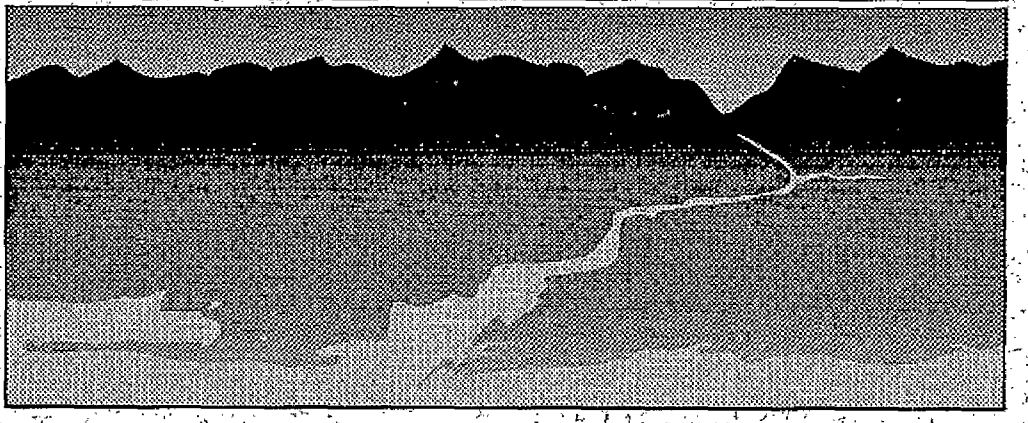
FINAL REPORT
TC 9161-03

LOWER COLUMBIA RIVER



BI-STATE PROGRAM

DATA MANAGEMENT

DATA MANAGEMENT SYSTEMS EVALUATION AND RECOMMENDATIONS

MAY 28, 1993

Prepared By:

TETRA TECH

TETRA TECH

FINAL REPORT
TC 9161-03

DATA MANAGEMENT

DATA MANAGEMENT SYSTEMS EVALUATION AND RECCOMMENDATIONS

MAY 28, 1993

Prepared For:

The Lower Columbia River
Bi-State Water Quality Program

Prepared By:

TETRA TECH
15400 NE 90TH, SUITE 100
REDMOND, WASHINGTON 98502

ACCESSIBILITY INFORMATION: THIS PUBLICATION IS AVAILABLE IN ALTERNATE FORMAT (E.G., LARGE PRINT, BRAILLE) UPON REQUEST. PLEASE CONTACT EITHER ED SALE IN OREGON DEPARTMENT OF ENVIRONMENTAL QUALITY'S PUBLIC AFFAIRS OFFICE AT (503) 229-5766, OR KURT HART IN WASHINGTON DEPARTMENT OF ECOLOGY'S PUBLIC AFFAIRS OFFICE AT (206) 459-6712 TO REQUEST AN ALTERNATE FORMAT

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

This document was prepared by Tetra Tech, Inc. of Redmond, Washington and Ms. Roberta P. Feins, an independent environmental information consultant, for the Lower Columbia River Bi-State Water Quality Program under Contract Nos. C0091228 (WDOE) and 78-91 (ODEQ). Mr. Don Yon and Ms. Cordelia Shea of Oregon Department of Environmental Quality (ODEQ) and Mr. Neil Aaland of Washington Department of Ecology (WDOE) served as the Program Coordinators for the Bi-State Program. Dr. Steven Ellis served as Tetra Tech's Program Manager for the Lower Columbia River Project. Mr. Gary Braun of Tetra Tech served as the Project Manager for this Work Assignment. Funding for this project was provided to the Bi-State Program by WDOE, ODEQ, public ports in Washington and Oregon, and the Northwest Pulp and Paper Association.

Ms. Roberta P. Feins was the primary author of the report. Contributions to the report and technical review were provided by Mr. Braun. Illustrations were prepared by Ms. Kim Shaty. Word Processing and report production were performed by Ms. Lisa Fosse and Ms. Rosemary O'Brien. Reviewers of a draft version of this report included: Don Yon (ODEQ), Ms. Kate Dempsey (WDOE), Ms. Linda Hastings (WDOE), Mr. Neil Aaland, Mr. Dennis Kirk (ODEQ), Mr. Doug Terra (ODEQ), and Ms. Cordelia Shea (ODEQ).

EXECUTIVE SUMMARY

The Bi-State Lower Columbia River Water Quality Program (Bi-State Program) is a four-year program created by the state legislatures of Washington and Oregon. The program's goals are to characterize environmental conditions in the lower Columbia River, identify existing problems and impaired uses, and develop recommended solutions.

Because data and information are needed to meet these goals, the Bi-State Program has recognized the need to develop an integrated framework for managing information. Such a framework, or data management system, includes data, hardware and software, computer programs, and quality assurance procedures that are tied together by a set of organizational arrangements.

This report develops recommendations for a data management system to meet the Bi-State Program's short-, medium- and long-term needs:

- Short-term needs to manage the data that have been collected to date by the Bi-State Program itself
- Medium-term needs to manage, analyze, and distribute data collected by the Bi-State Program, and other related data about the lower Columbia River
- Long-term needs to ensure cooperative sharing of all available information on the River, in order to improve environmental decision-making.

SHORT-TERM NEEDS (2 MONTHS TO 1 YEAR)

The recommended approach to meeting short-term needs is to place monitoring data collected by the Bi-State Program into an existing data archive or data transfer format. A number of data archives and data

transfer formats were evaluated; the Puget Sound Ambient Monitoring Program (PSAMP) data transfer format is recommended as the best for program needs. Oregon requirements for Geographic Information System data should also be used. This approach would cost the Bi-State Program $2,000 - $5,000 per year.

To implement this approach, Bi-State Program cooperative agreements and contracts should contain requirements for submittal of project data in PSAMP format. Use of the PSAMP quality assurance approach should also be required.

MEDIUM-TERM NEEDS (1 YEAR TO 5 YEARS)

The recommended choice of a medium-term system depends on the Bi-State Programs's answers to the following questions:

- How frequently do Bi-State Program staff need to access and use Program data?
- Does the Bi-State Program want to, and is the Bi-State Program able to manage its own database system?

Three possible approaches were identified:

1) Maintain and store the data in the data transfer format: If the Program staff do not need frequent data access, data could be maintained in the data transfer format selected to meet short-term needs. This option is the simplest and lowest-cost option (<$5,000/year) for the Bi-State Program, but also limits the ability to use and analyze data.

2) Select a database system to be managed by the Bi-State Program: If the Bi-State Program needs frequent, direct, and flexible access to information, then an existing database system should be selected and managed by the Bi-State Program. This approach is the most expensive one, requiring from $15,000 to $35,000 in set-up costs, and about $15,000 a year to maintain. However, it would provide the Bi-State Program with the most extensive capabilities for data

analysis and presentation, and local control. The PSAMP database, and NOAA's Coastal Ocean Mapping, Planning and Assessment System (COMPAS) are both workable alternatives; while the PSAMP database would be less expensive and is currently available, COMPAS (currently under development) will provide impressive capabilities.

3) Select a database system managed by another Organization: If the Bi-State Program needs frequent access to information, and does not want to manage its own database management system, then an existing database system managed by another organization should be selected. A mainframe-based Federal database system would be the best approach if the Bi-State Program wanted to minimize local responsibility for and control over the data, and minimize costs ($7,500 in set-up costs, $7,000 - $22,000 per year). ODES is the recommended option for this approach, although it would cost more to use than STORET.

If the Bi-State Program wanted to work with a local database, and does not need on-line analysis capabilities, then working out an arrangement with the Northwest Power Planning Council and the BPA for use of CRCIS (the Columbia River Coordinated Information System), would be an attractive and possibly inexpensive alternative.

LONG-TERM NEEDS (GREATER THAN 5 YEARS)

The Bi-State Program may not continue to exist into the long-term time frame. However, the needs for evaluation and management of the lower Columbia River will remain. The Programmatic Needs Assessment revealed that there is a need for better information, better access to and sharing of existing information, and better use of information in order to manage the river's resources.

A number of technological changes are underway that will facilitate meeting these goals. Using computer networks, databases can be linked together in ways that facilitate data integration, and link together programs within and between agencies, states, Federal, local, and tribal governments, businesses, and individuals.

The Bi-State Program should encourage the discussion of data sharing and data integration as part of the Steering Committee's development of a long-term institutional framework for Columbia River monitoring and management. A working group of data experts from the states and interested Federal agencies should meet to explore issues, and make recommendations to the Steering Committee for a long-term approach. The Program should seek to take advantage of and encourage efforts to improve access to existing agency data needed to make management decisions.

1.0 INTRODUCTION

The Bi-State Lower Columbia River Water Quality Program is a four-year program created by the state legislatures of Washington and Oregon. The program's goals are to:

- Characterize water quality, sediment quality, and biota in the lower Columbia River
- Identify water quality problems in the lower Columbia River.
- Determine whether beneficial/characteristic uses are impaired.
- Develop solutions to problems found in the river below Bonneville Dam.

