

ornl

ORNL/TM-13269

OAK RIDGE NATIONAL LABORATORY

LOCKHEED MARTIN



Implementation of a Consolidated, Standardized Database of Environmental Measurements Data

RECEIVED

OCT 29 1996

OSTI

MASTER

MANAGED AND OPERATED BY
LOCKHEED MARTIN ENERGY RESEARCH CORPORATION
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

ORNL-27 (3-96)

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37831; prices available from (423) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government of any agency thereof.

Computational Physics and Engineering Division

**Implementation of a Consolidated, Standardized Database
of Environmental Measurements Data**

T. L. James
Energy, Environment, and Resources Center
University of Tennessee
Knoxville, Tennessee 37996

B. C. Zygmunt
Computational Physics and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

Date Published: October 1996

Prepared by the
OAK RIDGE NATIONAL LABORATORY
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORP.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-96OR22464

*Operated by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract DE-AC05-96OR22464

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED



DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	v
LIST OF ACRONYMS	vii
ABSTRACT	ix
1. INTRODUCTION	1
2. OVERVIEW	2
2.1 BENEFITS OF CONSOLIDATION	4
2.2 RESISTANCE TO CONSOLIDATION	4
2.3 OVERCOMING RESISTANCE	5
2.4 DEFINING THE CONSOLIDATED DATABASE	5
3. STANDARDS	7
3.1 DATA ELEMENTS	7
3.2 CODES	8
3.3 COORDINATES	10
3.4 SITE-WIDE STANDARDIZATION	10
4. QUALITY CONTROL ISSUES	11
4.1 DATA INTEGRITY	11
4.2 DATA QUALIFIERS	11
5. TRANSMITTING DATA	13
5.1 FORMAT ISSUES	13
5.2 OREIS READY-TO-LOAD (RTL) FORMAT	13
6. DATA USERS	14
6.1 DATA AVAILABILITY	14
6.2 EDUCATED USERS	14
7. EVOLUTION OF A CONSOLIDATED DATABASE	15
7.1 USER EXPERIENCE	15
7.2 MANAGERIAL DECISIONS	15
7.3 STAFF EXPERIENCE	15
7.4 NEW DATA TYPES	15
7.5 CONFIGURATION CONTROL	16
7.6 PERFORMANCE TUNING	16
8. HISTORICAL DATA	17
9. CONCLUSION	18

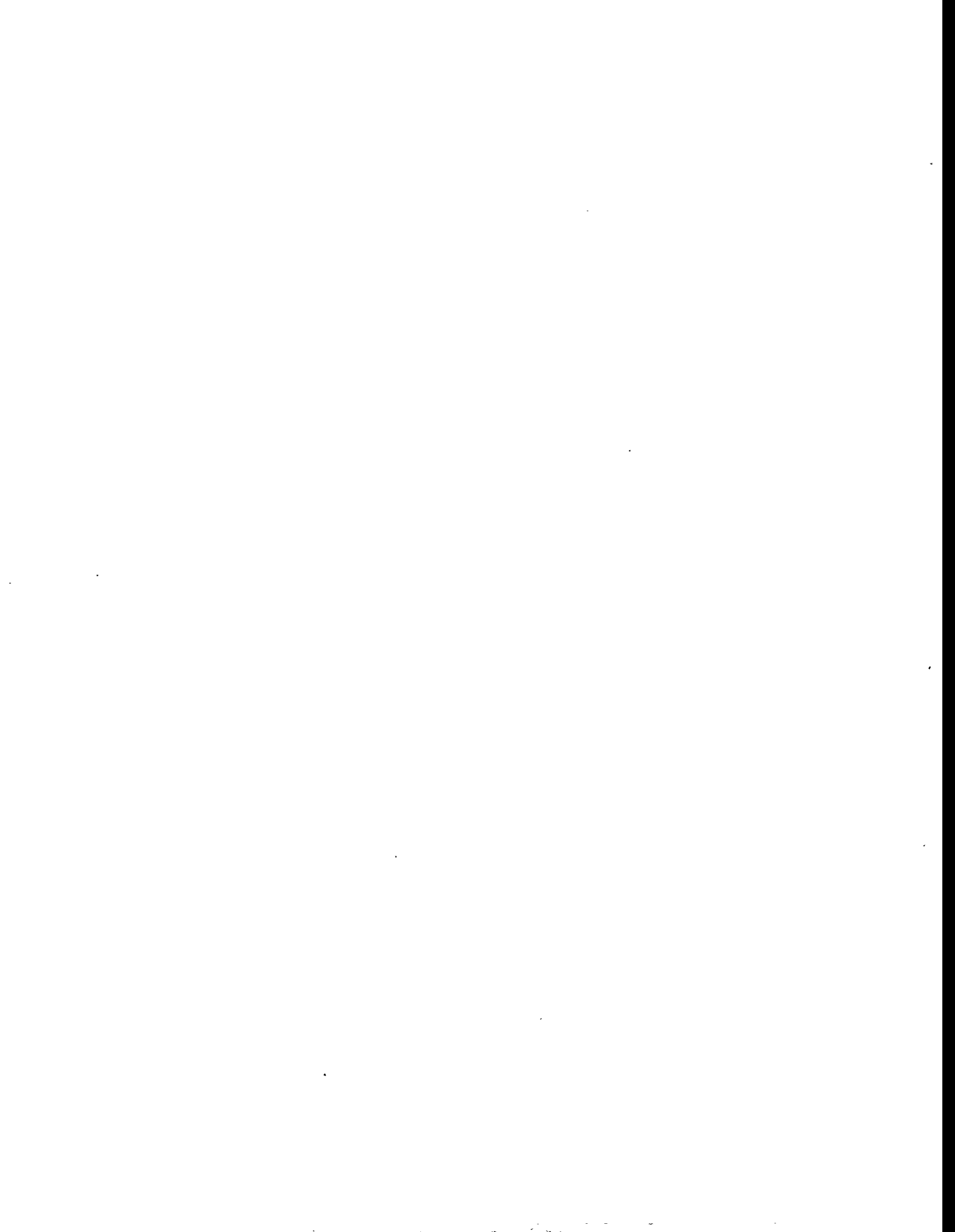
APPENDIX A : PROCESS FLOW DIAGRAM FOR THE PROCESS AND REVIEW OF DATA TRANSMITTED TO OREIS	19
APPENDIX B: PROCESS FLOW DIAGRAM FOR CHANGES TO OREIS DATA	23
APPENDIX C: RELATED BIBLIOGRAPHY	27

LIST OF FIGURES

Fig. 1. OREIS Data Sources	3
Fig. A. 1. Process Flow Diagram for the Process and Review of Data Transmitted to OREIS	19
Fig. B. 1. Process Flow Diagram for Changes to OREIS Data	23

LIST OF TABLES

Table. 1. Examples of OREIS Standardization	9
---------------------------------------------------	---



LIST OF ACRONYMS

CASRN	Chemical Abstract Services Registry Numbers
D&D	Decontamination and Decommissioning
DOE	U.S. Department of Energy
EPA	Environmental Protection Agency
ER	Environmental Restoration
GIS	Geographic Information System
NOAA	National Oceanographic and Atmospheric Administration
NPDES	National Pollutant Discharge Elimination System
OREIS	Oak Ridge Environmental Information System
ORNL	Oak Ridge National Laboratory
ORO	Oak Ridge Operations
ORR	Oak Ridge Reservation
PEMS	Project Environmental Management System
QA/QC	Quality Assurance/Quality Control
RAD	Radionuclide
RI/FS	Remedial Investigation/Feasibility Study
RTL	Ready-To-Load
S&M	Surveillance and Maintenance
TDEC	Tennessee Department of Environment and Conservation
USGS	United States Geological Survey
WWW	World Wide Web



ABSTRACT

The three U.S. Department of Energy installations on the Oak Ridge Reservation (Oak Ridge National Laboratory, Y-12 Plant, and K-25 Site) were established during World War II as part of the Manhattan Project that "built the bomb." That research and work in more recent years have resulted in the generation of radioactive materials and other toxic wastes. The U.S. Department of Energy's Oak Ridge Operations has been mandated by federal and state agreements to provide a consolidated repository of environmental data and is tasked to support environmental data management activities on the Oak Ridge Reservation and the Portsmouth, Ohio, and Paducah, Kentucky, plants. The Oak Ridge Environmental Information System (OREIS) was initiated to fulfill these requirements.

