and developing a database structure to manage those data is intended to serve as a useful starting point for providing long-term access to data.

Data Usability Considerations

Beyond defining data needs, identifying available data, and establishing a policy for providing access to data, an equally important part of the project is to determine the usability of available data for assessing human health risks and ecological impacts related to LANL operations. Establishing data usability, in this context, refers broadly to the process of identifying the data that are appropriate for assessing risk, adopting procedures to guide selection of relevant data, and understanding the processes used to ensure overall quality of the data. EPA guidance for data usability describes five basic issues that affect the use of data in risk assessment: (1) ensuring

data source can provide data of acceptable quality, (2) defining detection limits, (3) using data qualifiers, (4) developing a method to account for background, and (5) establishing consistency in data collection (EPA 1992). Based on initial discussions with data stewards and the understanding of available data, *RAC* has identified similar issues related to both identifying subsets of "usable" data and developing methods for application of those data that require consideration.

In general, the guidance provided by EPA (1992) provides a broad overview of a number of important issues that should be considered during the planning stages of a sound data collection program; however, it offers a limited amount of assistance in answering many of the site-specific issues related to using currently available data as part of this project. For this project, there are a number of issues to consider with regard to establishing procedures to guide data use, document data review, track changes to data, interpret data qualifiers, and use detection limit values. Ultimately, decisions will need to be made, considering stakeholder input, about the most appropriate way to address these items as part of this project, which will direct the overall approach for managing the data. The following sections identify and provide preliminary discussion related to several of these issues.

Data Review Process

The data-collecting groups at LANL use an internal system for reviewing (validating and verifying) data for accuracy, completeness, and quality. The approaches are not necessarily consistent across groups, making cross-comparison of the data more complex. *RAC* is in the process of obtaining documentation from each group to support the methods that are used to validate the data. Current procedures at LANL for validating and verifying data sets depend not only on the group responsible for the data, but also on the eventual use of the data. For example, the procedures used by the Ecology (ECO) Group are different from those used by the Water Quality and Hydrology (WQH) Group, the Meteorology and Air Quality (MAQ) Group, or the Environmental Restoration (ER) Project. The ER Project data being used for the upcoming LA/Pueblo Canyon report have undergone additional review not completed for other data in the ER database. This secondary level of validation is not currently documented in the database, so there is not a straightforward mechanism for identifying which subsets of data have undergone this increased level of scrutiny.

It will be important to understand the levels of validation and verification performed on each data set and to determine an appropriate method for using existing data with variable degrees of review. Eventual adoption of a consensus methodology for data validation and verification (i.e., a single set of clearly documented procedures) would be very beneficial in achieving consistency across all data sets. Based on discussion by LANL (2002c), a common set of data review procedures is a feasible goal.

Although the need to verify and validate data to ensure their reliability and accuracy is understood, it also is considered beneficial for stakeholders and users to have access to data before an extensive validation process, as discussed previously. The comparison that is made possible by providing users with data that are raw in form and data that have undergone validation is an important one to be able to make. The process that takes the data from preliminary to final form should be transparent. In reviews of data to date, *RAC* has been unable to readily discern these review steps, which vary by data-collecting group. Generally, several

DRAFT

processes are used in sequence, but documentation of that sequence in a summary form that is useful or clearly understandable for stakeholders has not been identified.

Recommendation. That all data stewards develop a concise summary of the processes and steps and the order in which they are completed to generate a "usable" data set, starting with the preliminary data as received from the analytical laboratory. This summary would be no more than a few pages, with reference to more detailed procedures as appropriate. Ultimately, the process needs to be transparent and understandable to everyone. The same level of detail may not be used or required by all data stewards in all cases, but it would be very beneficial if all data stewards could adopt the same basic process or sequence of steps to develop a "usable" data set.

Changes to Data

Discussions with LANL data stewards have indicated that some data (e.g., measurement values, location or sample IDs, matrix designations) may be changed following a review of the values within the context of institutional (or other) knowledge. Both for this project and for data use in general, it is necessary to plan for and develop a means to identify all changes, to document the reason for the change, and to propagate those changes to data sets that rely on the most current (or valid) set of values. Particularly in a public study, it is very important to be able to readily identify what has changed and clearly communicate why it may have been appropriate to make the change.

Recommendation. That all data stewards adopt a method for clearly identifying and tracking changes to data within their data management structure that is consistent with the review processes used to validate and verify the accuracy and quality of the data.

Qualified Data

Qualified data are analytical results that come back from the analytical laboratory with a flag indicating that the analytical laboratory identified a possible or definite problem with the data point. In some cases, the data users (e.g., LANL) also apply flags to the results. The flags include, but are not limited to, the values shown in Table 3.

Table 3. Examples of Qualifiers Applied to Analytical Data

Qualifier	Description
B	Reported value was less than the required detection limit (MAQ)
E	Reported value is estimated because of the possible presence of interference (MAQ)
N	Spiked sample recovery is not within limits (MAQ)
U	The analyte was not detected
J	The analyte was detected, but the concentration value is expected to be more uncertain than usual
J-	The analyte was detected, but the concentration value is expected to be more uncertain than usual with a potential negative bias (ER)
J+	The analyte was detected, but the concentration value is expected to be more uncertain than usual with a potential positive bias (ER)
R	The data value is rejected

A consistent approach for using analytical laboratory and LANL data qualifiers to identify and select results relevant for assessing risk is necessary. Currently, each Risk Reduction and Environmental Stewardship (RRES) group has a different internal approach of dealing with qualified data, which may differ depending on the eventual use of the data. For example, the ER Project currently eliminates R-flagged results (i.e., rejected results) from analysis, but does incorporate other flagged results (e.g., J- or U-flagged) in their analyses. In contrast, WQH personnel may eliminate both R- and U-flagged results from consideration, depending on the intent of the analysis. EPA guidance suggests that both qualified and unqualified data can almost always contribute to a better overall understanding of risk as long as the uncertainty associated with the data and the impact of that uncertainty are explained (EPA 1992). For this project, *RAC*'s intent is to use all data except R-flagged values, which corresponds to the current ER Project procedure and guidance offered by EPA. Additionally, a consistent method for identifying and qualifying results as detections will be important to establish.