During its two-year existence, the Bi-State Program has compiled and analyzed historical monitoring data and conducted a reconnaissance survey to describe conditions on the lower Columbia River. In the future, additional reconnaissance surveys and baseline monitoring are proposed.

Because of a need for continued monitoring and an ongoing need for information to support management decisions of the Bi-State Program and of various agencies involved in management of the lower river, the Bi-State Program has recognized the need to develop a framework for managing information. Such a framework, or data management system, includes:

- Clearly defined types of information
- Computer equipment, and software
- Applications (e.g., specialized computer programs to perform needed functions)
- Procedures (e.g., data entry, quality assurance).

All of these elements are tied together by a set of organizational arrangements that define who collects and distributes information, who uses it, who ensures its quality, etc.

In the fall of 1992, the Bi-State Program issued a work assignment to help define its data management system. The Work Assignment included the following tasks:

- Perform a Data Management Needs Assessment to define required and preferred features of the Data Management System
- Evaluate existing Databases or Data Management Systems and recommend one for Bi-State Program use
- Demonstrate the recommended system to the Bi-State Program Steering Committee and staff.

A more detailed discussion of the overall project process and schedule is available in the Data Management Work Plan (Tetra Tech 1992).

The Data Management Needs Assessment was completed in January of 1993 (Tetra Tech 1993). The Programmatic Needs Assessment (Webster 1992) and a series of interviews were used as research material. Interviews were conducted with representatives from within and outside of the Bi-State Program to provide different views on optimal data management strategies.

The Needs Assessment identified key programmatic and technical issues that would affect the selection of a Data Management System, including institutional considerations (how long and in what form would the Bi-State Program exist), existing data management activities, and existing data standards of Washington and Oregon state governments. A wide list of user needs and requirements was defined. These requirements were further broken down into those needed to meet short-term Bi-State Program objectives (within the next 4 years), and long-term Bi-State Program objectives (beyond 4 years). Short-term requirements were those needed to support characterization and analysis by the Bi-State Program itself, while long-term requirements were those needed to support more coordinated interagency management of the lower Columbia River. Elements of a potential data management system were

categorized as required (fundamental to the utility of the proposed system) or preferred (not fundamental to the utility of the proposed system, but would enhance it).

Once the Needs Assessment was complete, the next step was to evaluate existing databases or data management systems, determine how they met the defined requirements, and recommend a system for Bi-State Program use. However, the list of requirements in the Needs Assessment was so extensive, and the list of potential users so broad, it was clear no system could meet all these needs. An additional meeting was held with Bi-State Program staff to narrow down the objectives of the data management system, modify the list of requirements, and develop a final list of systems to be evaluated.

At that meeting, the decision was made to focus more narrowly on the needs of the Bi-State Program itself. The evaluation criteria were modified to reflect the priority that the selected data management system directly support Bi-State Program needs. It was agreed that existing data management systems should be evaluated according additional "technical" criteria, that would assess features such as system design and performance. Recognizing that the Bi-State Program itself had some very immediate needs for data management, time scales were modified to reflect options for meeting short-, medium- and long-term needs:

- The short-term (2 months to 1 year) data management objectives of the program are to manage the data that have been collected or compiled through the program itself. This includes reconnaissance survey data (water/sediment/tissue), loading calculations, and baseline monitoring data.

- In the medium-term (1 year to 5 years), the data management objectives include managing, analyzing, and distributing data collected by the program, and other related data about the lower Columbia River, and encouraging the distribution of information on the river to interested parties, including the public and other agencies.

- In the long-term (greater than 5 years), the data management objectives of the program are to ensure cooperative sharing of all available information on the lower Columbia River, in order to improve environmental decision-making.

The modified time scale and modified list of elements were then used to evaluate existing systems for use by the Bi-State Program. The purpose of this report is to present the results of the system evaluations, and to develop recommendations for meeting short-, medium, and long-term data management needs. The report is organized into four major sections:

- Section 1: Introduction (this Section)
- Section 2: Short-term Needs/Approaches
- Section 3: Medium-term Needs/Approaches
- Section 4: Long-term Needs/Approaches.

Each of the three latter sections is organized in a similar manner. First, the short-, medium-, or long-term need is restated and identified. This is followed by a description of the alternative approaches or options. Following that section is the description of the actual evaluation process and results of the systems evaluation based on the required, preferred, and technical elements. Finally, each section concludes with the recommendations based on the system evaluations.

2.0 SHORT-TERM NEEDS/APPROACHES

The short-term (2 months to 1 year) data management objectives of the Bi-State Program are to manage the data that have been, or will be, collected or compiled through the program itself. This includes data from the reconnaissance survey on chemicals in water, sediment, and tissue, benthic abundance data, loading calculations, and baseline monitoring data. The Bi-State Program itself has no immediate needs to use the data, but does need to make the data available to interested parties.

2.1 RECOMMENDED APPROACH

There are two general approaches the Bi-State Program could take to address the short-term data management needs. These include:

- Selection of a data archive/transfer format
- Selection of a data management system without use of an interim archive/format.

While both approaches would meet the short-term needs, the latter approach requires that the Bi-State Program immediately address the many outstanding issues identified for the medium-term options (see Section 3.0). As this action by the Bi-State Program does not appear possible without additional evaluation and definition of Bi-State Program data management goals, the recommended approach to making existing Bi-State Program data available is to place data into a standard form that can be acquired by interested users.

Advantages of this approach include:

- The Bi-State Program would not need to expend time and effort to maintain Program data on a computer in accessible or usable form

- The Bi-State Program would not need to reformat the data each time a request for data is received, so that a low level of staff effort would be required
- Requestors would receive data in a standard, documented form
- Requestors can reformat and analyze the data as desired
- Low cost ($0 - $10,000/year).

Disadvantages of this approach include:

- The requestor must be familiar with the data format and know how to retrieve and use the data
- The data are not immediately available for use, but require retrieval and conversion. Data cannot be listed, graphed, mapped, or analyzed without transfer to another format
- The data are not immediately available to the Bi-State Program staff for their use.

In light of short-term program uncertainties and limited available budgets, this approach seems to be the best for meeting user needs for the next 6 months to one year.

Figure 1 shows the choices that must be made to implement this option. First, the Bi-State Program must decide whether to place data into a data archive or into a data transfer format. Then the Bi-State Program needs to select a particular archive or data transfer format for use.

2.2 DATA ARCHIVE OR DATA TRANSFER FORMAT?

A data archive is an established database system, such as the EPA STORET database, where data are stored until needed. Datasets are loaded into an archive and join a pool of existing data.