Most OREIS data are generated by projects that deal with environmental compliance, surveillance, and remediation activities. The primary use of OREIS data is to provide access to project results by regulators. A secondary use is to serve as background data for other projects.

This report discusses the benefits of a consolidated and standardized database; reasons for resistance to the consolidation of data; implementing a consolidated database, including attempts at standardization, deciding what to include in the consolidated database, establishing lists of valid values, and addressing quality assurance/quality control (QA/QC) issues; and the evolution of a consolidated database, which includes developing and training a user community, resolving configuration control issues, incorporating historical data, identifying emerging standards, and developing pointers to other data. OREIS is used to illustrate these topics.

1. INTRODUCTION

Lockheed Martin Corporation manages the U.S. Department of Energy (DOE) Oak Ridge Operations (ORO), which includes the three Oak Ridge installations, as well as the Environmental Restoration (ER) programs at the DOE plants in Portsmouth, Ohio, and Paducah, Kentucky. DOE ORO has been mandated by the Federal Facilities Agreement and the Tennessee Oversight Agreement to provide a consolidated repository of environmental data and to support environmental data management activities at all five installations. The Oak Ridge Environmental Information System (OREIS) was initiated to fulfill these requirements.

OREIS provides both state and federal regulators with access to standardized, consolidated environmental data. The data may be accessed through a standard tool-set, a World Wide Web (WWW) interface, standard reports, and user-defined requests.

OREIS also serves as a long-term repository for these data and provides access to a large user community. This community includes project staff, subcontractors, and interested third parties.

There are some basic principles and specific issues that have been fundamental to the success of OREIS. Future consolidated database efforts can learn from and build on the OREIS experience. This report begins with a discussion of consolidated database principles and then focuses on key concerns. OREIS is used throughout the report to illustrate these topics.

2. OVERVIEW

The purpose of a consolidated database is to bring together and unify dissimilar data. According to the dictionary, *consolidation* is "to combine into one, to make strong and stable." This combination may include information stored in different formats and media, data that were collected for different purposes, or data maintained at different sites. A consolidated database can be a database of environmental measurements located on separate, but networked, machines or on a single computer. The data may originate from a collection of sampling information files, electronic or otherwise (e.g., log books) or pointers to files; tables with disparate or similar structures in project-level database management systems; or raw measurement data.

The definition of consolidation includes becoming "strong and stable." A consolidated database gains strength and stability because it encourages the use of a single set of tables, files, or pointers rather than permitting multiple copies of data, which are difficult to maintain. But the true strength of most consolidated databases comes from the standardization of the data. In an environmental measurements database, this standardization may include, but is not limited to, resolving different naming conventions for chemical compounds, standardizing date formats, and requiring a minimal set of specific fields used to identify and define measurement results.

In OREIS, this consolidation involves joining tabular data from sources such as environmental restoration projects, compliance programs, risk analysts, state agencies, and national monitoring organizations. (See Fig. 1.) OREIS makes these data available at a centralized location and in a standardized format. The data are available to remote as well as local users. These data traditionally have been stored in a variety of electronic formats or as log-books and other "hard copy" materials. The data could be viewed only by a very limited number of persons. A person in search of data to support ongoing work would have to ask a number of questions:

What data exist?
Where are the data?
Are there hard copies?
Can I get to them?

Who would know?
Are the data in electronic format?
Can the floppies be read? By what software?

2.1 BENEFITS OF CONSOLIDATION

On the Oak Ridge Reservation (ORR), the consolidation of the data is especially important because of the general trend towards combining similar functions of the three separate plants. Having the data on-line in a single relational database management system supports this new initiative. Data transmittals to OREIS must be standardized to fit into OREIS, and this standardization gives data for the entire reservation more of the same "look and feel" to the users of the data.

Another reason to implement a consolidated database is economics. It is more expensive to maintain duplicate sets of data than a unified data repository. Additional resources must be committed to keep the data consistent when they are stored in separate systems. Costs associated with data storage include hardware and software purchases and maintenance. But the most significant expenditures are for the personnel needed to create, update, and maintain these redundant systems. A related expense is the cost to the user of extracting and consolidating data from separate projects. Each user of the data probably will not be able to build on previous work and will have to start from the beginning of the extraction and consolidation process. The expense

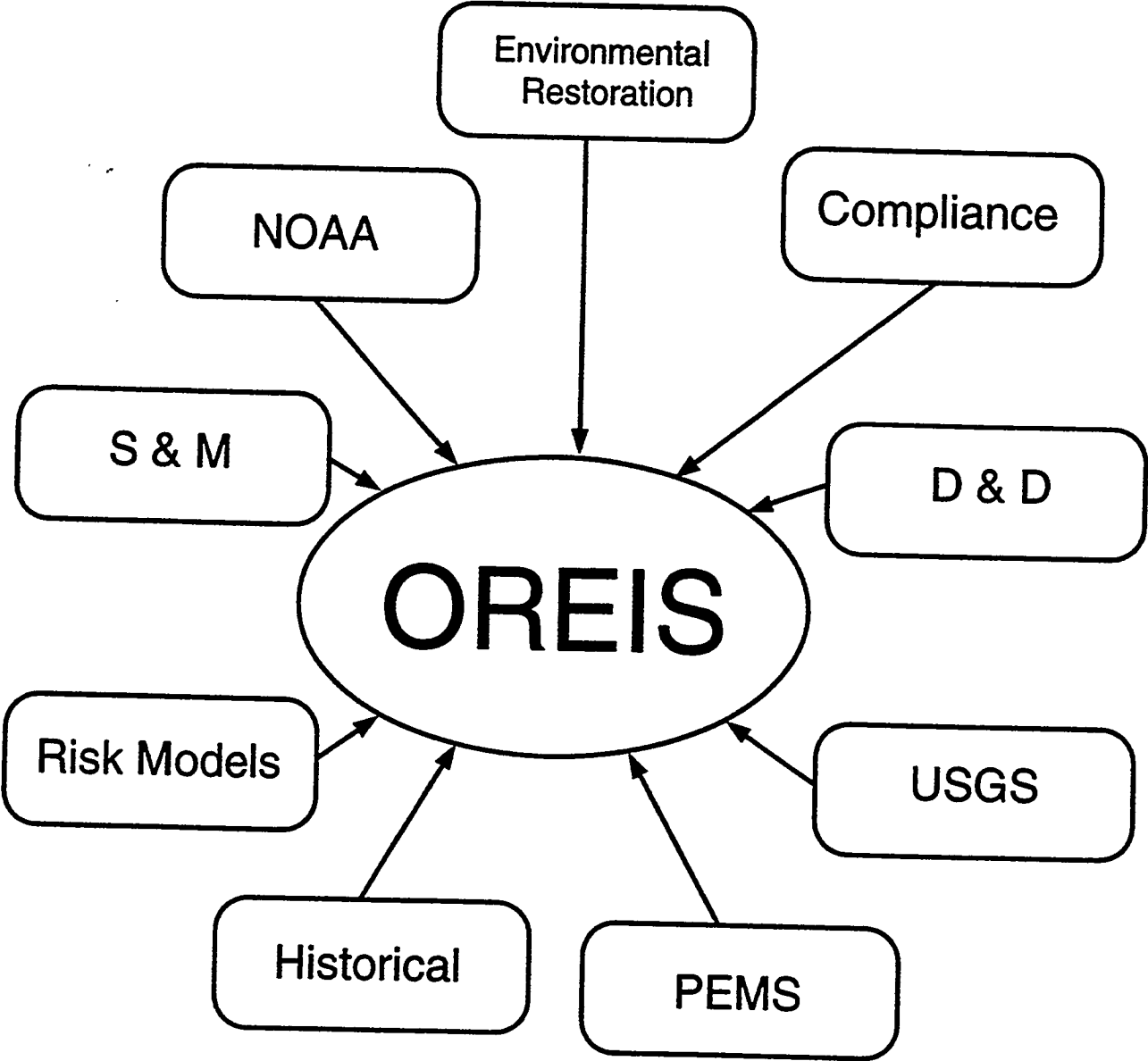


Fig.1. OREIS data sources.

of this duplication and the likelihood of errors and data inconsistencies multiply with each iteration.