Detection Limits

An important issue related to data qualifiers is the issue of detection limits. If a data point has been qualified as less than some value, it will be important to evaluate the detection limit associated with the analytical process to determine a numerical value for that data point to use in the risk assessment. Certain detection limits are required by the group requesting analysis, but instrument detection limits also may hinder detection of a material. EPA guidance indicates that the two most important types of detection limits to quantify for the purposes of a risk assessment are, for chemicals or radionuclides, respectively, the instrument detection limit and sample quantitation limit, or lower limit of detection and minimum detectable concentration (EPA 1992). The instrument detection limit, or lower limit of detection, is a detection limit related entirely to the characteristics of the counting equipment. The sample quantitation limit (SQL), or minimum detectable concentration (MDC), corresponds to the limit practically achievable given the instrument, the analytical method, and the type of sample (EPA 1992). The SQL and MDC are expected to be most useful in the risk assessment evaluation because they account for sample

DRAFT

characteristics, sample preparation, or analytical adjustments that may differ from sample to sample (EPA 1989).

Various data groups at LANL use different methods for quantifying results below the detection limit, including using the detection limit as the analytical value or using some fraction of the detection limit. It will be important to establish a common method for using the SQL and MDC to quantify results less than the detection limit. Guidance from EPA (1989) prescribes use of one-half of the SQL as a proxy concentration if there is reason to believe that the material is present in a sample at a concentration below the SQL, and use of the SQL value itself if there is reason to believe the concentration is closer to it than to one-half the SQL. Because it will be difficult to make this determination, at least initially, *RAC* intends to use a conservative approach and rely on the SQL (or MDC, for radionuclides) value.

Quality Assurance

The quality assurance (QA) checks and procedures that accompany data are important to ensure that data are accurate and representative of actual environmental conditions. Quality assurance results are related to procedures done at the analytical laboratory as well as samples prepared and submitted by the sampling organization. For data-collecting groups, quality assurance takes many forms. Collocated, split, blank, spiked, and duplicate sampling are all part of a comprehensive quality assurance program. It will be important to maintain such information in the database so that comparisons and quality checks can be made. Different groups, however, manage these data in different ways. Some groups identify these samples using "dummy" location IDs that do not correspond to an actual sampling location, while others identify blanks using the matrix or sample type code for the sample. Collocated samplers are given different location IDs in some groups and the same location ID in other groups.

RAC intends to incorporate these data into the database in a uniform fashion to facilitate evaluating quality assurance samples in a consistent manner across all data sets. This does not suggest that a complete validation and verification process be implemented as part of the risk assessment. It is important to maintain the ability to evaluate quality assurance as necessary as it may relate to individual sample results. Therefore, determining the relevance of quality control samples within the context of this project and the long-term objectives of data use and management by the various data stewards is another important consideration.

Recommendation: Develop an overview document of the quality control procedures comprising the specific sampling and analysis programs. It is anticipated that these overview documents would include a description of each quality control procedure or sample and its intended purpose and its use in ensuring overall data quality.

Identifying Biases in Data

To enable efficient and accurate data analysis, it is important to clearly flag results that are known to be biased or otherwise erroneous and have an impact on interpretation and analysis so that users of the data are aware of these issues. Some of this information is discussed on an annual basis in the Environmental Surveillance Reports (ESRs); however, much of the knowledge about issues that may impact data interpretation resides with the data stewards and is not reflected

in the databases. A mechanism to link this information with the actual data would enable efficient identification of these important issues.

Some examples of biases impacting existing information include (primarily based on discussions included in LANL 2002a):

- Change to automated surface water samplers in 1999 (complicates comparing to earlier sample results)
- 1999 ^{90}Sr results for most media are generally considered erroneous, based on discussions with LANL data stewards
- Spurious high perchlorate values reportedly due to matrix interference using ion chromatography method (before April 25, 2001); presence of sulfates interferes with analysis (LANL 2002)
- 2001 total suspended particulate (TSP) results (air)
- 2000 false positives for ^{238}Pu, $^{239,240}Pu$, ^{241}Am in groundwater
- 2000 sediment radionuclide concentrations
- 1999 change in laboratory analytical procedure created systematic errors in ^{90}Sr analyses (McLin and Lyons 2002).

Request for Stakeholder Input. It is important to identify all issues that may impact data interpretation but that may not be clearly tied to the data. As a result, careful review and input, particularly by data stewards, with regard to known biases or errors inherent in currently compiled data is requested. In addition to identifying the actual bias, input on options or recommendations for accounting or adjusting for these issues, if appropriate or possible, is also requested.

Data Application Considerations

In addition to the considerations discussed above that impact identification and selection of relevant and accurate data, a number of issues related to the actual application, or use, of available data for this risk assessment project also will guide the approach for managing data. Although these issues are most obviously related to development of the risk assessment methodology, consideration of their impact on the information needed at this stage in the project is important to ensure consistency with the desired endpoints for applying the data. The following sections identify and provide preliminary discussion of several important considerations that will guide the eventual application and use of available data.

Documenting the State of Knowledge

Determining an appropriate methodology for representing the current, or contemporary, state of knowledge about environmental conditions in the vicinity of LANL is one of the fundamental issues impacting the risk assessment methodology being developed for this project. Establishing an appropriate time period (e.g., three to five years) over which measurements are considered representative for documenting a contemporary state of knowledge will be an important part of this process because of the dynamic environmental conditions present at LANL. The question of when data are no longer representative of current site conditions will be important to address, and

DRAFT

it is possible that the answer may vary by site or location depending on the relative stability of conditions across the site and the occurrence of such things as localized flooding events.