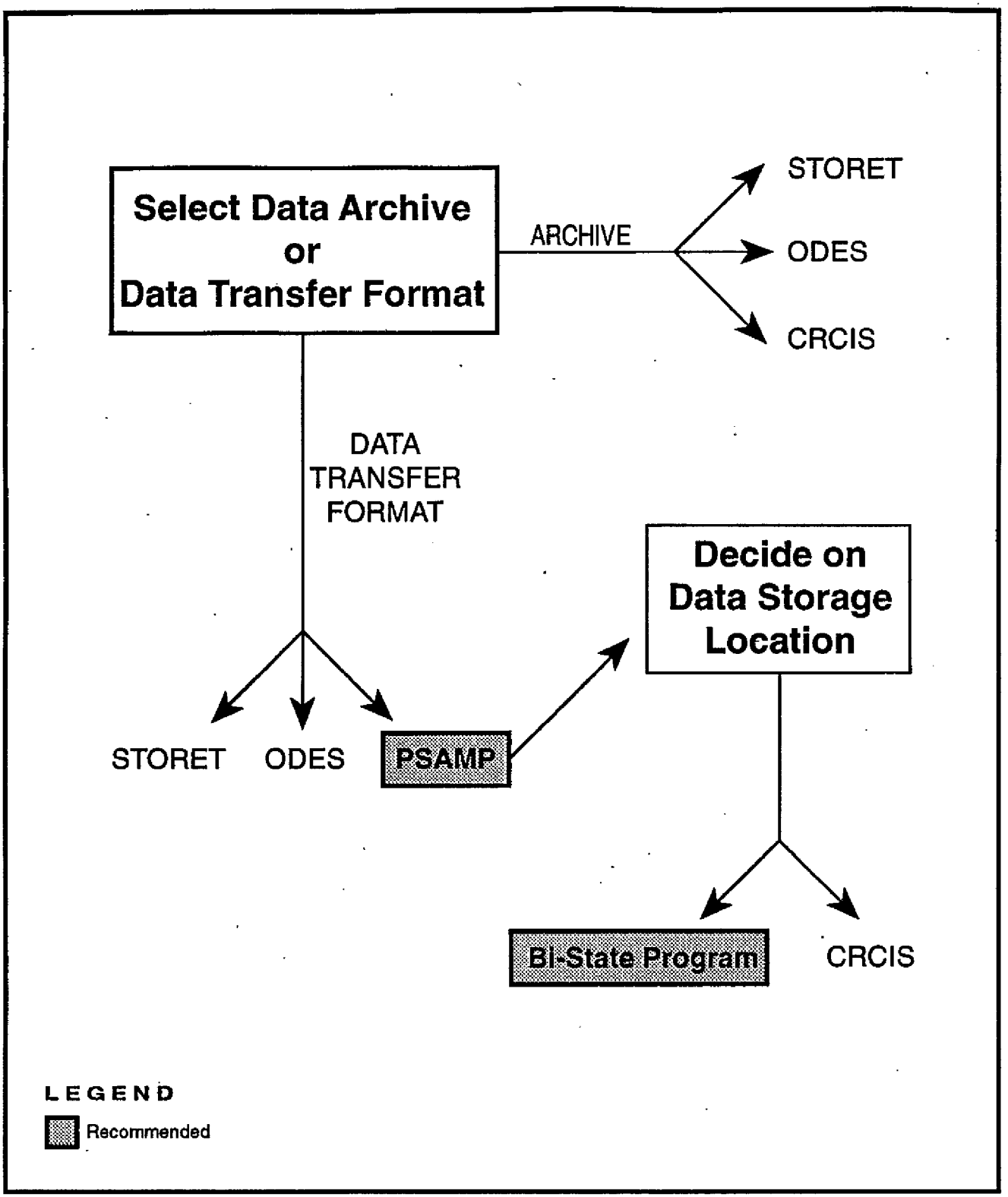


Figure 1. Short-term Data Management Options and Recommended Approach.

A data transfer format is a specification that describes the structure and content of a data file used for data transfer. Such formats usually specify the order of information on each line of a data file, as well as the codes, formats, and maximum width of each piece of information reported. Datasets exist in data files, and remain separate - they do not join a larger database.

Getting data into a data transfer format is a necessary step in getting data into a data archive. The data from a monitoring survey are entered into a computer using the correct codes, spacers, and required information. Existing computerized data can be reformatted manually or by using special computer programs. This step can cost from $0-$2,000 per monitoring study, depending on the form the data are in. Costs can be minimized by clearly specifying the required format in advance of issuing a contract for data collection.

At this point, if the data are to remain in the standard format, they remain in discrete files and are stored on disks or other media. Copies of these files can then be sent on disk or by modem to those requesting data.

If the data are to go into a data archive, they are submitted to the archive in one or more electronic files. Archive staff use computer programs to transfer the information from the files into a large database. The database can then be accessed by anyone with authorization to view or download data. Costs to move data from discrete formatted files into the archive can take from $500 to $10,000 per monitoring study.

Using a data archive would have the following advantages:

- Data archives usually have specific data formats and standard procedures for receiving, adding, and maintaining the database. Standard procedures for data access (logging on, downloading information) have also been developed

- Persons requesting data could be directed to the data archive, where they would have primary responsibility for acquiring the data in a form useful to themselves

- Once the data are in the data archive, there is no ongoing cost to the Bi-State Program

- Putting data into a data archive contributes to the pool of monitoring information available for a variety of uses.

Using a data archive would have the following disadvantages:

- Some archives cannot store all of the information associated with a monitoring study: a data type may not be supported or certain quality assurance or identifying information cannot be stored. Depending on the archive selected, this information could be permanently disconnected from the rest of the dataset

- There is an initial effort needed to submit the data to the archive and to answer questions that may arise in adding data to the archive. This means that the initial cost of putting data from a monitoring study into an archive could be from $500 to $10,000 higher than using a data transfer format. However, once the data are in the archive, there is no further cost to the Bi-State Program

- A person requesting data must get authorization to use the archive, learn to log-on to the archive, find the desired information, and learn how to down-load it. Getting access to the archive, and to the desired data set can sometimes be a problem

- If the Bi-State Program needed the data for analysis, staff would have to go through the same process as any other requestor - logging in, downloading information.

Using a data transfer format would have the following advantages:

- Data would be maintained on disk or tape in both Bi-State Program offices, so that the information would be immediately accessible to Oregon and Washington Bi-State Program staff

- The Bi-State Program could directly distribute the data to requestors, and answer questions about the data as they arose

- It would be simple for a person requesting data to get data files via diskette or modem

- There would be lower initial costs to use this approach, since data do not need to be submitted to an archive.

Using a data transfer format would have the following disadvantages:

- Bi-State Program staff would have responsibility for making copies of the data and distributing files via diskette or modem (1/2 to 1 hour of time per request). The overall time and cost of this approach would not be onerous, but would be an ongoing responsibility

- Potential data users would have no way to know what Columbia River Bi-State Program data exist.

The choice of a medium-term data management approach (Section 3.0) may also affect the selection of a data archive or a data transfer format, since some of the formats evaluated are required for data entry into some of the systems evaluated for medium-term use.

2.3 EVALUATION OF DATA ARCHIVES

In order to meet Bi-state Progams needs, a data archive must:

- Be an existing database system or archive

- Store all of the data types to be archived in an appropriate form

- Store essential data elements, and quality assurance information (i.e., sample dates, station locations, data qualifiers, analysis methods)

- Provide for easy transfer of data into this format, and relatively simple transfer from this format to files for use in reporting, analysis, mapping, etc.

- Be familiar to the user community

- Be reasonably easy to access or read for users using DOS compatible personal computers (i.e., a connection is available by computer modem)

- Maintain data identity throughout the transfer process, so that data can be tracked back to the original source if necessary

- Be a stable format that is likely to be around in several years, and is supported by an organization or agency.

2.3.1 Evaluation Methods

The following data archives were evaluated:

- STORET
- Ocean Data Evaluation System (ODES)
- Columbia River Coordinated Information System (CRCIS)

Each alternative was rated based on how it met the required, preferred, and technical elements according the following scale:

3 - Meets user requirements (as is, or with minor changes)
2 - Nearly meets user requirements
1 - Barely meets user requirements
0 - Does not meet user requirements

Table 1 shows the results of the evaluation. Each option is discussed briefly below.

TABLE 1. - SHORT-TERM NEEDS:
DATA ARCHIVE EVALUATION

Evaluation Elements	STORET	ODES	CRCIS
Stores Data on:	1.5[a]	2.5	0.5
Fish health/mortality	N	Y	N/Y
Contaminants in biota	N	Y	N/Y
Wildlife and fish communities	N	N/Y	N/Y
Invertebrate communities	Y	Y	N/Y
Sediment contaminants	Y	Y	N/Y
Radionuclides in sediment	Y	Y	N/Y
Water column contaminants	Y	Y	N/Y
Water quality	Y	Y	N/Y
Point sources	N	N/Y	N/Y
Pollutant loadings	N	N/Y	N/Y
Data Quality Assurance Information	0	3	0.5
Easy Transfer	1	1	3
Familiarity	3	2	1
Easy Access	3	2	1.5
Maintain Data Identity	2	2	3
Long-Lived, Supported	3	2	2
Minimal Cost	2	1	2
TOTAL UNWEIGHTED SCORE	15.5/24	15.5/24	13.5/24

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of evaluation elements may be found in Appendix A.

2.3.2 STORET

STORET is EPA's major water quality database, and has been in existence for at least 15 years. Datasets are transferred to STORET by an agency "owner" in a standard submittal format. STORET is the most widely used water quality database in the U.S. It is located on an EPA mainframe computer in North Carolina.

Because STORET is so long-lived, it uses out-dated computer technology and its capabilities for data formatting and down-loading are very limited. In addition, STORET would not be able to handle all of the data types and all of the information the Bi-State Program would want to archive. STORET was not originally designed to handle sediment, bioassay, or tissue data and does so in a clumsy way, that does not capture minimum essential information. The quality assurance information that can be stored for any data type is very limited.

EPA is currently proceeding with a major project to update and redesign STORET. The intent is not to rebuild STORET as a complete data management and analysis tool, but to build a more user-friendly data archive, with extensive capabilities for transferring information to user computers for further manipulation and analysis (King, R., 31 March 1993, personal communication). The project team anticipates testing the design and the use of Oracle in a prototype system by the end of 1993 at the earliest. No firm date for completion of the update has been set.

To put data into STORET, the Bi-State Program would need to reformat existing data or reenter data. (There are several microcomputer packages that can output data in STORET format). With an agency "owner" ID and a user ID, the data can then be uploaded to STORET via modem. STORET staff run some limited checks on the data, and, once corrected, data are made available for use by anyone who knows the agency "owner" ID. Costs to add a dataset to Storet would include: costs to put data into STORET format, costs to upload data to Storet (minimal), and costs to make any changes after STORET data checking.

As an example, water column data, sediment pollutant data, grain size data, and fish tissue data from the Columbia River Reconaissance Survey could be stored in STORET (Benthic Abundance Data could not). It would cost from $1,500-$2,000 to reformat the existing data (in spreadsheets), and $0 - $500 for editing after data checking.