Consolidation also improves the quality of the data. OREIS looks for anomalies and corrects them across datasets. For example, OREIS checks to make sure all longitudinal values are negative, as they should be in the western hemisphere.

Another major advantage of data consolidation is the ability to replicate analyses. In the past, reports and the analyses they contained could not be reproduced exactly. This inadequacy led to questions about the validity and accuracy of the data. Regulators and others now will be able to generate the analyses contained in reports from a single data repository.

Previously, the absence of unifying characteristics, such as standard codes and formats, made it difficult for projects to share data and required reports to be rewritten for each project. Even within projects, there was often a lack of standardization and consolidation seen in the form of multiple data structures, different codes that mean the same thing, and duplication of records resulting from, for example, project staff maintaining individual project datasets. One of the main functions of OREIS is to provide a standardized structure and set of valid values that can accommodate data from a variety of projects. These valid values frequently are stored as codes, such as codes for units of measure, specific analyte types, and sampling devices. This feature benefits the users by providing them access to data in a consistent format using standard codes.

2.2 RESISTANCE TO CONSOLIDATION

Despite the benefits, there is often resistance to the consolidation of data. Most people initially are opposed to changes in the way they do business, especially if they feel their input has not been given enough consideration. OREIS was mandated by state and federal agreements. The providers of the data were not given the opportunity to evaluate whether or not a consolidated database should be implemented, thus adding to the initial resistance to the consolidation effort.

In addition, data providers often feel ownership of their data. This ownership is expressed by a desire to maintain control of and access to the data. Historically, when projects were smaller and their associated data did not need to be shared with other projects, this method was the way business was done. That viewpoint, however, is not consistent with the consolidated approach.

Another reason for resistance is that consolidation will lead to downsizing/rightsizing, with fewer people needed to work with the data. Some of this work, which includes updating and maintaining the data and answering user requests, will be taken over by the consolidated database staff and will mean fewer jobs at the project level.

And finally, although saving money is a benefit of consolidation, the costs to the projects of "retrofitting" the data to match the consolidated database structure can be substantial. This retrofitting includes format changes, code revisions, keeping data that were never kept before, and putting data into an electronic format.

2.3 OVERCOMING RESISTANCE

OREIS has managed to overcome many of these objections by showing data providers and project staff how the implementation of a consolidated database can be used to benefit them by reducing data management tasks, responding to user requests, providing standard codes and formats, and creating a larger context in which consistency checks of the data can be performed. One of the most critical data management tasks assumed by OREIS is configuration control. This

step includes establishing a procedure to initiate data changes, notify users of proposals, incorporate user feedback, implement the changes, and document the process. Other data management tasks assumed by OREIS include data archival, data backup and recovery, system and database administration, and providing access to the data. OREIS supports data users with a wide range of computer-skill levels. These users include project managers who may have very little computer experience, regulatory agency personnel with some computer proficiency, and project staff who are skilled computer professionals. A point-and-click interface is available to novice users along with training, on-line documentation, and staff support. In addition to these services, the more experienced user may access the data directly using the suite of software available on the OREIS computers.

Although using standard codes and formats may not provide direct benefits to an individual project that is collecting samples, if the data are viewed in a larger context with other projects, data analysis and consolidation will be simplified. Similarly, consistency checks of the data within a project reveal certain types of anomalies, but checks across projects may reveal other discrepancies.

An unexpected result of OREIS is that ownership of the data has been redefined. Even though the data no longer reside on project-controlled hardware and the data provider is freed from answering requests and doing maintenance, the data cannot be changed without the permission of the data provider. In this sense, the provider still owns the data, even though the data are part of OREIS.

2.4 DEFINING THE CONSOLIDATED DATABASE

One of the first steps in implementing a consolidated database is to establish the scope of the contents. At the highest level, OREIS includes data that can be of use to secondary users and that are used to draw an accurate picture of the environment at any given time, including analyzing human health and ecological risk in the environment. A decision was made that OREIS would include the following: data that have been cleared for public release (i.e., no classified data); summarized data, when appropriate (e.g., daily flow readings summarized from results taken at 15-minute intervals) and data that are required by the U.S. Environmental Protection Agency (EPA) and DOE and to support Federal Facilities Agreement milestone projects. More specifically, OREIS receives data from current and historical Environmental Restoration (ER) projects, including decontamination and decommissioning (D&D), surveillance and maintenance (S&M), and remedial action/feasibility study (RI/FS) data; and compliance data from the ORO. The compliance projects include the National Pollutant Discharge Elimination System (NPDES) and programs that monitor air emissions, surface water, radionuclides (RADs), groundwater, biota, and storm water.

For each of these types of projects, OREIS contains geographic information system (GIS) data, field and laboratory measurements, and metadata. The GIS data include line (vector) and image (raster) data. OREIS vector data include roads, rivers, and buildings; one type of OREIS raster data is the aerial photographs of the five plants. Field measurements include results of analyses performed in the field, such as pH and temperature. As their name implies, laboratory measurements data are generated by the analyses performed by laboratories on samples taken at points of interest.

One way both types of measurements are grouped across projects is by media type. For example, someone might be interested in all the groundwater samples at K-25. The measurements in OREIS contain a variety of media types including the following:

- groundwater,
- surface water,
- soil,
- sediment (associated with surface water),
- sludge,
- effluent air,
- ambient air,
- biota,
- drill cuttings,
- filter residue,
- drilling fluid,
- porewater,
- scrapings,
- swab or wipe,
- solid waste,
- equipment wash water, and
- wastewater.

Other types of measurements include results of samples taken from storage tanks, control samples (i.e., field blanks, trip blanks, and rinsates), and meteorological data [e.g., National Oceanographic and Atmospheric Administration (NOAA) and United States Geological Survey (USGS)]. In addition to analytical results, OREIS contains construction information for tanks, wells, boreholes, coreholes, and location/station information such as coordinates and station names.

Another data type in OREIS is a set of human risk models and the data needed to run them. These data are not accessible to the entire OREIS community, but to a subset of approved users. Some proposed future directions for OREIS include photographs, well-construction logs, health physics data, waste inventories, and sampling data from the Tennessee Department of Environment and Conservation (TDEC).

Metadata, which are data about data, are an important component of OREIS. The data transmittal form provides the minimum metadata for a dataset. This information includes the names of the data provider and a project contact, the name of the project, cautions about using the data, the project site, the dates covered by the data, and reports associated with the data. In addition, metadata may include a copy of the report associated with the data and other relevant information.

As the data types were defined, a software development methodology called data modeling was used to determine the "best" structure for information storage and retrieval. As OREIS evolves, the data modeling process continues to accommodate the changes.

3. STANDARDS

A variety of standards must be established before a consolidated database can be implemented. These may include the following:

- a minimum set of data elements required to identify the various data types,
- standard units,
- valid analytes and methods,
- other codes, and
- a set of coordinate systems for representing geospatial data.

These standards were established in OREIS through a multiphase process that included reviewing existing environmental database codes and structures, implementing modules from these where possible, adopting ORR code standards, and conducting extensive modeling sessions with environmental professionals. OREIS continues to work with data users and subject matter experts to refine and expand OREIS standards.

3.1 DATA ELEMENTS

OREIS contains different data types (e.g., NPDES, biota), and each type may require a unique set of data elements (also known as fields, variables, or columns). NPDES data results must be associated with a significant digit, because the regulators who monitor the results must have this information to interpret the values correctly. Most other projects do not report significant digits. Biota data, unlike all other types of measurements in OREIS, are not required to include a numeric results value. Instead, biota analyses may be reported as categorical results, such as "E" ("Enlarged") or "F" ("General discoloration").

Although some data elements are specific to a particular data type, there are others that are common to all types. These common elements can be grouped according to the kind of information they contain, as follows:

- project
 - project name
 - project site
- station/location
 - station name
 - station type
 - coordinates and coordinate system
- field measurement
 - date
 - media type
 - analyte
 - results
 - units
- laboratory measurement
 - date

- media type
 - sample type
 - laboratory identifier
 - sample matrix
 - analysis method
 - analyte
 - results
 - units
- well construction
 - well type
 - well component type and material
 - discontinuity type and depth
 - monitored interval name, depth, and material
 - geologic formation identifier

Information contained in these data elements is used both to identify individual records in the database and to group similar data (e.g., all data from a particular media, such as groundwater). Standardization of coded values in OREIS makes it much easier to do such grouping across projects.