RAC is in the process of identifying and compiling available data to document the contemporary state of knowledge. To limit the amount of data requested while still capturing important knowledge about existing conditions in the environment, the identification and eventual use of the data has been constrained to results related to sampling conducted from 1997 to the present. This period of record was selected because it limits the impact of complicating factors that are known to exist for certain earlier sampling, and it encompasses the time period of the Cerro Grande Fire, which has had a measurable impact on local environmental concentrations. This time period is more inclusive than relying on only the most recent and current data sets and will help *RAC* develop a better understanding of the general environmental conditions that currently exist at LANL. Assessing data from this more inclusive time period should enable a more complete understanding of the current state of knowledge, particularly in those cases where current measurements are either not representative or were not made. This approach also is consistent with the current ER Project approach of focusing on post-1998 data to guide its efforts. It is also consistent with an approach discussed by WQH (2003a), which *RAC* supports, of basing important programmatic decisions on an accumulated weight of evidence over multiple sampling events rather than isolated or one-time occurrences. Although *RAC*'s focus is on post-1997 data, some sites or areas only can be characterized by data collected before 1997. As a result, it will be necessary to incorporate some data for earlier years.

It will be important to establish the state of knowledge (i.e., through measured or estimated concentrations at specific locations), so that it can be used quantitatively to assess risk. Because of the public nature of this study, it is very important to document the state of knowledge with actual data so the process is as transparent as possible and the assumptions that are made in the risk assessment process can be readily understood.

Establishing Adequacy of Monitoring

It is useful to demonstrate that the appropriate materials are monitored for in the appropriate locations. As a result, it may be necessary to develop lists of analytes used or released as a function of location (e.g., watershed). Options for this include review of Toxic Substances Control Act (TSCA) and Toxic Release Inventory (TRI) reporting to identify potentially used and released materials. It may also be useful to evaluate the positively detected materials at all PRSs in combination with knowledge of historical releases potentially impacting a given watershed to demonstrate the rationale for monitoring in specific areas. This would enable a cross-reference of potentially released materials with monitored materials to help identify data gaps, or unnecessary monitoring and associated data collection. It also may be necessary to identify degradation products to ensure that monitoring efforts are sufficient to detect the presence of potentially important degradation products. For example, some degradation products are more toxic than the original material (e.g., tetrachloroethane).

Identifying Materials of Concern

An important part of documenting the contemporary state of knowledge is a transparent and readily understood mechanism for identifying the materials of concern (i.e., chemicals or

radionuclides) that are of LANL origin and that may pose some measurable level of risk to public health or the environment. Appropriate techniques for accomplishing this aspect of the work may be through comparisons to background values or various media-specific standards and limits. Comparison to these values, as appropriate, may represent one of the most efficient and useful mechanisms for identifying and ranking materials of concern using a risk-informed process. Another option for ranking could be through examination of data summaries, such as frequency of detect tables (ER 2001b). A companion report to this Task 4 report is currently being written to discuss the importance of identifying materials of concern, and it presents some options for developing defensible methods for this identification.

Deriving Representative Background Values

Background, for the purpose of this project, is defined as the concentrations of radionuclides and chemicals that are present in the environment around LANL, but that are not attributable to LANL operations. Determining an appropriate methodology for estimating representative background concentrations is an integral part of evaluating either the presence or lack of operational impact from LANL on the surrounding environment. Several documents have been generated by LANL that report background values for a number of radionuclides and metals (Ryti et al. 1998, Longmire et al. 1996, McLin and Lyons 2002). *RAC* has reviewed these reports to determine the potential usefulness of the reported background values. These reviews focused on understanding the statistical analyses, calculations, and data sets, including the locations and time periods of sampling, upon which the background values are based. In most cases, it was determined that the reported values are not directly relevant to this work. For example, the data sets used to calculate the reported background values for fallout radionuclides include data from locations along the perimeter of LANL; therefore, they are not representative of background concentrations in areas not impacted by LANL. Furthermore, most of the reported background values represent the upper, 95% confidence limit on the 95th percentile of the measured concentrations (i.e., represent the upper limit of the background distribution).

*R*AC is also investigating the option of representing background through development of a consistent method of using routine environmental monitoring data to develop a continuous running average, range, or other statistical representation of background (perhaps for a three- to five-year period to be consistent with documenting the state of knowledge). Samples from regional (i.e., background) stations are currently collected for several different media on a routine basis. The data management design that is being developed for this project will enable an automated process by which these data could be used to represent background levels. This methodology would maintain values representative of current conditions, develop values for media other than soil or sediment, and avoid the need to refer to a number of different studies related to background, which all report somewhat different values. The WQH Group is currently working on a procedure to use environmental surveillance data to develop representative background levels for regional surface waters, using a method to normalize water concentrations with suspended solids measurements. *RAC* intends to evaluate and incorporate the results of the WQH effort into this project to the extent possible. A supporting report is currently in progress as part of Task 4 to discuss the importance of establishing background levels of radionuclides and chemicals in the LANL region, describe the intended use of those values as part of this project, and evaluate the information available to support this effort.

DRAFT

It will also be necessary to develop appropriate methods for making comparisons of measured concentrations of chemicals and radionuclides with background concentrations. For example, a simple ratio may be useful as part of an initial screening process to identify potential materials of concern, whereas more detailed statistical comparisons may be more appropriate for confirming or refuting facility impact. Guidance offered by EPA (2002) may be useful in this regard.