2.3.3 Ocean Data Evaluation System (ODES)

ODES is a database developed in the mid-1980's by Tetra Tech, Inc. and AMS Inc. for the EPA Office of Marine and Estuarine Programs, Washington DC. Originally designed to store monitoring data from the 301(h) ocean discharge program, ODES is currently used to store a variety of marine and estuarine data. It is located on an EPA mainframe computer in North Carolina.

ODES was designed as a data archive and a data analysis system. It can store all of the data types required by the Program. ODES can store quality assurance information including data from lab blanks, matrix spikes, duplicate analyses etc. Before a data set can be archived in the ODES database, it goes through a series of automated data checks, and technical review to assure data quality.

ODES tools allow data selected by area, and date to be listed in various formats, and down-loaded. User access to ODES is more limited than STORET - one must request a user ID from EPA staff.

To put data into ODES, the Bi-State Program would need to reformat existing data or reenter data. (There are several microcomputer packages that can output data in ODES format). Monitoring data for each separate data type and each three month period are identified as a dataset. Each dataset is then submitted to EPA. ODES staff run extensive computerized checks on the data before loading, and also conduct a technical review of each dataset. Once correct, data are put on-line and made available to anyone with an ODES user ID. Costs to add a dataset to ODES would include: costs to put data into ODES format, costs to upload data to ODES, costs of the data checking and technical review (estimated cost: $1,000 per dataset), and costs to make any changes after ODES data checking.

As an example, the Columbia River Reconaissance Survey would comprise six ODES data sets (Sediment Pollutant Data, Benthic Abundance Data, Grain Size Data, Fish Bioaccumulation Data, Water Quality Data, Bacteria Data). It would cost from $1,500-$2,000 to reformat the existing data (in spreadsheets), $6,000 for data checking and technical review, and additional time to make changes or corrections to the datasets after data checking.

2.3.4 Columbia River Coordinated Information System (CRCIS)

CRCIS is a system of databases maintained by the Northwest Power Planning Council, the Bonneville Power Administration (BPA), and state environmental, and resource agencies. CRCIS is a sub-set of the

Northwest Environmental Database (NED), which was originally developed to provide information on potential hydroelectric sites on rivers in the Northwest. CRCIS is being developed because of issues surrounding the Columbia River and endangered runs of salmon. Staff at the Northwest Power Planning Council have indicated a willingness to archive and distribute data on the lower Columbia River, since they believe it may contribute to understanding conditions for fish in the river.

CRCIS is not a data archive in the same sense that STORET and ODES are. CRCIS does not have formats or procedure for submitting data, does not add data to a larger data pool, would not provide on-line access to data, and has no capabilities for automatically down-loading information. CRCIS databases do not currently store detailed monitoring data. CRCIS could meet the needs of the Bi-State Program by accepting formatted files from the Bi-State Program, archiving them and distributing them when requested. The Bi-State Program would still be responsible for defining the content and format of the data files, and ensuring data quality before the data go to this distribution point.

Costs to the Bi-State Program to use CRCIS are unclear. To archive data with the CRCIS, the Bi-State Program would need to select a data format and reformat existing data or reenter data ($1,500-$2,000). Data would then be sent to CRCIS, and made available to anyone requesting it. BPA staff have indicated a willingness to archive and distribute the data; however, some kind of cooperative agreement between the Bi-State Program and NWPPC/BPA would probably be required.

2.4 EVALUATION OF DATA TRANSFER FORMATS

In order to meet Bi-State Progams needs, a data transfer format must:

- Be an existing, published format
- Provide specifications for storing all of the data types to be archived
- Store essential data elements, and quality assurance information (i.e., sample dates, station locations, data qualifiers, analysis methods)

- Provide for easy transfer of data into this format, and relatively simple transfer from this format to files for use in reporting, analysis, mapping, etc.

- Be familiar to the user community

- Be reasonably easy to read for users using DOS compatible personal computers (i.e., the formatted files fit onto floppy disks, data can be loaded into standard databases, or spreadsheets.)

- Maintain data identity throughout the transfer process, so that data can be tracked back to the original source if necessary

- Be a stable format that is likely to be around in several years, and is supported by an organization or agency

- Minimize the cost of getting data into the format or archive.

2.4.1 Evaluation Methods

The following data transfer formats were evaluated:

- STORET data transfer format
- Ocean Data Evaluation System (ODES) data submittal format
- Puget Sound Ambient Monitoring Program (PSAMP) data transfer formats
- Spreadsheets

Each alternative was rated according the following scale based on the required, preferred, and technical elements:

3 - Meets user requirements (as is, or with minor changes)

2 - Nearly meets user requirements

1 - Barely meets user requirements

0 - Does not meet user requirements

Table 2 shows the results of the evaluation. Each option is discussed briefly below.

TABLE 2. - SHORT-TERM NEEDS:
DATA TRANSFER FORMAT EVALUATION

Evaluation Elements	STORET	ODES	PSAMP	Spreadsheets
Organizes and Stores Data on:	1.5[a]	2.5	2.5	1
Fish health/mortality	N	Y	Y	N/Y
Contaminants in biota	N	Y	Y	N/Y
Wildlife and fish communities	N	N/Y	N/Y	N/Y
Invertebrate communities	Y	Y	Y	N/Y
Sediment contaminants	Y	Y	Y	N/Y
Radionuclides in sediment	Y	Y	Y	N/Y
Water column contaminants	Y	Y	Y	N/Y
Water quality	Y	Y	Y	N/Y
Point sources	N	N/Y	N	N/Y
Pollutant loadings	N	N/Y	N	N/Y
Data Quality Assurance Information	1	3	2.5	1
Easy Transfer	1.5	1.5	3	1
Familiarity	3	2	1.5	3
Easy Access	2	2	2	3
Maintain Data Identity	1	2	3	3
Long-Lived, Supported	3	2	1.5	1
Minimal Cost	2	1	2	2
TOTAL UNWEIGHTED SCORE	15/24	16/24	18/24	15/24

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of evaluation elements may be found in Appendix A.

2.4.2 STORET Data Transfer Format

Datasets are transferred to STORET in a standard submittal format. These formats are widely used by Federal and state agencies, and are probably the most common format for the transfer of environmental information.

The STORET formats were not originally designed to handle sediment, bioassay, or tissue data and do so in a clumsy way, that does not capture all required information. Because of their age, the formats are based on a 80-column punch card format that is out-dated and a hierarchical data structure somewhat complex to prepare and to read. Very little data quality assurance information can be stored with the data. It is likely that STORET formats will always be supported, however, they may change once the STORET redesign is complete.

If STORET format is to be used for the Bi-State Program, the Program needs to work with data collectors to agree on standards for data collection, quality assurance, and reporting that are more stringent than those required for entry into STORET. Otherwise data will not be comparable.

Costs to put data in STORET format would include data entry or reformatting costs (there are several microcomputer packages that can output data in STORET format). Estimated costs are $1,500 - $2,000 per monitoring study, if the data need to be reentered or reformatted.

2.4.3 ODES Data Submittal Format

ODES data submittal formats are a variation of formats that were developed for use in the National Oceanographic Data Center archives in the 1970's. The formats were updated in the 1980's for the ODES system. They can handle almost all of the required data types.

ODES data submittal formats are based on an 80-column punch card format that has been modified so data are not restricted to 80 columns. ODES also has a hierarchical data structure that is somewhat complex to prepare and to read. Putting data into ODES format usually requires renaming stations and samples which makes tracking original data slightly more difficult. ODES format is not commonly known or used in the Northwest, particularly since there are no 301(h) program permittees in the area.

Costs to put data in ODES format would include data entry or reformatting costs (there are several microcomputer packages that can output data in ODES format). Estimated costs are $1,500 - $2,000 per monitoring study, if the data need to be reentered or reformatted.

2.4.4 PSAMP Data Transfer Formats

The Puget Sound Ambient Monitoring Program (PSAMP) has published a set of standard data formats for the exchange of Puget Sound monitoring data among agencies. The Puget Sound Dredge Disposal Analysis Program recommends a related, but different set of formats for reporting sediment chemistry, bioassay, and invertebrate community data.

The PSAMP data transfer formats were developed in 1988 for use in transferring data among agencies involved in monitoring Puget Sound. The formats have been and are being used for other programs as well, including the Puget Sound Estuary Program (PSEP). The 1991 Puget Sound Water Quality Management Plan calls for the use of PSAMP data transfer formats in:

- The Puget Sound Ambient Monitoring Program
- Monitoring watersheds under the non-point program
- Exchange of NPDES permit monitoring data
- Department of Ecology intensive surveys.

Despite these requirements, the formats are not widely used in the user community outside of the PSAMP program.

The formats were designed to allow data to be transferred among databases using different hardware and software, and cover all of the types of data identified in the user requirements. The formats are based on a relational data structure that separates information into clearly defined files and allows different kinds of information to be linked. NODC codes are used throughout. General quality assurance information about a data set is stored; however, the formats are not designed to transfer data on matrix spikes or blanks.