3.2 CODES

Codes are "abbreviations" that represent frequently used values (e.g., "AA" for "Ambient Air"). In addition to saving space in a database, the use of codes enforces a standard way of representing data elements. Failure to implement coded values may result in data that look different but mean the same thing. Some frequently coded data elements include the following:

- media type,
- sample type,
- station type,
- laboratory identifier,
- well component,
- analytical method,
- units,
- coordinate system, and
- sampling device type.

Because one of the goals of a consolidated database is to provide the same look and feel to the data, standardization of codes is important. Table 1 gives examples of different coded values that map to a single OREIS code.

In order to further the ORR standardization process, OREIS is leading an effort to define a standard unit for each unique combination of analysis type, analysis method, media type, and analyte. This has always been a goal of OREIS, but it has been difficult to implement because of the large number of valid combinations. Recently, OREIS has defined an initial set of relationships among three of these data elements: analysis type, analysis method, and analyte. Adding units to this relationship matrix is particularly difficult because some users have a regulatory requirement to maintain the original units and results in their data. Because OREIS

does expect to complete this matrix at some point, OREIS has two sets of result/unit fields: one for the original data and one for the OREIS standardized values. For example, the OREIS standard could be to maintain water flow measurements in gallons per day, while a regulatory driver may require that the data be reported as liters per hour. The OREIS data model allows both to be retained.

Table 1. Examples of OREIS standardization

OREIS name	OREIS code	Other name	Other codes
ANA_METHOD	EPA-624	ANALYSIS METHOD ANALYSIS_METH METH	EPA-624-M EPA-624 Jul82 CLP-624 624
PARAMTR	100027	ANALYTE CHEMICAL	<i>p</i> -Nitrophenol Phenol, 4-nitro- 100-02-7
UNITS	mg/kg	UNIT MSMT_UNIT DEPTH_UNIT	MG/KG milligram per kilogram mg/KG
MAT_TYPE	GRVL	MATERIAL MAT COMP_MAT	Gravel G GL
LAB_CODE	ORNL	LAB ANA_LAB	O X-10

Two other codes that must be standardized are analytes and methods. The original list of valid analytes was obtained from an ORR standards committee and was based on the Chemical Abstract Services Registry Numbers (CASRNs). Many analytes of interest to OREIS were outside the scope of this committee and did not appear in its list. OREIS followed the committee's format when adding analytes, using CASRNs when available or generating nonstandard identifiers when required. Periodically, the OREIS analyte table is reviewed to eliminate multiple listings of an analyte (due to chemical synonyms).

The OREIS methods have a similar history. They began as a list obtained from an ORR standards committee. When required by projects, OREIS adds methods following the original format. The standardization of methods is a critical and difficult effort. Most methods originate from an EPA or other standard method; however, individual laboratories may modify or rename a particular method. The problems in trying to standardize these methods include nonstandard naming conventions (e.g., failure to include a prefix in the method number), conflicting method definitions, inconsistent tracking of method revisions, and incomplete method numbers and definitions.

3.3 COORDINATES

OREIS includes a geospatial component. To implement this component, coordinates for sampling points must be standardized in order to display them across projects and overlay them on OREIS images. On the ORR, there are five different northing/easting grid systems (K-25, Y-12, ORNL, administrative grid, and state plane) used to locate sampling stations. OREIS converts the original coordinates provided by the project to state plane, administrative grid, and latitude and longitude. These converted values, along with the original coordinates, are stored in the OREIS database.

3.4 SITE-WIDE STANDARDIZATION

OREIS staff participate in several ORO-wide standardization committees that are working toward these same goals: standard units, analytes, methods, and other codes. The focus of these committees is to establish valid values and formats to represent environmental data. Each group comprises professionals from a variety of backgrounds and organizations, which may have their own sets of standards. The challenge is to integrate these discrete sets into a unified way of doing business. OREIS, other data systems, and data users have benefited from these initial steps toward ORO-wide standardization.

4. QUALITY CONTROL ISSUES

Two broad categories of QC issues must be addressed in a database management system: the integrity of the data once they are in the system and the flags used to explain and qualify individual results. The first is an issue of maintaining the data once they are in the consolidated database and ensuring that changes are carried out only in a prescribed manner. The second concerns the interpretation of the data using data qualifiers.

4.1 DATA INTEGRITY

Data integrity is maintained in OREIS through a set of procedures and practices that address the following:

- awareness of possible security risks associated with the use of the database management system,
- limiting access and privileges of users,
- internal data processing, and
- configuration control of both the data and the data structures.

Because of inherent security concerns with the type of database access provided by OREIS, a security bridge has been placed between the OREIS database server and the rest of the world. This configuration allows connections to the database only from "trusted" (approved) workstations. Access to OREIS is managed with a procedure that requires potential users to go through an approval process that includes obtaining a Lockheed Martin Energy Systems computing account. As part of the approval process, each user is assigned appropriate resources and privileges. For example, although all OREIS users can read the data, only the database administrator can change it.

Internal data processing procedures ensure that any transformations to the data are documented and reviewed before being sent to the data provider for approval. Appendix A, "Process Flow Diagram for the Process and Review of Data Transmitted to OREIS," is a process flow diagram of the procedure for the process and review of data transmitted to OREIS. Once the data are loaded in the OREIS database, they are under configuration control. All changes must be approved by the data provider. Appendix B, "Process Flow Diagram for Changes to OREIS Data," illustrates the steps required for changing OREIS data.

Two additional elements under configuration control are the data model and coded values. These elements undergo review by the user community before they are implemented to assess the impact of proposed changes.

4.2 DATA QUALIFIERS

One of the primary mandates of OREIS is to store data of "known" quality and make those data readily available to environmental professionals, state and federal regulators, and the public. In order to avoid misinterpretation of measurement results, the results must be evaluated in conjunction with the qualifiers. Three types of qualifier information are contained in the OREIS database. One type of qualifier indicates the usability of the data. Another type of qualifier

contains codes that document conditions associated with the results of the analysis. The third indicates the results of the validation process.

OREIS has a policy not to store data that were qualified as "rejected" during the validation process. It is assumed that these data were not used in making decisions. In addition, there is the potential for using a "bad" data result out of context and drawing an erroneous conclusion. However, issues have arisen over time that have caused OREIS to review its policy regarding "rejected" data. These issues are as follows.

- OREIS is mandated to keep data that are used in a Federal Facilities Agreement deliverable or are used to support the conclusions contained in the report; the "rejected" data may have been used to support some portion of the report.
- It is possible for data to be rejected by a project yet still be useful to secondary users. For example, a project may collect data and require definitive-level results. Another project that only has a need for screening-level data could accept data the first project rejected.

Another ongoing question is whether OREIS needs to keep any QC data that are used to calibrate the laboratory equipment. Potential users and risk assessors may need these data in order to determine usability.

5. TRANSMITTING DATA

Initially, OREIS would accept data in any electronic format. This decision was made to encourage people to send data to OREIS. Even though this approach made sending data easier for the data generators, it made OREIS processing very difficult and more likely to result in errors. Now, OREIS accepts data only in a limited number of formats. In addition to requiring specific electronic formats, OREIS has defined data structures within those formats that make the data "ready-to-load."

5.1 FORMAT ISSUES

The first data packages received by OREIS took months to process. Many problems were encountered in trying to map the data into the OREIS structure: Fields had the same name but meant something different; fields with dissimilar names meant the same thing; analytes and analysis methods did not match the OREIS list of valid values for a variety of reasons, including spelling, capitalization, synonyms, and revision numbers; there were missing values for required fields; QC samples were difficult to match to regular samples; and inconsistencies, such as outliers, unit conversion errors, duplicate records, and missing values, existed within the dataset. In response to these issues, OREIS developed a specification for transmitting data. Most of the responsibility of preparing the data to OREIS specifications now has been passed to the projects: the project staff are the most familiar with the data and best qualified to resolve questions.