Establishing an Appropriate Level of Detail for Risk Calculations

There are a large number of radionuclides and chemicals related to LANL operations, as well as variable habits and lifestyles of members of the public in the LANL region. As a result, it will be important to develop a level of detail for the risk calculations that is adequate for representing potential exposures, but not overly cumbersome to achieve as a practical matter. For example, the level of detail achieved by and the effort associated with propagating exposures through a large number of potentially consumed produce items may not be necessary to adequately estimate potential exposures. As another example, risk estimates can be made for specific tissues or for the whole body. It is likely that a single significant figure will be used for estimating risk values and hazard quotients. It will be important to develop defensible justification for the level of detail and associated parameter values necessary for estimating potential health impacts. The time and effort required to compile the parameters needed for completing the calculations will depend on the level of detail determined to be appropriate for developing an accurate representation of human health risk and ecological impact.

Propagating Uncertainty: Quantitative or Qualitative

Significant uncertainty may be associated with estimated risks. It will be important to identify a level of uncertainty analysis that is consistent with available data and with the overall objective of uncertainty analysis in this project. EPA (1992) notes that most risk assessments end up with a quantitative assessment of risks and a qualitative assessment of uncertainty, and that for most risk assessments, it may not be possible, nor necessarily advantageous, to develop a quantitative uncertainty analysis. Appropriate techniques for incorporating the uncertainties associated with unqualified data with the uncertainty of a less than or qualified value, or with the uncertainty of a modeled or predicted value will need to be developed.

Request for Stakeholder Input

All of the above data application considerations are important factors that impact the direction taken to develop a credible risk estimation methodology for this project. As a result, careful stakeholder review of these considerations and proposed approaches, along with ideas about considerations not discussed here, are requested.

Data Management

This section outlines the approach that *RAC* is taking, with input from all data stewards, to create an integrated and consistent protocol for data management as part of this project. Although

these actions are necessary for this project, they also may improve the ability for LANL, NMED, and other interested stakeholders to evaluate data comprehensively, efficiently, and quickly. This enhanced capability can assist greatly with answering questions internal to LANL and from members of the public about the environmental condition in and around LANL and about LANL's impact on risk to human health and the environment. A consistent protocol for data management also will lead to an improved position for providing access to and dissemination of data, along with the supporting information needed to understand and evaluate the data (i.e., a single data dictionary that is comprehensively applicable to all data sets). The objectives of Task 4 are consistent with other efforts aimed at increasing information sharing capabilities and adopting standards for nomenclature and format (EPA 1993 and 2000).

The fundamental design elements for data management during this project are most heavily influenced by the data needed for risk assessment (see Data Needs, Availability, Considerations, and Application). As a practical means of facilitating data migration during this project and in the future, though, the current content and structure of available data are also guiding the data management process. It is important to understand that all data-related issues cannot be anticipated at this time. As such, it is essential for everyone to remain as flexible as possible, particularly early in the process as *RAC* and stakeholders begin to more completely identify complicating issues. This report discusses the current approach, the steps that are being taken to implement that approach, and the progress made to date. The database tables and their design are currently a work in progress. There are a number of issues to resolve before beginning to finalize the structure and framework for managing and compiling the data that will be used as part of this project, many of which have been discussed in the preceding sections of this report.

Purpose and Utility of Consistent and Integrated Data Compilation

Development and eventual adoption of a consistent data-compilation protocol by all data-collecting organizations would greatly enhance the utility and usability of the data. Such a system not only achieves a common format for the data that are compiled, but also makes it possible for anyone, particularly those independent of the data-collecting organizations, to interpret the data in an efficient, timely, and thorough manner. Collection and management by multiple groups and organizations using separate and different methods is a significant hindrance to maximizing data utility. A recent effort by the ER Project and WQH Group at LANL to adopt a consistent database format is providing a valuable stepping-stone toward creating an integrated approach for use as part of this project (LANL 2002b). This format also is being incorporated to some extent into NMED data collection. While this is an important step in the right direction, there are still a number of obstacles to overcome before a consistent format for all of the data needed for this project is achieved.

A critical component of any data set is the accompanying data dictionary, which describes the contents of the fields included in each table and defines the codes that are used to populate those fields. A consistent and integrated protocol means that a single data dictionary would be used by all data-collecting agencies. This also serves to make datasets transparent to all users, which is critical for effective use of data.

Once a consistent format is achieved, queries can be written and other useful tools or applications incorporated to automate a number of useful functions regardless of which data set is

being used. Following are a number of specific examples of potential uses for data that a consistent format and structure could enable and, in many cases, automate:

- Rapid and efficient examining of temporal trends by analyte, medium, or location/location group (e.g., scenario, watershed, canyon reach). Queries and forms can be designed to automate figure and graph generation within the database application. Trends at specified locations or groups of locations can be compared directly with trends at background or regional locations, detection limits, or standard/limit comparison values. This could help expedite preparation and presentation of data in the annual ESR.
- Efficient linking of analytical results with sampling location coordinates to facilitate interpolation or extrapolation procedures and contour/isopleth creation.
- Using data efficiently to calibrate atmospheric, surface water flow, and erosion models, as well as validate modeled concentrations and contaminant transport.
- Comparing results at specific locations or groups of locations with background or health standard and regulatory limit comparison values using ratios, sorted from highest to lowest. This represents one option for a screening mechanism to focus on the most important materials of concern.
- Comparing collocated, split, duplicate, blank, etc., samples, as well as comparing (or selecting) results from different methods. Two important advantages of data collection by multiple and different organizations (i.e., collocated and split samples) are to provide a mechanism for validating or confirming analytical results and to allow a means for conducting additional sampling that may not be part of an existing program. Designing the system with this in mind would enable automation of these comparisons.
- Grouping results based on subsets of locations (e.g., locations applicable to a specific scenario) appropriate for estimating risk to the public or environment and measuring change in risk. Totals queries can be written to calculate automatically the mean, standard deviation, and range of values for a defined group of results. Automatic calculation of other statistical parameters, such as the median, also may be useful.
- Linking results for materials of concern to appropriate carcinogenic risk and slope factors and noncarcinogenic hazard quotients. Compilations of analyte-specific properties such as partition coefficients, bioaccumulation and transfer factors, and molecular weight also can be linked with corresponding analytical data for modeling and risk estimation purposes.
- Developing and using a consistent methodology to determine whether an analytical result should be considered as a detected result when it is close to the detection limit.
- Developing a single specification/protocol for data provided by all analytical laboratories used by LANL, NMED, and other data-collecting organizations to facilitate migration and upload of electronic data deliverables (EDDs).
- Providing ready access to key pieces of data needed for decision-making as well as for dissemination of data to and access of data by the public.
- Ensuring that all appropriate quality assurance (QA) or other review and validation procedures are completed using a consistent methodology, and identifying whether results have been finalized or should still be considered preliminary.
- Examining analytes measured as a function of such things as data-collecting agency, sampled medium, or purpose of data collection. Also evaluating analytes measured at