Costs to put data in PSAMP format would include data entry or reformatting costs (there are several microcomputer packages that can output data in PSAMP format). Estimated costs are $1,500 - $2,000 per monitoring study, if the data need to be reentered or reformatted.

2.4.5 Spreadsheets

Monitoring data could be maintained in archived spreadsheets for distribution to interested users. While spreadsheets are a commonly used, stable and widely known type of software, in themselves they do not specify a format for the data stored therein. The only requirement for spreadsheet design is that data be displayed in rows and columns. However, in order to readily exchange data, a transfer format must specify the content of information in each cell of the spreadsheet, along with codes, etc. Spreadsheets in themselves have no data quality assurance features.

Spreadsheets are actually a software medium for transfer of already formatted data. Data in STORET, ODES, or PSAMP format could be transferred in a spreadsheet from one user to another. However, it is often very difficult to get data into or out of a spreadsheet, due to problems with missing data, and brand-specific formatting requirements.

The spreadsheets that are currently used to store existing data, are not in a standard form; while the data can be used in the spreadsheet, it would be difficult for the requester to transfer the information to any other software.

2.5 RECOMMENDATIONS

Selection of a Data Transfer Format is the recommended short-term option. This option is recommended over a Data Archive System because distribution of data in a data transfer format will provide Bi-State Program staff, committee members, and others with sufficient access to data in the short-term; is the least expensive cost option; and does not preclude the use of an archive system in the future. There are 2 recommended alternatives for data storage and distribution:

- Bi-State Program staff can maintain properly formated data on floppy disks and, when requested, can provide the data on floppy disks to interested parties

- The Bi-State Program can develop an arrangement with the Northwest Power Planning Council by which CRCIS is the repository for Bi-State Program data.

While nothing precludes providing CRCIS with a copy of the Bi-State Program data, the former approach is recommended. Bi-State Program staff should be the ones responding to requests for information, since they can answer questions about the data, and will be the first to know about problems with or updates to the information. Should the Bi-State Program cease to exist, using CRCIS as the archive for information would be the recommended alternative.

Under either alternative, no matter who distributes the information, the monitoring data will still need to be in an appropriate electronic format. The PSAMP format is recommended for use as a data transfer format. The Bi-State Program may want to create its own version of the PSAMP format, for example, to accommodate the management of point source and pollutant loading information.

In order to implement the use of PSAMP format as a data transfer format, Bi-State Program cooperative agreements and contracts should contain requirements for submittal of project data in PSAMP format. In addition, the PSAMP quality assurance approach, which specifies certain minimum reported data elements, required accuracy of spatial locations, and required technical review of data should also be implemented.

2.6 GEOGRAPHIC INFORMATION

Coordinate-based geographic information (e.g., the location of sampling stations) can be reported in PSAMP format. However information on resource locations or information required to map areas cannot be reported in PSAMP format.

Geographic data compiled to date by the Bi-State Program has been archived at the Oregon State Service Center. The Service Center charges by the hour for retrieval and use of the information. Neither ODEQ nor Ecology have any centralized GIS capabilities, though each agency has several copies of PC Arc/Info.

The Bi-State Program should adopt Oregon standards for GIS information (Oregon State Map Advisory Council 1990). Requirements for submittal of data according to these standards should be placed in Bi-State Program contracts and cooperative agreements.

The Oregon GIS requirements are general and do not specify the format in which geographic data should be exchanged. There are some draft recommendations for reporting of estuarine and marine resource data, developed by the Washington Geographic Information Council (1990). The Bi-State Program should work with Washington and Oregon state agencies to ensure that different types of geographic resource information are submitted in formats compatible with emerging agency GIS designs.

If CRCIS is selected as a data archive, geographic data can be stored in BPA's Geographic Information System. The Bi-State Program could then make arrangements with BPA to distribute the geographic data, or to prepare maps for the Bi-State Program as needed. No information on potential costs of this approach are available.

3.0 MEDIUM TERM NEEDS/APPROACHES

In the medium-term (1 year to 5 years), the Bi-State Program will need to store information collected from any ongoing monitoring programs. The Bi-State Program will also need tools to manage, analyze, and distribute data on the lower Columbia River to its various audiences: the scientific and business communities, government agencies, legislatures, and the public.

Staff of the Bi-State Program have indicated that it is not their intention for the program to become a repository of all data on the lower Columbia River. The Bi-State Program will want primarily to manage monitoring information; some spatial or monitoring data from other sources may be included for analysis purposes.

3.1 RECOMMENDED APPROACH

There are a confusing number of alternative approaches to meeting the Bi-State Program's medium-term needs. There is no one single approach that will meet all the needs; the Bi-State Program must make decisions on priorities, costs and desired outcomes. Two major questions are:

- How frequently do Program staff need to access and use Program data?
- Does the Bi-State Program want to and is the Bi-State Program able to manage its own database system?

Once these questions are answered, an approach, and a data management system can be selected (Figure 2).

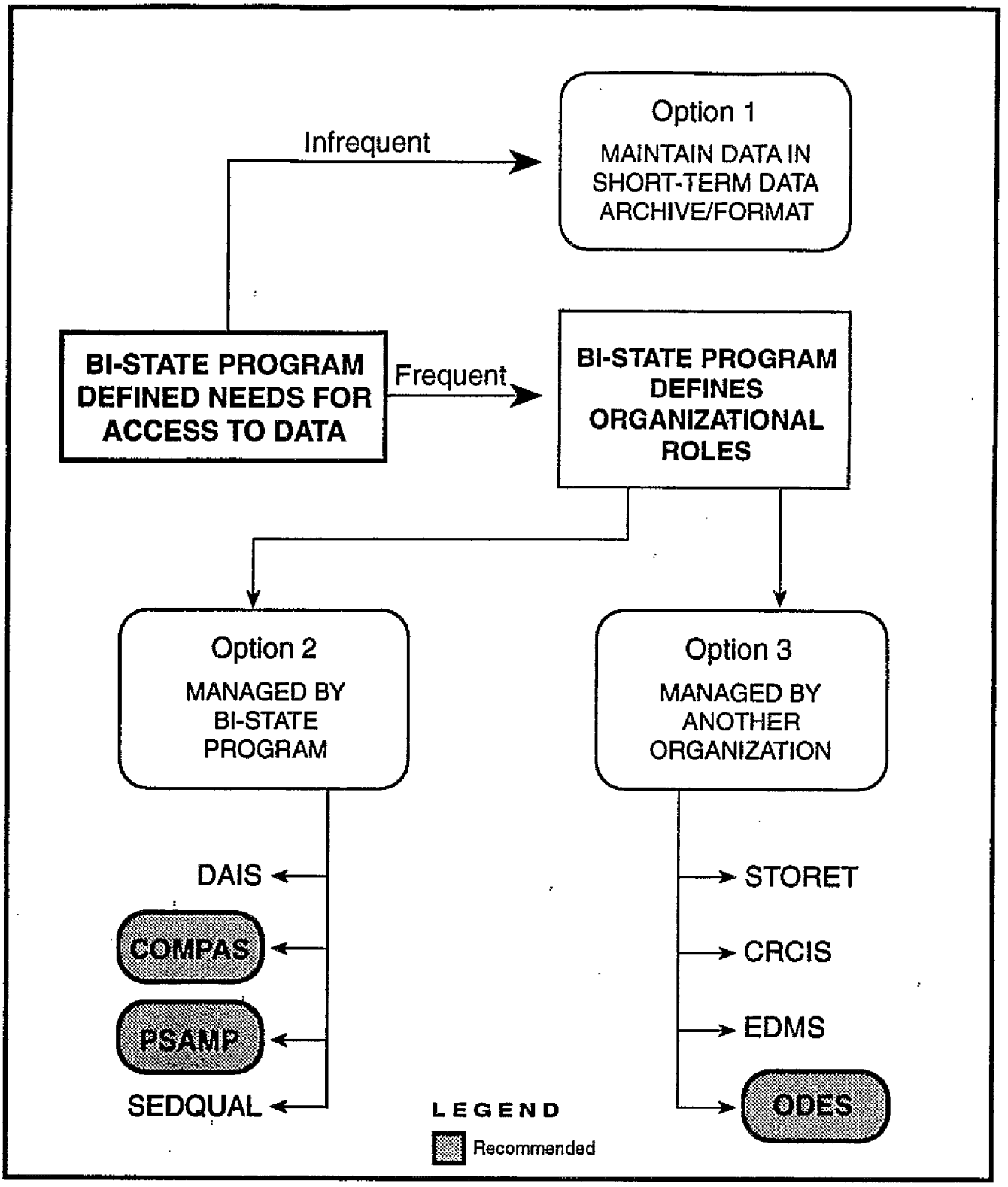


Figure 2. Medium-term Data Management Options and Recommended Approach.

3.1.1 Define Needs for Access to Data

Monitoring data, resource data, and other Columbia River information have been and will continue to be collected by the Bi-State Program. Bi-State Program staff may be content to have the data compiled and presented in published reports, or they may wish to have access to the information in a computer. Having the information in a computer would allow staff to add, compare, and combine data in new and different ways.