5.2 OREIS READY-TO-LOAD (RTL) FORMAT

The RTL is the specification for transmitting data to OREIS. It includes descriptions of the OREIS fields, names of the required fields, file structures, and a model for linking QC and regular samples. Data provided in RTL format can be processed by OREIS staff in days, rather than months, because there are fewer issues to resolve: Project staff have already made the decision about how to map fields, values have been assigned for all required fields, and the links between QC and regular samples have been defined. Even with the RTL and OREIS lists of valid codes, problems with the data may still exist. Frequently, the problems are caused by misinterpretation of the RTL or incorrect usage of the codes.

Even though the RTL structure must evolve because of changes in the database model, an important goal is to keep the RTL as static as possible. In order to minimize the impact of model changes on the data generators, OREIS staff update the RTL only when changes to the model are significant. Other changes can be handled internally in the processing of the data. Current RTL documentation and code lists are available on the WWW via the OREIS home page.

6. DATA USERS

Initially, the OREIS user community was defined by the federal and state mandates to create a consolidated repository for environmental data. This community consisted of regulators at both the state and federal level and DOE. The user community has expanded to include risk modelers, ER project and compliance program staff, subcontractors, and local researchers.

6.1 DATA AVAILABILITY

Use of the data has grown as more data have become available. Scientists, analysts, and other secondary users recognize OREIS as a source for data. In addition, OREIS has provided data to two out-of-state groups working on projects to re-evaluate the contaminant exposure to people who worked on or near the ORR 20 years ago. These groups came to OREIS because OREIS can provide a variety of data from many projects in a standardized electronic format.

OREIS is continuing to expand its user community as more and new types of data are loaded. Users can access data supplied by a variety of projects and, because of the standardization of the data which has taken place, the same variables, codes, and definitions are found in each. The NPDES and ambient air compliance data are being supplied regularly to OREIS. As a result, these groups will generate their parts of the next ORR annual report from data within OREIS. In addition, current projects are generating reports from the data in OREIS. These data may include new sample data and information from historical projects.

6.2 EDUCATED USERS

An educated user community is a requirement for a successful consolidated database. Users may become frustrated if they have not been given all the information necessary to retrieve the data and interpret them correctly. One of the things a user must know is the relationships among the data fields. For example, in OREIS the user must examine both the result and result qualifier fields when evaluating results because the qualifier can indicate problems with a given result. Similarly, in the case of NPDES data, if the significant digit value is not applied to the result, it can appear that a result is out of compliance when it is not. In addition, users need a basic understanding of the data model and a general understanding of the types of fields that are coded and the code values.

To ensure that users have the information they need to make use of OREIS data, OREIS offers training on a variety of subjects. These include an introduction to the data model and hands-on experience with the OREIS software. OREIS also offers modules on each of the OREIS procedures. In addition, OREIS maintains electronic information, accessible through the WWW and file transfer protocols, about the valid values for coded fields, the data model structure, forms, reports, RTL file formats, hints, and answers to frequently asked questions. In an evolving system, it is a challenge to keep the users updated on all of the changes. One of the ways OREIS does this is through an electronic news group. Each proposed change is posted to the group, and comments about the change are shared among the users. User input is the critical factor in deciding whether or not a proposal will be implemented.

7. EVOLUTION OF A CONSOLIDATED DATABASE

As a consolidated database evolves, a number of issues arise. Decisions must be made about the direction of the evolution. Users, managers, and staff will have input on these decisions that must be addressed. As new types of data are identified, modifications to the code tables, the model, and documentation may be needed. Changes to the codes and model can require the retrofit of existing data and can affect the way in which data generators submit data. Because of the continual growth and evolution of a consolidated database, it is essential that the system be under configuration control and that all changes are communicated to data providers, regulators, and database users. The growth of the database may affect performance as more users access the data and the number of records increases.

7.1 USER EXPERIENCE

User experience contributes to the evolution of a consolidated database. As the users explore the system and become familiar with its capabilities, they are able to envision additional functionality and new data sources. This experience helps provide direction for the growth and expansion of the database system.

7.2 MANAGERIAL DECISIONS

Evolution also is driven by managerial decisions. These decisions can expand the scope of the database, such as the choice to include D&D data. These data initially were not part of OREIS, but with the changing funding environment, management opted to integrate D&D data into OREIS. Another management decision required the design of an interface between OREIS and the Project Environmental Management System (PEMS), which is a project-level database management system that was already sending data to OREIS. The interface was intended to provide more real-time access to project data by the regulators. The managerial decisions with the greatest impact are financial ones. Currently, these are typically a reduction in resources.

7.3 STAFF EXPERIENCE

Because of their familiarity with the database, OREIS data management staff contribute to the evolution of the database. They are able to recognize potential problems with proposed revisions to the model, inconsistencies in codes and values across projects, performance problems in retrieving data, and candidates for standard data products.

7.4 NEW DATA TYPES

OREIS began as a repository of ER projects data and has evolved to support other programs. Some new data types include measurements from the ambient air monitoring program and underground storage tank construction information. With most new types, it was necessary to add codes, change the model, and modify the rules for data submission.

As OREIS has evolved, other types of data have become available. For example, several projects have digitized well and bore logs; other projects have created WWW pages with project descriptions and photos. These data are of interest to OREIS users. For a variety of reasons, especially to avoid redundancy, it may not be possible or desirable to put the data into OREIS format. OREIS points to these other sources via its WWW home page and its WWW-based user interface.

7.5 CONFIGURATION CONTROL

As new types of data are brought into the database, modeling enhancements may need to be made. Code lists may expand. New applications may be required and current applications may have to be modified. Documentation may need to be updated or new procedures may need to be added. Configuration control, which is always an issue, becomes even more critical. Virtually all components of OREIS are under configuration control, including the following:

- data model;
- valid values;
- project data;
- documentation;
- software, commercial and customized; and
- system archival and backup strategies.

A part of configuration control is the need for keeping the data current with the model. All retrofits involve basically the same steps.

- Summarize what is currently in the database. The statistics generated in this step will be used to verify the transformation of the data.
- Determine how to convert data from the old structure to the new.
- Write programs to move data from the old structure to the new.
- Move a test set of data into the new model.
- Verify that the transformation of the test set is correct.
- Move all of the data into the new structure.
- Verify that the transformation is correct by using the statistics from the first step.

7.6 PERFORMANCE TUNING

As the database and number of users grow, performance tuning becomes crucial. The OREIS database model includes many indexes that were designed to enhance system performance. Sometimes indexing is not enough, and individual queries must be optimized. Query optimization is a very complex process that requires training beyond that of most users. OREIS staff are available to provide assistance in this and other technical areas. In addition, hardware and software upgrades may be required to improve database response and to provide for more users of the system.

8. HISTORICAL DATA

A consolidated database seldom has the choice of collecting new data exclusively; almost always there are historical records that must be accommodated. In nearly all cases, historical data are more expensive to reformat to conform to the consolidated structure because they often do not meet current database specifications. Even data that were collected six months ago may not meet today's standards. At OREIS, historical data are subjected to less stringent requirements than the data from current projects. For example, analytical methods are not required for historical data if the methods were not kept or are irretrievable. Other required fields may not be available. In addition, it may be difficult or impossible to determine the meanings of some of the codes, causing the loss of part of the original information contained in the data. But there is a wealth of information in historical data, and it is worth the effort to accommodate them.

Another issue that arises with historical data is how to deal with duplicate records. If two projects reference information from the same source and both projects use the data in a report, both projects may send these data to OREIS. In order to avoid the need to store duplicate measurement records, changes to the model have been made so that a single result can be associated with multiple projects. In this type of structure, in which a record can be "owned" by more than one project, configuration control issues become more complex. For example, if one project wants to change a "shared" result, all the data providers involved must be notified and must approve the change.