PRSs (or other source areas) impacting a given location (e.g., watershed) versus the analytes measured at locations that could be potentially impacted by those PRSs. These types of assessments can be used in identifying data gaps or discrepancies or in focusing data collection to ensure that the most appropriate measurements are being made for samples collected at each monitoring location.

- Clearly flagging results that are known to be biased or otherwise erroneous and that have an impact on interpretation and analysis so that users of the data are aware of these issues.
- Comparing filtered and unfiltered water sample results, along with suspended sediment concentration (SSC) or total suspended sediment (TSS) measurements for the same sample, to develop site-specific estimates of distribution coefficients (K_ds).
- Examining the age or persistence of different materials of concern in key PRSs to project future behavior based on previous historical impact, as defined by monitoring results.

Importantly, the eventual uses for the data must be considered in designing the protocol for managing those data. There may be other aspects of data use that have not been considered here. It also may be useful to examine the capability and application of the SMART product discussed by LANL (2002b) as a data extraction, analysis, and reporting tool to accompany the ER Project database (ERDB).

Request for Stakeholder Review. It is important to identify other uses for data to ensure the design adopted for use as part of this project can serve as many beneficial purposes as possible in the future. Stakeholder input on uses for data is requested.

Process for Achieving an Integrated Data Management System

Developing and adopting a consistent and integrated data management system will take considerable time and effort, particularly given the extremely large amount of currently available data. The design of the system must first be carefully considered to maximize the utility of data. Next, a series of steps must be taken to move from the current situation of multiple, and distinctly different, data management systems to a single system that can incorporate all relevant data. For this project, work has begun on the series of steps to get from multiple systems to one system, some of which overlap and are occurring simultaneously.

The initial step has been to set up meetings with data stewards to become familiar with and identify the specific data that are collected and the purposes for which they are collected. This process was started with LANL data stewards, primarily because they will provide the largest sets of available data. However, the importance of data collected by other organizations, particularly NMED, is recognized, and *RAC* is just now beginning this process with the department, as indicated in Table 2 of this report.

RAC also has established a number of priority tasks and has begun to work on them. The two primary tasks are to (1) develop a comprehensive list of analytes for which analytical results are available for all organizations and media and (2) obtain location coordinates for environmental sample results that are relevant or necessary for completing both the contemporary and prospective risk assessments, with a focus on the data needed for the contemporary assessment.

DRAFT

These two priority tasks, their importance, and current progress on achieving them are discussed in greater detail below.

RAC has begun developing a database design to facilitate efficient use of data for this project. Although the primary focus of this design is related to the pieces of information needed for this project, *RAC* also is attempting to anticipate and consider as many potential uses as possible to maximize functionality of the application. For example, an evaluation by NMED personnel of the RRES-ER database and a recent data submittal to the NMED Oversight Bureau suggests that a protocol or specification for data dissemination would be very beneficial to ensure that all important pieces of information are provided as part of filling requests for data.

Because various LANL groups maintain the majority of available data, the structure and design of the application being developed is based in large part on the current design of the ER and WQH databases in an attempt to minimize difficulties that may be encountered during data migration. It is anticipated that the recent efforts by ER and WQH to adopt a consistent design will be of significant benefit in developing a system that can efficiently compile existing data. Use of existing structures, codes, and definitions will not only facilitate data migration, but also will increase the potential for long-term acceptance and use of the tool developed for this project. To assist with compilation of the various data sets both now and in the future and to identify changes necessary to achieve an integrated set of data, *RAC* is linking individual elements of the design being developed for this project with the elements in the current databases used by the ER, WQH, and MAQ groups at LANL. This linkage will be expanded to incorporated ECO, NMED, and other data sources so that differences in the various data management systems can be readily identified.

To accompany the database developed with the stakeholders for this project, *RAC* is creating a single comprehensive data dictionary to provide supporting information and descriptions for the tables and fields contained in the database. Hyperlinks will be incorporated in the electronic version of the data dictionary to provide rapid and efficient access to the lookup tables containing definitions for the codes used to populate many of the fields, as well as the tables linked to the primary data table. This will increase the database's utility and understandability. Reliance on a single data dictionary will significantly simplify the process of interpreting the data and providing data understandable to anyone.

Another step that could be taken to improve the efficiency of data compilation would be to develop a single specification for electronic data deliverables (EDDs) from all analytical laboratories that could be used by all data-collecting organizations. To date, EDD specifications have been received from the MAQ, WQH, and ER groups at LANL. Initial review of these documents indicates inconsistencies in the requested format of data among these three groups that could be readily resolved and would provide greater continuity among the datasets. For example, some groups request that results for quality control (QC) samples (blanks, duplicates, and spikes) and quality control analytes not be reported in the EDD file. When samples are diluted or reanalyzed for quality control reasons, however, the results are designated a value in the QC type field that indicates the sample handling that was required. For other groups, duplicates and replicates are included in the EDD with a notation in the comment field about the type of QC sample that is represented (MAQ 2003, ER 2001b, WQH 2003b) This is a fundamental reporting difference that affects the data contained in the individual databases.