If Bi-State Program staff anticipate a frequent (i.e., more than once a month), need to access and work with Bi-State Program data, then the Bi-State Program should decide on an organizational approach (Section 3.1.2), and select a database management system. Conversely, if Program staff anticipate never using data on a computer, or using data once a year, the Bi-State Program should simply maintain copies of Program data in the same standard format used for the short-term approach (Figures 1 and 3).

In maintaining the archive approach, each new year of data would be submitted in the transfer format to the Bi-State Program staff, and maintained on its own set of disks (or other magnetic media). These disks could be provided to staff or contractors doing analyses on program data.

The primary advantage of this approach is its low cost and low "bother" to the Bi-State Program staff. Costs incurred would be those required to get data entered into the standard format (which may not cost anything extra if requirements are written into contracts), and to make copies of the data for those requesting them (less than $100 a year in supplies, and maybe 1 hour a week of labor).

The primary disadvantage of this approach is that Bi-State Program staff would not have access to the data, and would not be able to readily view the data, ask questions about conditions and trends, build reports, and do analyses.

3.1.2 Define Organizational Roles

If the Bi-State Program wishes to use an existing data management system, a decision needs to be made about whether the system will be managed by the Bi-State Program staff, or managed by another organization.

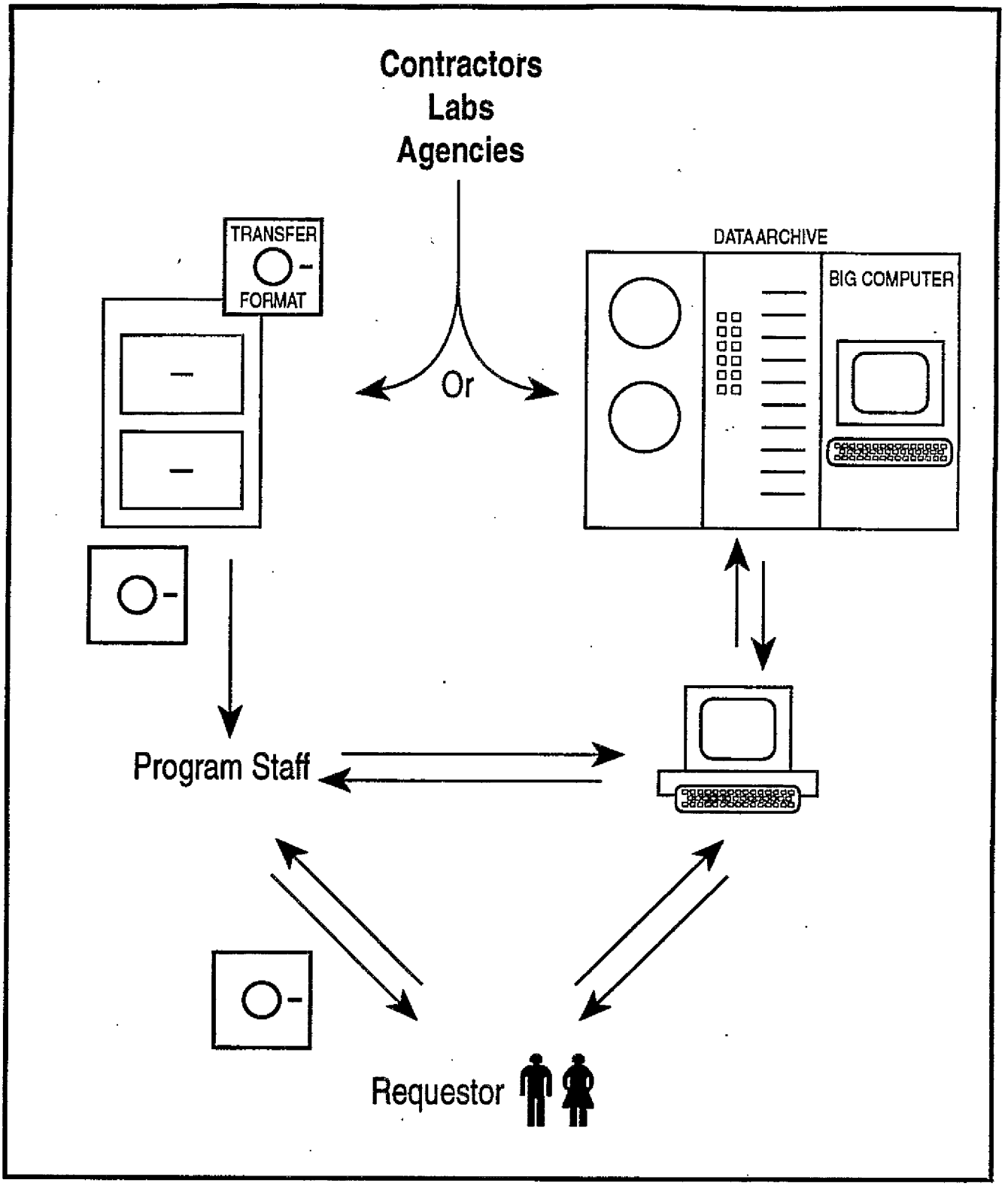


Figure 3. Medium-Term Approach #1. Maintain Data in the Archive/Transfer Format.

3.1.2.1 Use a Program-Managed Database System (Figure 4). This approach would involve placing monitoring data collected by the Bi-State Program into an existing database system with menus and functions designed for environmental information. The system would be one that is managed and maintained by Bi-State Program staff.

The approach would require Bi-State Program staff (or other designated staff) to take an active role in operating the database system. Yearly monitoring data could be submitted by agencies or labs in a standard readable format for inclusion in the database. Staff would be responsible for adding data, checking data quality, conducting analyses, and providing data to requestors.

Costs of this medium-term approach would include initial costs to acquire hardware and software (@$5,000), set up the database management system (@$5,000-$20,000), and obtain training ($5,000-10,000). Ongoing costs would include staff time, which would be divided between managing the database, and working with the database. Managing the database could cost @$6,000 -$10,000/year (10-15% of a staff-person's time). Working with the database (conducting analyses, preparing reports, etc.) would be part of the Bi-State Program's regular activities.

If this approach were selected, a way of dividing or sharing information between Washington and Oregon would need to be worked out. If each state agency (Washington Ecology, Oregon DEQ) maintained a copy of the database, then the costs listed above would be doubled. One agency could maintain the master copy of the database, and the other agency could use a modem or Wide Area Network to log-in and use the database. Since no Wide Area Network link is in place right now, it would cost from $10,000-$20,000 to set up adequate communications between the agencies.

Advantages of using a Program-managed database include:

- Data would be readily accessible by program staff in a form that allows direct and flexible retrieval and analysis. This would contribute considerably to the effectiveness of the Bi-State Program, and its ability to understand the lower Columbia River