9. CONCLUSION

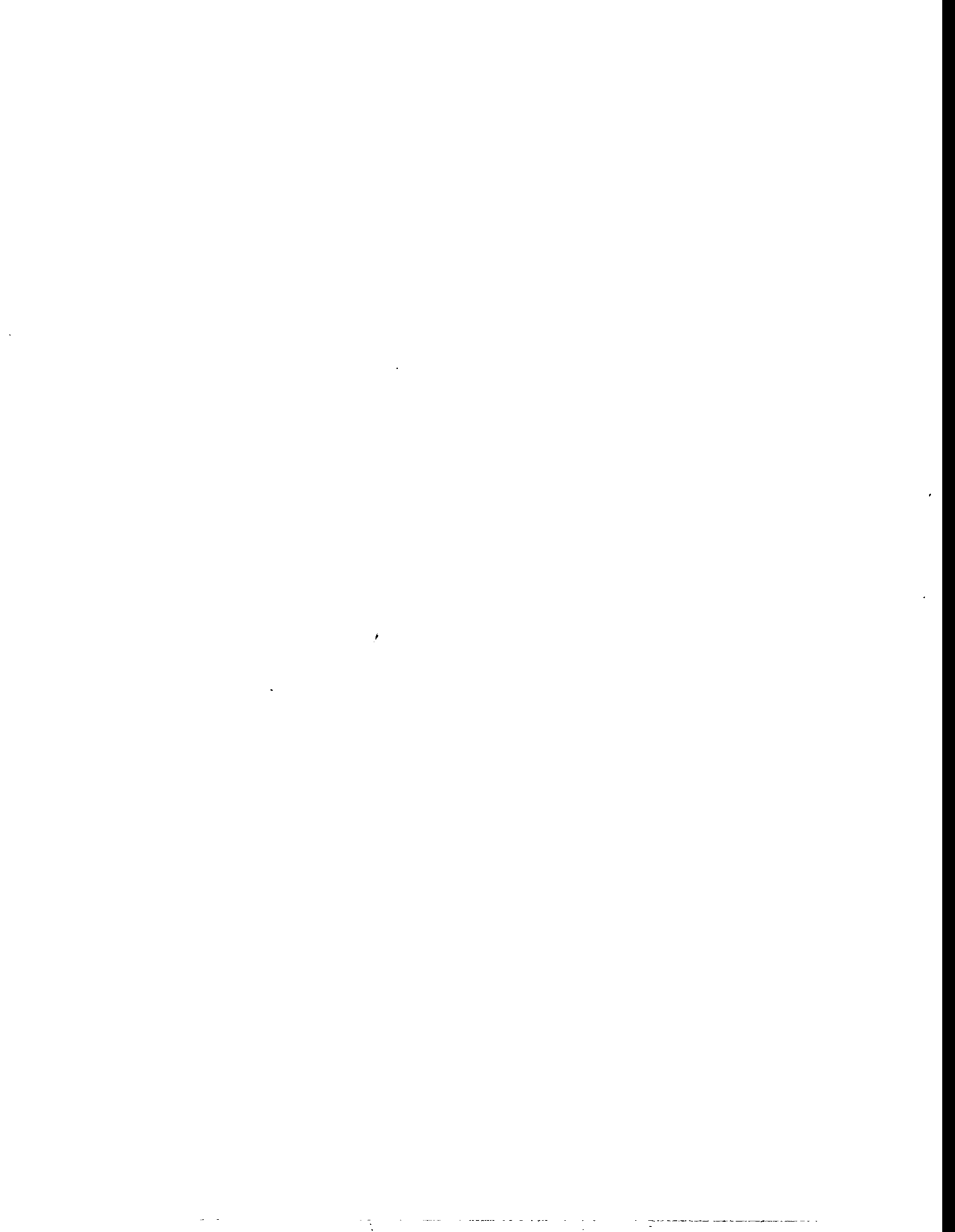
Considerable effort has been required to consolidate data in OREIS, but the benefits are being recognized. The data from both ongoing and historical projects can be accessed electronically in a single location. Codes have been standardized to give the data the same "look and feel." As users have become familiar with OREIS, they are more involved in its development. Their concerns range from comments on proposed codes to opinions on the scope and direction of OREIS. With their increased involvement often comes a feeling of ownership and a desire to see OREIS succeed.

OREIS still challenges. Consolidation and standardization will continue to be issues as more data from varied sources come to OREIS. Secondary use of data is growing in importance as projects rely less on new sampling efforts and base decisions on historical data that are easier now to access. User involvement will continue to be critical to the success of OREIS, and new technology, such as the WWW, will influence future OREIS directions.

Even though OREIS is proving to be a successful implementation of a consolidated database, the success brings new issues that must be addressed. OREIS has grown larger and faster than anticipated. Controlling this growth in a way that keeps OREIS responsive to the needs of its existing user community while accommodating new user requirements is a challenge. In contrast to the increases in scope and the amount of data that must be processed and loaded, OREIS is experiencing reductions in budget and staff. These reductions mean that OREIS may not be able to afford all the refinements and modifications to the database and software identified by users.

APPENDIX A:

**PROCESS FLOW DIAGRAM FOR THE PROCESS AND REVIEW OF DATA
TRANSMITTED TO OREIS**



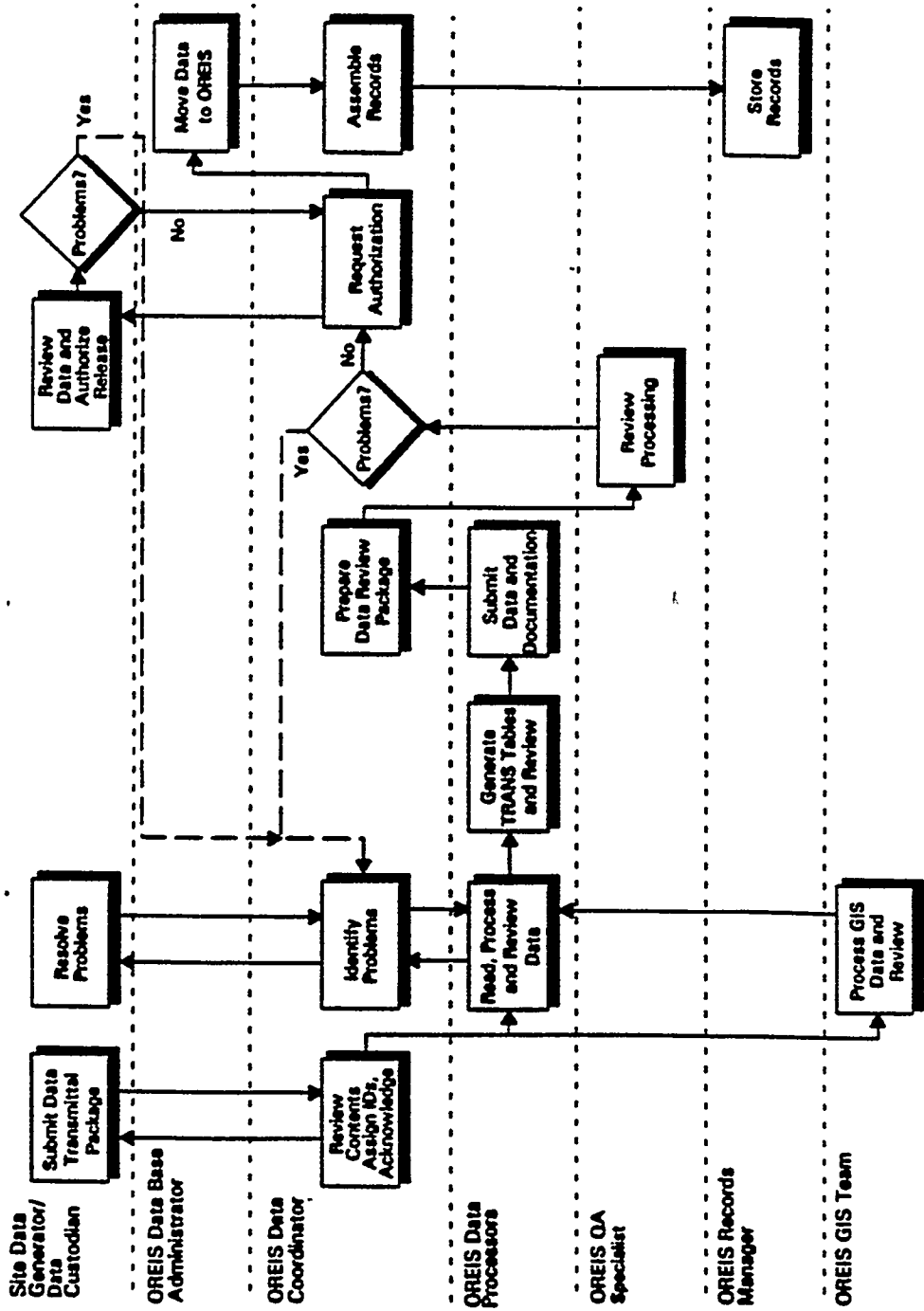
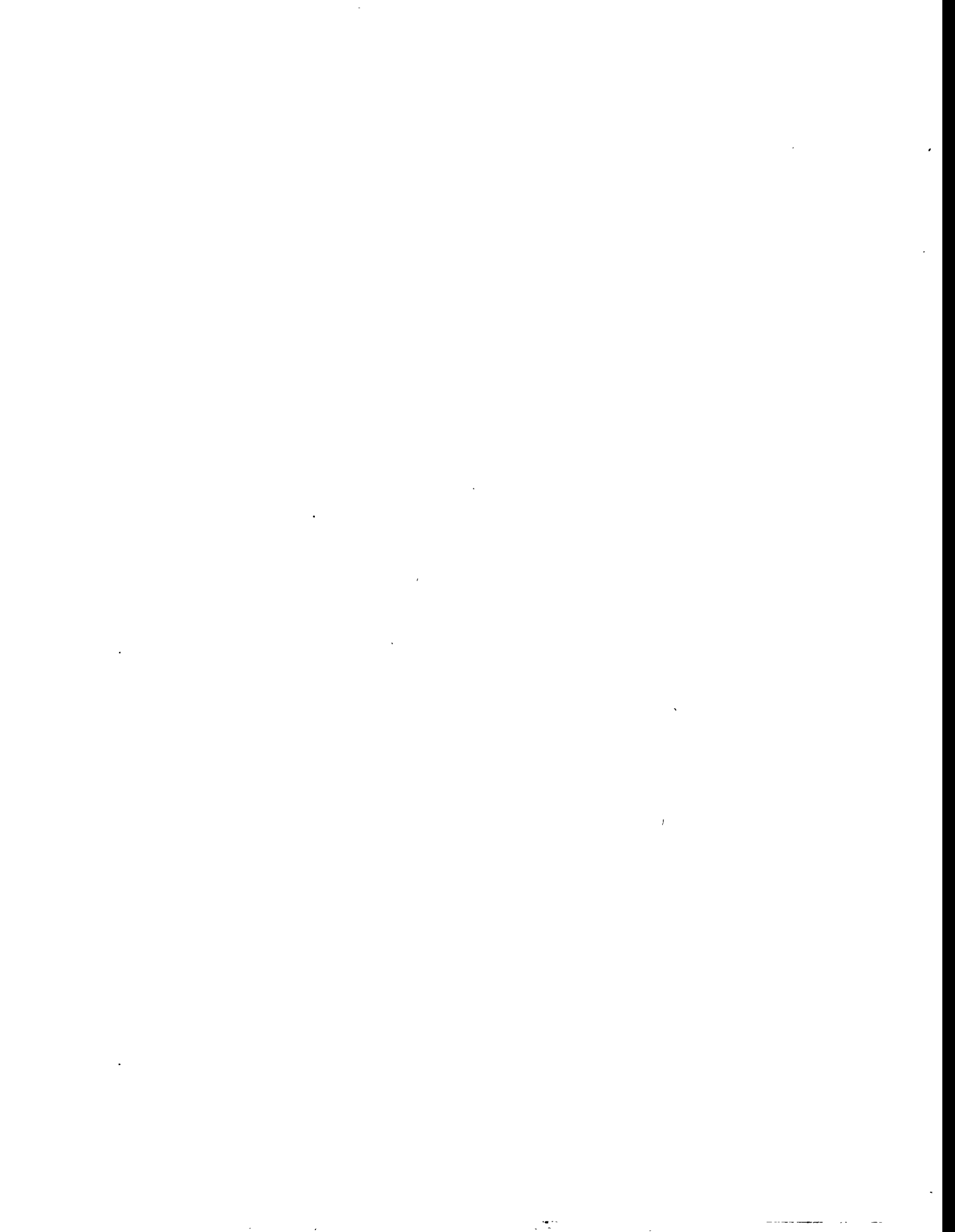


FIG. A.1. Process flow diagram for the process and review of data transmitted to OREIS.



APPENDIX B:

PROCESS FLOW DIAGRAM FOR CHANGES TO OREIS DATA



REQUESTER/USER

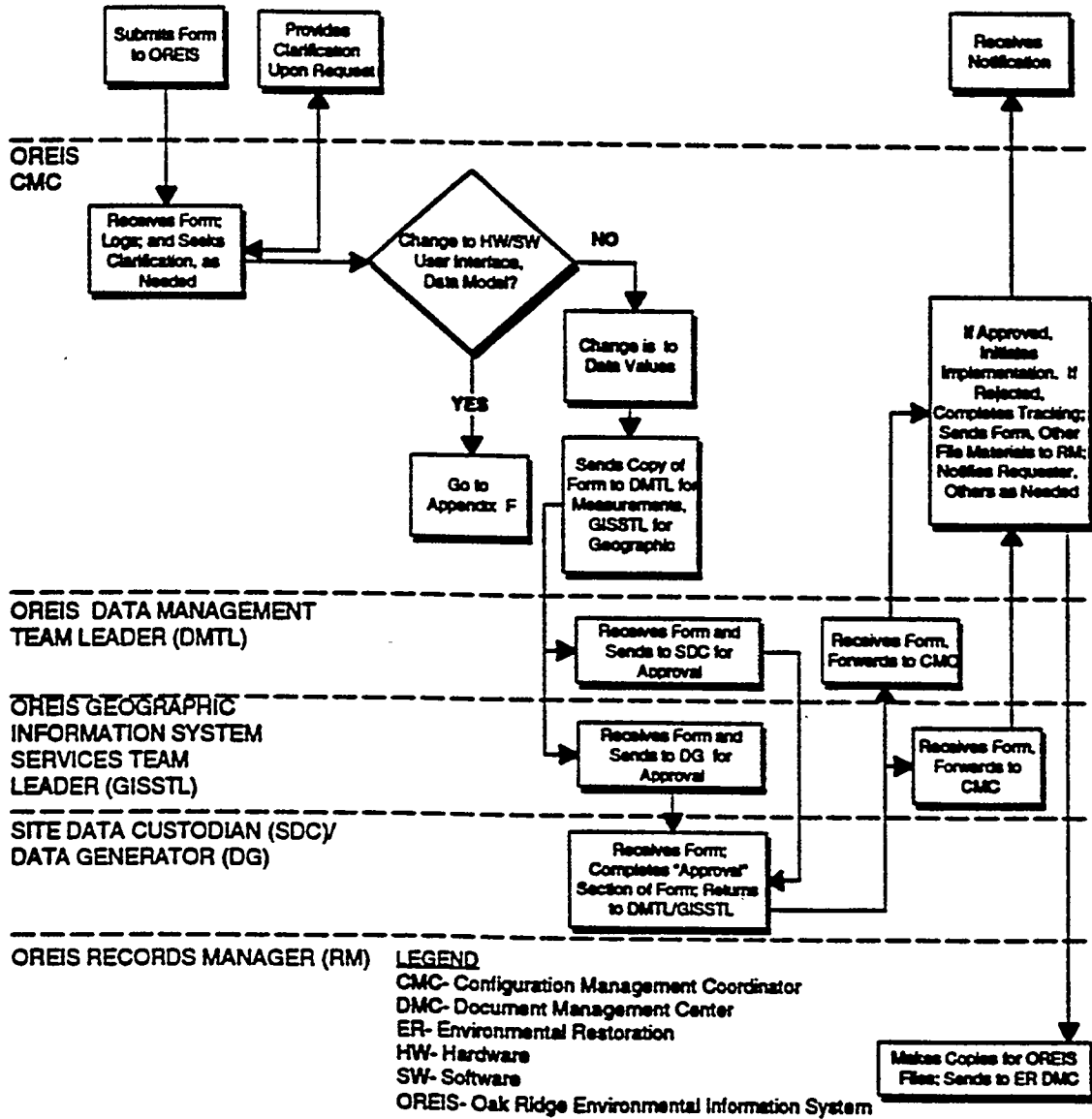
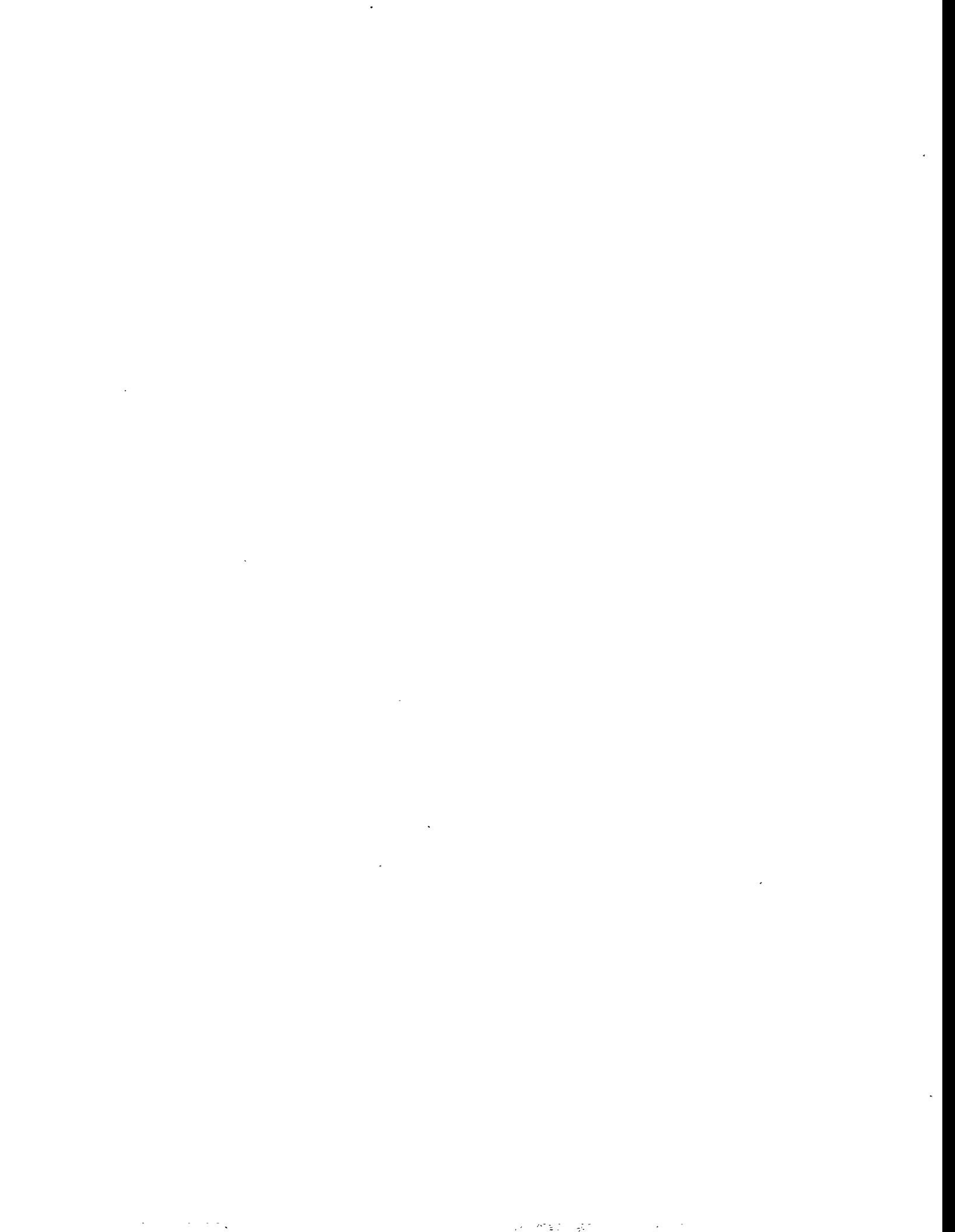