Also in the individual EDD requests, quantities that are less than the detection limit are required to be reported in different ways. Some groups require that the MDL, quantitation limit,

or actual result populate one field and another field be populated with a U qualifier. This same result reporting strategy is used by another group, but a field is populated with a less than symbol instead of a qualifier (MAQ 2003, ER 2001b, WQH 2003b). This also affects the way data are contained within the databases and migrated into the *RAC* database.

Additionally, the EDDs for different data-collecting groups at LANL specify different designations for the analyte names. As noted in the priority tasks below, these disparate designations for analyte name make combining data tables from different groups significantly more challenging. LANL (2002c) indicates that WQH and ER intend to adopt the same EDD specification in the future, and that the current WQH EDD is the preferred specification for that purpose. The WQH and ER effort should serve as a valuable stepping stone toward creating an EDD that can be utilized by all data stewards, but the effort to create a single EDD specification should recognize that eventual goal in its design and incorporate input from other data stewards as necessary. The current WQH and ER effort to create and adopt a single database information system is based on nearly three years of effort (LANL 2002b). The database design was developed to be intentionally generic to allow accommodation of many types of environmental data, and the eventual goal was to enable exchange or consolidation of all data (LANL 2002d). The WQH and ER database design is also very similar in structure to the database currently used by NMED. For these reasons, *RAC* considers it to be the most appropriate starting point for achieving a single EDD specification that can be utilized universally by organizations collecting analytical data for sampling related to measuring conditions in the LANL environs.

Recommendation: Indicate a LANL-wide EDD specification that applies to all environmental sampling analytical data (i.e., all media) and require that all LANL data-collecting efforts utilize it, beginning as soon as practically achievable. Require that analytical labs move toward submitting EDDs. The EDD specification should be based on the current WQH EDD, modified as necessary to be consistent with the current WQH/ER database structure and nomenclature. This specification should be expected to undergo periodic review and modification to address changing or new data needs and management practices, as well as any stakeholder concerns that may not have been considered during its initial development. It should also be expected that variable levels of time and effort to adopt a new specification will be required for different data collectors depending on each organization's current data collection and management practices.

Priority Tasks

As discussed above, work was begun on certain tasks considered to be priorities for moving forward with the data management task of this project. These tasks include creating a single comprehensive analyte list, obtaining coordinates for sampling locations, and compiling values to be used for comparing to measured concentrations in collected samples. The following sections provide details about the importance of each of these priority tasks.

Analytes

The first priority task identified has been to establish a single comprehensive list of analytes for which environmental monitoring results are available during the time period of interest. A

DRAFT

single comprehensive list of analytes (analytes is used here generically to refer to all materials and parameters for which measurements may be made in an environmental sample) is absolutely essential to enable an efficient and comprehensive assessment, both now and in the future. Because different groups use various nomenclatures to refer to the same analyte, it is important to establish a single listing of analytes to enable a comprehensive assessment of results for a given analyte across all relevant environmental media and all data-collecting organizations. It is recognized that the same analytes and parameters are not measured for all media. It also is recognized that results for all analytes or parameters for which measurements are made will not be needed for this project. However, developing a specific list for this current project is difficult at this point, and the best approach seems to be one that is comprehensive and covers all measured analytes and parameters.

Currently, analyte lists have been obtained from all LANL data stewards, and a single list that includes all analytes measured by one or more groups has been generated. Separate fields have been created to capture the name, code or abbreviation, and chemical abstract service (CAS) number for each analyte. The list of analytes is nearly complete, based on information that has been received from LANL data stewards. The primary remaining part of this task will be to incorporate the lists that have been obtained or will be obtained from NMED.

RAC also has established general groups of analytes (e.g., radionuclides, inorganics, organics) and more specific subgroups (VOC, PAH, PCBs, pesticides, herbicides, high explosives, metals, etc.) to assist with sorting and reviewing the list according to like materials. The groups and subgroups that have been designated are largely based on the groupings used by the ER and WQH groups at LANL. In some cases, other groups analyze for materials not analyzed by ER or WQH, and this information is incomplete. The groupings used at LANL note which analytical suite provides results for each analyte; in a number of cases, results for the same analyte are obtained through different analytical methods. In these cases, the analytical subgroup may be best assigned according to the analytical method that provides the lowest detection limit. This would enable automatic selection of the most appropriate result when more than one result is available through multiple analytical methods.

For this project, the intent of establishing groups and subgroups (if appropriate) is to create a system that lends itself to sorting the many analytes into logical groupings that are most helpful for examining the list with regard to a risk assessment, as well as understanding what materials are being monitored and evaluating the appropriateness of that monitoring. EPA (1989) describes an approach for grouping chemicals by class, such as by structure-activity relationships or other similarities. This type of approach may be useful as part of this project.

A valuable future component of this task may be to determine and document the source or primary sources of each material to assist with flagging those that would be expected to be solely or largely LANL-derived, or those whose presence is not likely a result of LANL operations. This could be quite useful for screening and flagging important materials of concern.

Another important future task will be to categorize each analyte determined to be a material of concern (i.e., one for which risk will be calculated) as radionuclide, chemical carcinogen, or chemical noncarcinogen, and compile the appropriate risk coefficients, slope factors, and reference doses. The risk coefficient and slope factor are used to estimate an upper-bound probability of an individual developing cancer as a result of a lifetime of exposure to a particular level of a potential radionuclide or chemical carcinogen, respectively. Reference doses are used to calculate a hazard quotient, which is a ratio of the level of exposure of a chemical of concern over

a specified time period to the reference dose for that chemical of concern derived for a similar exposure period. In addition, *RAC* is creating a table to compile various analyte-specific properties, including molecular weight, boiling point, melting point, half-life, density, octanol-water partition coefficient, equilibrium partition coefficient (sediment/water), water solubility, Henry's Law constant, bioaccumulation factor, soil-to-plant transfer factor, milk transfer factor, and meat transfer factor.