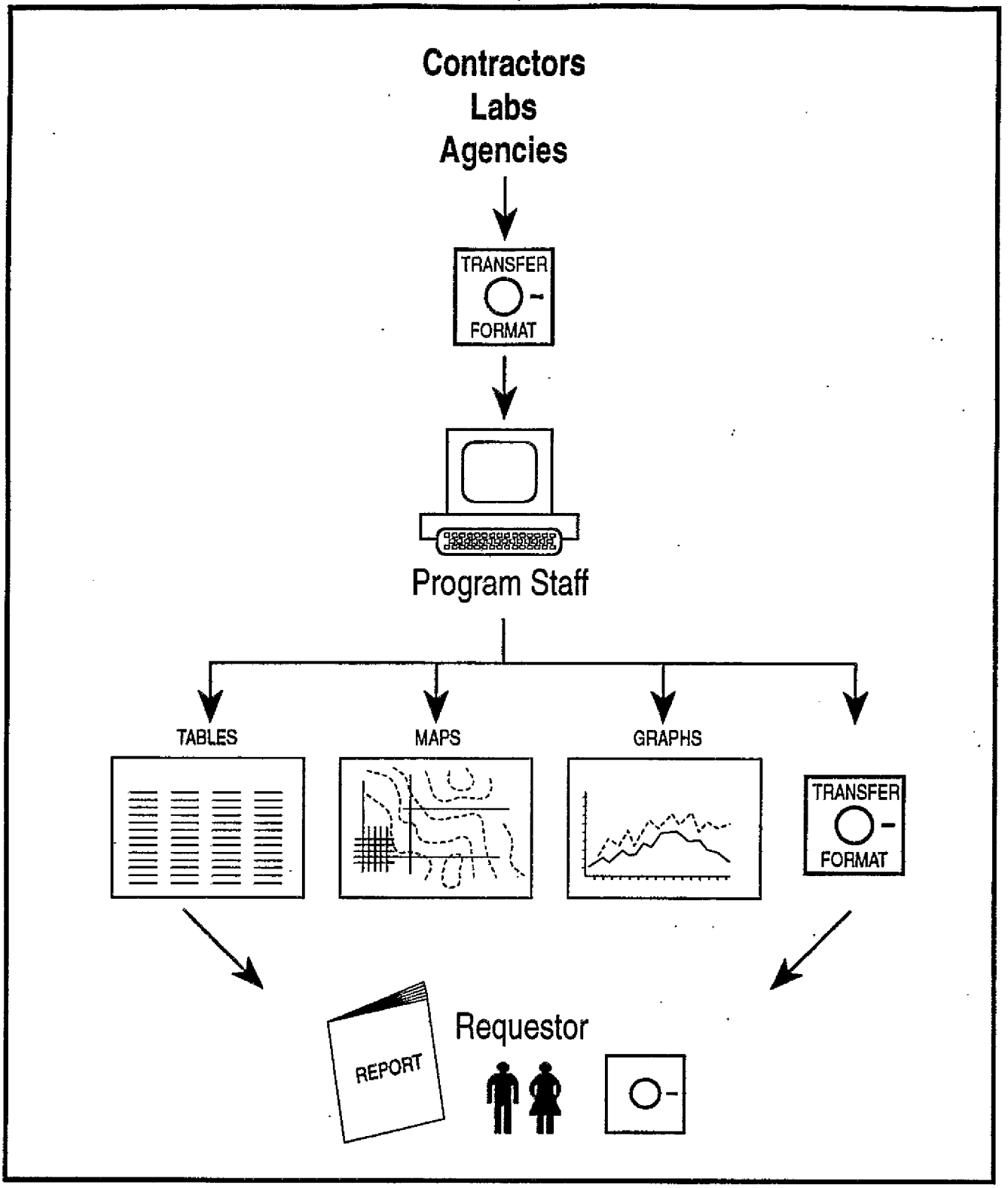


Figure 4. Medium-Term Approach #2. Use a Bi-State Program-Managed Database System.

- Agency staff could provide data in a standard transfer format to requesters, and also use the data themselves.

Disadvantages of this approach include:

- Bi-State Program staff levels, and expertise might not be adequate to manage data
- A Bi-State Program-based data management system would not be integrated with other agency databases
- Costs.

3.1.2.2 Use a Database System Managed by Another Organization (Figure 5). This approach would involve placing monitoring data collected by the Bi-State Program into an existing database system, managed and maintained by an organization other than the Bi-State Program. This option would require less active involvement on the part of Bi-State Program staff. Their activities would be limited to submitting data, dealing with quality assurance issues, and learning to retrieve data.

Costs of this approach would include acquiring hardware and software for communicating with the remote database, (@$2,500), and training ($5,000). Ongoing costs would include costs of any communications lines or long-distance connections (< $2,000/year), costs to add datasets to the database system (@$5,000-$20,000/year depending on the amount of data collected). Working with the database (conducting analyses, preparing reports, etc.) would be part of the Bi-State Program's regular activities.

Advantages of this approach include:

- Costs to the Bi-State Program would probably be lower than the costs to maintain a stand-alone database system
- Bi-State Program staff would not need the expertise to manage a database system

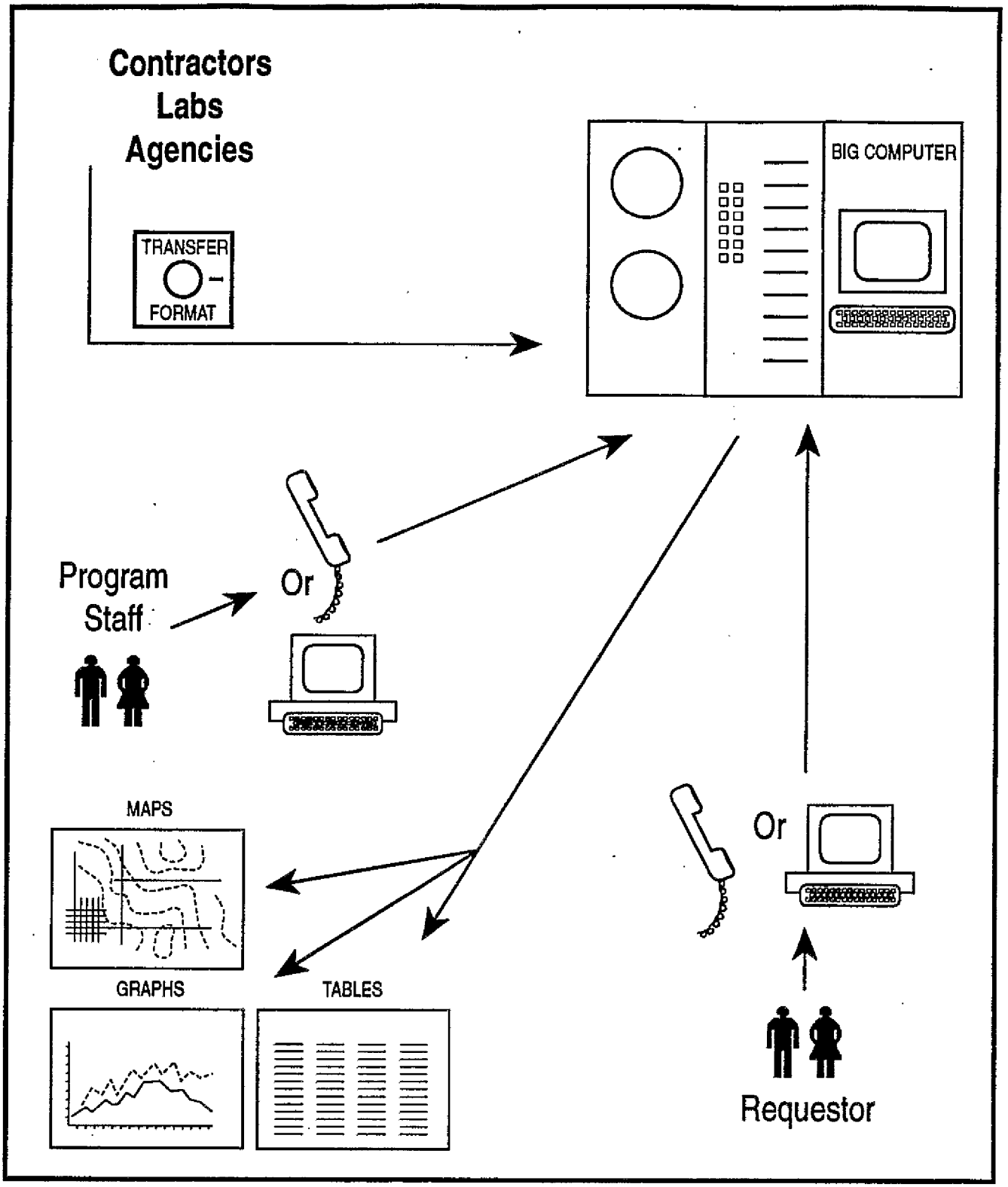


Figure 5. Medium-Term Approach #3. Use a Database System Managed by Another Organization.

- Data would be accessible to program staff, and other interested users
- Washington and Oregon would not need to manage separate copies of the database.

Disadvantages of this approach include:

- There may be costs to using another organization's existing database system
- Organizational arrangements would need to be worked out
- The Bi-State Program would have limited control over the distribution of data or results
- The Bi-State Program will have limited influence over the way data are stored in the selected system
- The Bi-State Program will have limited influence over changes to the selected database system. These changes may not always be in the Bi-State Program's best interest.

3.2 EVALUATION OF BI-STATE PROGRAM-MANAGED DATABASE SYSTEMS

In order to meet Bi-State Program needs, a database system managed by the Bi-State Program must have required, preferred, and technical elements (features) as defined in the Needs Assessment Report (Tetra Tech 1993) and modified through meetings with Bi-State Program staff:

- Required Elements are fundamental to the utility of the proposed data management system to the Bi-State Program

- Preferred Elements are those that enhance, but are not fundamental to, the utility of the proposed data management system

- Technical Elements are those elements that, though not always seen by a user, affect system performance, flexibility and life-span. The Technical Elements were added as a means of factoring the professional judgement of the system evaluators (the consultants) into the scores.

The technical appendix (Appendix A) contains detailed definitions of the Required, Preferred and Technical elements.

3.2.1 Evaluation Methods

A number of existing database systems, that could be managed by Bi-State Program staff were evaluated:

- NOAA's COMPAS (Coastal Ocean Mapping, Planning and Assessment System)

- Ecology's SEDQUAL (Sediment Quality Database)

- DAIS (Seattle District Corps of Engineer's Dredge Analysis Information System)

- PSAMP (Puget Sound Ambient Monitoring Program Database system).

Preliminary research indicated that the following could not be evaluated:

- EPA Region X Columbia River Dioxin Database: While EPA has begun to compile information on dioxin in the Columbia River, the project is still in its initial planning stages. No decisions about preferred hardware or software have been made, no data have been entered, and no consideration has been given to the development of menus, application programs, etc. While the Bi-State Program will probably want to use data compiled by this project, no "system" as such can be evaluated for Bi-State Program use.