Fig. B.1. Process flow diagram for changes to OREIS data.



APPENDIX C:
RELATED BIBLIOGRAPHY



ER procedures

- ERWM/ER-P2701, Rev. 2 Transmitting ER Data in Ready-To-Load (RTL) Form to the Oak Ridge Environmental Information System (OREIS)
- ERWM/ER-P2703, Rev. 1 Submitting, Reviewing, and Approving Changes to the Oak Ridge Environment Information System (OREIS)
- ERWM/ER-S2704, Rev. 0 Data Product documentation for the Oak Ridge Environmental Information System (OREIS)
- ER/C-S2710, Rev. 0 Format Requirements for Environmental Restoration Map Products

OREIS procedures

- OREIS/P-1101, Rev. 0 Receiving, Processing, and Maintaining Records for OREIS
- OREIS/P-2800, Rev. 0 Back-up, Archival, and Recovery Processes of OREIS
(PCN-1)
- OREIS/P-2901, Rev. 0 Processing and Documenting Data Product Requests
- OREIS/P-2801, Rev. 1 Processing and Review Associated with Measurements Data Transmitted to the Oak Ridge Environmental Information System (OREIS)
- OREIS/P-2802, Rev. 0 Processing and Review Associated with Geospatial Data Transmitted to the Oak Ridge Environmental Information System (OREIS)
- OREIS/P-2805, Rev. 0 Implementing Approved Changes and Conducting Testing Associated with the OREIS User Interface Software, Data Management Software, and Data Model
- OREIS/P-2806, Rev. 0 Implementing Approved Changes to Measurements Data Values and Reference Tables in OREIS
- OREIS/P-2807, Rev. 0 Implementing Approved Changes to Geographic Data Values and Geographic Data Structures in OREIS
- OREIS/P-2808, Rev. 0 Implementing Approved Changes to OREIS Documentation

ES/ER/TM documents

ES/ER/TM-34/R0	<i>OREIS Phase I - System Definition Document</i>
ES/ER/TM-66/R0	<i>OREIS Functional System Design Document</i>
(no number)	<i>OREIS Physical Design Information Package (internal updates to TM-66)</i>
ES/ER/TM-74/R1	<i>Quality Assurance Plan for OREIS</i>
ES/ER/TM-93/R0	<i>Oak Ridge Environmental Information System (OREIS) User's Guide (first official issue)</i>
ES/ER/TM-107	<i>OREIS Project Management Plan for FY 1994 (fiscal year updates to appendixes)</i>
ES/ER/TM-108/R1	<i>Annual Report of Operations for OREIS (1994)</i>
ES/ER/TM-116/R1	<i>Data Dictionary for the Oak Ridge Environmental Information System (OREIS)</i>
ES/ER/TM-137	<i>Results of the User Survey of Functional Requirements for OREIS</i>
ES/ER/TM-138/R0	<i>Results of the User Evaluation of OREIS Functionality</i>
ES/ER/TM-144/R0	<i>Technical Specification for Transferring Non-National Pollutant Discharge Elimination System (NPDES) Surface Water Data to OREIS</i>
ES/ER/TM-145/R0	<i>Technical Specification for Transferring Ambient Air Monitoring Data to OREIS</i>
ES/ER/TM-160/R0	<i>Technical Specification for Transferring Groundwater Well Construction Data and Groundwater Measurements Data to OREIS</i>
ES/ER/TM-164/R0	<i>Technical Specification for Transferring Air Emissions Data to OREIS</i>
ES/ER/TM-175	<i>Technical Specification for Transferring Non-National Pollutant Discharge Elimination System Data to OREIS (in draft)</i>

INTERNAL DISTRIBUTION

1. Central Research Library
2. R.C. Durfee
- 3-5. T.L. James
6. M.A. Kuliasha
- 7-8. Laboratory Records Dept.
9. Laboratory Records, ORNL-RC
Document Reference Section
10. ORNL Y-12 Research Library
11. ORNL Patent Section
12. D.H. Pike
13. C.H. Shappert
14. D.S. Tharpe
15. E.P. Tinnel
16. B.A. Watts
17. T.G. Yow
- 18-20. B.C. Zygmunt

EXTERNAL DISTRIBUTION

21. J.M. Coe, Science Applications International Corporation, 800 Oak Ridge Turnpike, Oak Ridge, TN 37831
22. Office of Scientific and Technical Information, Department of Energy, P.O. Box 62, Oak Ridge, TN 37831
23. Office of the ORNL Site Manager, Department of Energy, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831