Locations

The second identified priority task is to obtain sampling location coordinates for routine environmental monitoring locations, as well as locations sampled for site characterization purposes (e.g., potential release sites). Potential releases sites (PRSs) are areas within LANL that are potentially contaminated with hazardous or mixed wastes that are subject to the requirements of the Resource Conservation and Recovery Act (RCRA). Location information is needed to examine data (i.e., measured concentrations) spatially and to have a clear understanding of where samples currently are or have been collected. As with the analyte list, *RAC* has requested a listing that includes all sampled locations beginning with 1997. As noted above, however, data before 1997 represent the only information available for some locations. It likely will be necessary to obtain all available PRS-related sample location coordinates, but it will also be necessary to determine which subset of PRSs may be most relevant to incorporate into the contemporary risk assessment (e.g., those to which public access is likely, possible, remote, or impossible). *RAC* is currently in the process of obtaining this information from LANL data stewards and will begin to compile location information related to NMED sampling shortly.

There are a number of issues to consider with regard to compiling the location information. Most importantly, a mechanism must be available to link an individual environmental sample and its associated analytical results to defined location coordinates (X and Y plane). To the extent possible, the existing unique location identifiers maintained by each data group will be used; however, to avoid duplication it may be necessary to develop a separate field or combine two fields to use as a unique identifier (i.e., primary key) for each location. It is understood that coordinates may not be available for some biota and other measurements (e.g., elk or other wild game).

Based on experience, most coordinates maintained by the different LANL groups are in the following projection: New Mexico State Plane (NMSP), North American Datum (NAD) 83, New Mexico Central, feet. Some coordinates, such as those initially provided for the NEWNET stations, have been provided in a Geographic projection (i.e., latitude and longitude). For all coordinates that are provided, the projection must be identified so it can be converted to the coordinate system *RAC* is using. For this project, the following projection will be used: Universal Transverse Mercator (UTM), NAD 83, Zone 13, meters. Additionally, all coordinates will be converted and maintained in a number of different projections, to accommodate potential need for a specific projection as part of a different project.

RAC also is collecting and compiling available information to assist in identifying aliases or synonyms that may have been used at some point in time, but that do not correspond to current location nomenclature. This information is valuable to maintain for future incorporation of historical data that may include out-of-date location names. In addition, work is being done to identify locations within the context or purpose of sampling at each location. For example, this

DRAFT

type of information will be useful to clearly mark locations appropriate for establishing regional or background levels, or for characterizing existing sources of materials of concern. Subsets of locations also will be identified and marked as useful for specific tasks within this current project, which include estimating both contemporary and prospective risks to members of the public and assessing ecological impact. It also will be important to identify collocated sites where samples are collected by (or split between) two or more different organizations, so the process of comparing collocated or split samples can be automated.

Recommendation: That the ER Project identify specific PRSs at which public exposure is possible, and rank the potential for exposure at each PRS as likely, possible, remote, or impossible. At a minimum, characterization data will be required for each PRS where public exposure is likely and will be incorporated into the contemporary risk assessment.

Standards, Limits, and Background Values

Another priority task is tabulating health standards, regulatory limits, representative background levels, or other values useful for comparison to measured concentrations of chemicals and radionuclides in environmental samples and linking those values to the corresponding analyte and relevant environmental medium. Discussions with ER Project personnel regarding Environmental Screening Levels (ESLs) have revealed that conservatisms in developing and selecting ESL values has led to screening numbers well below routinely measured background levels. As a result, it will be important to identify those chemicals or radionuclides with regional background concentrations that approach or exceed comparison values to avoid inappropriate selection of materials of concern. Completion of this effort would occur after the comprehensive analyte list has been established, but identification and collection of these "comparison" values has already begun. Once all relevant comparison values have been identified, a summary table will be generated.

It also will be essential to collect and compile appropriate hazard and risk factors, but as noted above, it is likely that focused compilation of these "calculators" will be most efficient after a primary list of materials of concern has been established. Units of measure will need to be consistent with the units used to report measured environmental concentrations. It may be useful to incorporate the ability to select different units because standards and limits often use conventional units, and health impact factors often use System International (SI) units. Based on the discussion and comparisons made in the 2001 ESR, relevant "comparison" values include but are not limited to the following:

Water: EPA drinking water standards (40 CFR 141), NM drinking water regulations, DOE derived concentration guides (DCGs), NM wildlife habitat standard, EPA maximum contaminant limits (MCLs), New Mexico Water Quality Control Commission (NMWQCC) stream, general, livestock, watering, wildlife habitat, and groundwater standards, National Pollutant Discharge Elimination System (NPDES) Permit No. NM0028355

Sediment and Soil: EPA Preliminary Remediation Goals (PRGs), residential soil screening levels, ER screening action levels (SALs), regional/baseline statistical reference levels (R/BSRLs)

Air: 40 CFR 61 App E concentrations, DOE derived air concentration (DAC) for workers, action levels (investigation and alert)

Foodstuffs and Biota: dietary limits for protection of fish-eating animals, limits for protection of piscivores and fish, no-observable-adverse-effects-concentration (NOAEC), or level (NOAEL), or lowest-observable (LOAEL) (Sample et al. 1996), vegetation RSRLs

Other possible sources for comparison values identified to date include resources from NMED (2000), SC&A (2000), and Mauro (2003).

Request for Stakeholder Review. Stakeholder input on identifying relevant sources of comparison values is requested.

Conclusions

The data identification and management task of this project represents a significant effort and an important place for stakeholder involvement. *RAC* has attempted, in this report, to provide a progress update and to describe some of the important considerations that need to be resolved for this task to be successful. Stakeholder input is encouraged in the form of specific comments on the content of this report; identification of additional data sources; thoughts on the identified data use and application considerations; and any additional considerations related to data identification, use, and management that have not been discussed in this report.

DRAFT

References

Aanenson, J.W, H.A. Grogan, S.J. Maheras, H.J. Mohler, A.S. Rood, P.G. Voillequé, J.E. Till. 2002. *Independent Technical Audit of Los Alamos National Laboratory for Compliance with the Clean Air Act, 40 CFR 61, Subpart H in 2001. RAC* Report No. 6-DOJ-LANL Audit-2002-FINAL. *Risk Assessment Corporation*, Neeses, South Carolina. October.

Canepa, J. 1999. Los Alamos National Laboratory. Memo to K. West. Los Alamos National Laboratory. Subject: Recommendations for ER Project Integrated Information Management System. Memo E/ER:99-228. August.

U.S. Environmental Protection Agency (EPA). 1989. *Risk Assessment Guidance for Superfund, Volume 1, Human Health Evaluation Manual (Part A).* EPA/540/1-89/002. Office of Emergency and Remedial Response, Washington, D.C. December.

EPA. 1992. *Guidance for Data Usability in Risk Assessment. Parts A and B.* 9285.7-09A. Office of Emergency and Remedial Response, Washington, D.C. April.

EPA. 1993. *Minimum Set of Data Elements for Ground Water Quality.* EPA 813-B-93-001. Office of Water. September.

EPA. 2000. Agreement between the USGS Water Resources Division and the EPA Assessment and Watershed Protection Division on the Solution of the STORET/NWIS Compatibility Issue. World Wide Web address: http://www.epa.gov/STORET/usgs_memo.html. May.

EPA. 2002. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites.* EPA 540-R-01-003. OSWER 9285.7-41. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C. September.

Environmental Restoration Project (ER). 2001a. *Quality Assurance Checklist for Preparation of Data Sets from the ER Project Technical Database.* DI-4.28. ER2000-0402. Los Alamos National Laboratory, Los Alamos, New Mexico. March.

ER. 2001b. *Environmental Restoration Project Specifications for Electronic Data Deliverable Files. Draft.* LA-UR-01-5583. Los Alamos National Laboratory, Los Alamos, New Mexico. October.

ER. 2001c. Generating Standard ER Project Data Table Formats Using MS Excel Tools. DI-4.25. ER2000-0614. Los Alamos National Laboratory, Los Alamos, New Mexico. March.

Los Alamos National Laboratory (LANL). 2002a. *Environmental Surveillance at Los Alamos during 2001.* LA-13979-ENV. September.

LANL. 2002b. *Groundwater Protection Program Consolidated Data Management Workflow Analysis: Recommendation and Implementation Plan.* LA-UR-03-4699. September.

LANL. 2002c. *Groundwater Protection Program Consolidated Data Management Workflow Analysis: Results*. LA-UR-03-4698. September.

LANL. 2002d. *Groundwater Protection Program Consolidated Data Management Workflow Analysis: Executive Summary*. LA-UR-03-4692. September.

Longmire, P.A., S.L. Reneau, P.M. Watt, L.D. McFadden, J.N. Gardner, C.J. Duffy, and R.T. Ryti. 1996. *Natural Background Geochemical, Geomorphology and Pedogenesis of Selected Soil Profiles and Bandelier Tuff, Los Alamos, New Mexico*. LA-12913-MS. May.

Meteorology and Air Quality (MAQ). 2003. *Statements of Work for Electronic Data Deliverables*. LA-UR-03-5648. Los Alamos National Laboratory, Los Alamos, New Mexico.

Mauro, J. 2003. *Recommendations for Radioecological Screening Levels for Terrestrial Plants. Health Physics*, Volume 84: 6. June.

McLin, S.G. and D.W. Lyons. 2002. *Background Radioactivity in River and Reservoir Sediments Near Los Alamos, New Mexico*. LA-13603-MS. May.

New Mexico Environment Department (NMED). 2000. *Guidance for Assessing Ecological Risks Posed by Chemicals: Screening-Level Ecological Risk Assessment*. Hazardous and Radioactive Materials Bureau. March.

Rocco, J.R., K.R. Meyer, H.J. Mohler, J.W. Aanenson, L. Hay Wilson, A.S. Rood, P.D. McGavran, and J.E. Till. 2002. *Analysis of Exposure and Risks to the Public from Radionuclides and Chemicals Released by the Cerro Grande Fire at Los Alamos. Task 2.7: Estimated Risks from Releases to Surface Water*. Final Report, *RAC* Report No.4-NMED-2002-FINAL(Rev.1). *Risk Assessment Corporation*, Neeses, South Carolina. June.

Rood, A.S., J.W. Aanenson, S.S. Mohler, P.D. McGavran, H.J. Mohler, H.A. Grogan, and J.E. Till. 2002. *Analysis of Exposure and Risks to the Public from Radionuclides and Chemicals Released by the Cerro Grande Fire at Los Alamos. Task 1.7: Final Report on Estimated Risks from Releases to Air. RAC* Report No. 3-NMED-2002-FINAL(Rev.1). *Risk Assessment Corporation*, Neeses, South Carolina. June.

Ryti, R.T., P.A. Longmire, D.E. Broxton, S.L. Reneau, and E.V. McDonald. 1998. *Inorganic and Radionuclide Background Data for Soils, Canyon Sediments, and Bandelier Tuff at Los Alamos National Laboratory*. LA-UR-98-4847. September.

Sandy Cohen and Associates (SC&A). 2000. *Guidance for Assessing Ecological Risks Posed by Radionuclides: Screening-Level Radioecological Risk Assessment*. Document created for the New Mexico Environment Department Hazardous and Radioactive Materials Bureau. April.

DRAFT

Water Quality and Hydrology Group (WQH). 2003a. *Quality Assurance Project Plan for the Watershed Management Program.* LA-UR-03-4230. RRES/WQH-QAPP-WMP, R. 0. Risk Reduction and Environmental Stewardship, Los Alamos National Laboratory. June.

WQH. 2003b. *EDD specification.* LA-UR-03-5087. Los Alamos National Laboratory, Los Alamos, New Mexico.