Each evaluated system was scored on whether it contained the required, preferred and technical elements, described above (Section 3.2) and in Appendix A.

The evaluation was not based on judgment of the overall "quality" of a particular system, but on whether it meets the required, preferred, and technical elements. Each system was rated according the following scale for each element:

3 - Meets user requirements for this element (as is, or with minor changes)
2 - Nearly meets user requirements for this element
1 - Barely meets user requirements for this element
0 - Does not meet user requirements for this element

The technical appendix contains a detailed discussion of the ways that different systems evaluated have or don't have the Required, Preferred and Technical elements. Tables 3-6 show the scores for each element, and an overall unweighted score for each system.

There was no single system that had all the required elements; each system evaluated had pros and cons. System approaches to meeting the required elements differed widely, so that comparisons were often difficult to make. Each data management system is discussed briefly below.

3.2.2 COMPAS

COMPAS, the Coastal Ocean Mapping, Planning and Assessment System, was described as the "NOAA Desktop Information and Mapping System" in previous documents. This is an Apple MacIntosh microcomputer-based database in development by the NOAA Strategic Environmental Assessment Program in Rockville, MD. It combines database and mapping capabilities, using Oracle database management software and Atlas Pro mapping software.

As described by NOAA staff, COMPAS will contain mapping information on shoreline and coastal features; the features to be included (shorelines, towns, etc) can be customized for the Bi-State Program's needs. Each COMPAS version can also have different types of monitoring information stored in the Oracle database. A set of user-friendly menus will be available to set up the data files, manage data entry and editing, and to retrieve data from the files. By making a set of choices from a menu, a user will be

TABLE 3. DATABASE EVALUATION - MEDIUM-TERM NEEDS:
BI-STATE PROGRAM-MANAGED DATABASE SYSTEMS

Required Elements	COMPAS	SEDQUAL	DAIS	PSAMP
Existing Database	0.5[a]	3	3	3
Stores and Reports Data on:	3	1	0	2
Fish health/mortality	Y	Y	N	Y
Contaminants in biota	Y	Y	N	Y
Wildlife and fish communities	Y	N	N	Y
Invertebrate communities	Y	Y	N	Y
Sediment contaminants	Y	Y	Y	Y
Radionuclides in sediment	Y	Y	Y	Y
Water column contaminants	Y	N	N	Y
Water quality	Y	N	N	Y
Point sources	Y	N	N	N
Pollutant loadings	Y	N	N	N
Stores/Reports Geographic Data	3	0	0	0
Inexpensive to Modify	2	0	2	2
Inexpensive to Maintain	1	2	1	1
Consistent with Standards	NA	2	2	2
Data Quality Assurance	2	3	2	3
Import/Export Features	1	3	3	3
Compatible with DOS Computers	0	3	3	3
TOTAL UNWEIGHTED SCORE	12.5/27	17/27	16/27	19/27

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of required elements may be found in Appendix A.

TABLE 4. DATABASE EVALUATION - MEDIUM-TERM NEEDS:
BI-STATE PROGRAM-MANAGED DATABASE SYSTEMS

Preferred Elements	COMPAS	SEDQUAL	DAIS	PSAMP
Provide Results in Various Forms	3[a]	2	1.5	2
Generate Detailed and Summary Lists	3	3	2	2
Easy Access for Non-Experts	3	3	3	3
Mapping Capabilities	3	0	0	0
Stores and Reports Data on:	2	0	0	0
Dredging and disposal activities	N	N	N	N
Hydrologic data	Y	N	N	N
Changes in wetland resources	N	N	N	N
Changes in nearshore habitats	N	N	N	N
Monitoring Data can be Updated	2	2	2	2
Contains Most Recent Resource Data	1	0	0	0
TOTAL UNWEIGHTED SCORE	17/21	10/21	8.5/21	9/21

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of preferred elements may be found in Appendix A.

TABLE 5. DATABASE EVALUATION - MEDIUM-TERM NEEDS:
BI-STATE PROGRAM-MANAGED DATABASE SYSTEMS

Technical Elements	COMPAS	SEDQUAL	DAIS	PSAMP
State-of-the-Art Software	3[a]	3	3	2
Logical/Flexible Data Model	3	3	2	3
Adequate Performance	NA	2	2	2
Control	2	1	3	3
Necessary Functions:	2	2.5	2	2.5
Add/edit	Y	N	Y	Y
Automated checking	Y	Y	N	Y
Verify	N	Y	N	Y
Transfer in Any Standard Format	N	Y	Y	Y
Reports	Y	Y	Y	Y
Analyses	Y/N	Y	Y	N
Flexible Data Selection Criteria	3	3	1	2
Accessibility	2	3	2	2
TOTAL TECHNICAL SCORE	15/21	17.5/21	15/21	16.5/21

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of technical elements may be found in Appendix A.

TABLE 6. OVERALL SCORES - MEDIUM-TERM NEEDS: BI-STATE PROGRAM-MANAGED DATABASE SYSTEMS				
	Required (Total = 27)	Preferred (Total = 21)	Technical (Total = 21)	Overall (Total = 69)
PSAMP	19	9	16.5	44.5 (65%)
SEDQUAL	17	10	17.5	44.5 (65%)
COMPAS	12.5	17	15	44.5 (65%)
DAIS	16	8.5	15	39.5 (57%)

able to selectively retrieve monitoring data, calculate some simple statistics (e.g., count, mean, minimum), and place the data onto an existing shoreline map. Resource information can be indicated on the map using various colors. In addition, data will be able to be transferred to statistical packages, spreadsheets, charting software etc. for further analysis.

Advantages of using COMPAS include:

- Capabilities for presenting monitoring data and resource information on maps
- Use of Oracle database management software which provides extensive capabilities for data quality assurance and reporting
- Very user-friendly.

Disadvantages of using COMPAS include:

- COMPAS is designed for the Apple MacIntosh, and so can only be run from a DOS machine using a Local Area Network and protocol converters. Such an arrangement may be difficult to set-up and maintain
- Costs to use COMPAS will be higher than for other Program-managed databases, since more hardware will be required (@$10,000), more extensive set-up will be needed to handle Bi-state Program data types ($20,000-$30,000)
- COMPAS will not be available for use until March of 1994.

3.2.3 SEDQUAL

SEDQUAL is a microcomputer-based database system, developed in 1987-1988 by PTI, Inc. for EPA Region 10. SEDQUAL is designed for the calculation of Apparent Effects Thresholds for Puget Sound sediments, and is currently in use by the Department of Ecology Sediment Management Unit to manage Puget Sound sediment data.

SEDQUAL was developed for DOS microcomputers using FoxPro software. While data cannot be directly entered into SEDQUAL, they can be entered into a spreadsheet using a pre-defined format and transferred to the system. SEDQUAL allows sampling stations and chemicals to be grouped into temporary groups for retrieval and analysis. Menus lead a user through data grouping, data reporting, comparison to existing standards, and calculations of sediment quality values.

Advantages of using SEDQUAL include:

- User-friendly
- Very flexible data retrieval and data grouping capabilities.

Disadvantages include:

- Inconvenient data loading capabilities
- SEDQUAL does not contain many of the required data types
- SEDQUAL cannot be modified to handle these data types since SEDQUAL source code cannot be obtained for a reasonable cost (<$100,000).

3.2.4 DAIS

DAIS, the Dredge Analysis Information System, is a microcomputer-based system, developed in 1989-1991 by the Seattle District of the U.S. Army Corps of Engineers. It is designed for managing pre-dredge monitoring data under the Puget Sound Dredge Disposal Analysis (PSDDA) program. Detailed data on sediment quality and bioassay results are submitted in a pre-defined format by dredge permit applicants.

DAIS was developed for DOS microcomputers using FoxPro software. It stores an extensive amount of detailed data and quality assurance information about the sediment and bioassay testing. It uses this information to evaluate the results of the tests. There are a set of menu-driven programs for loading data

from submitted files and calculating statistics. There is a very limited set of automated data checks performed on the data.

Advantages of using DAIS include:

- User-friendly
- Automatically summarizes data
- Stores extensive quality assurance information.

Disadvantages include:

- Stores a limited number of data types
- Retrieval capabilities are limited. Data can only be reported by dredging project or dredging year
- No mapping capabilities.

3.2.5 PSAMP

PSAMP, the Puget Sound Ambient Monitoring Program database system, is a microcomputer-based database system developed in 1989-1990 (by consultants Roberta P. Feins and Nels Christianson) for the Puget Sound Water Quality Authority. It was designed to serve the Puget Sound Water Quality Authority as a central database for the program - to allow integration of monitoring data collected by a number of different agencies. It is also being used by two agencies (Ecology and Fisheries) to manage their own PSAMP data.

PSAMP was developed for DOS microcomputers using dBase IV software. It provides facilities for full-screen data entry and for loading data transferred in PSAMP data transfer format. There is a module for verifying data, for deriving statistics, for reporting information and transferring data to other systems.

Advantages of using PSAMP include:

- User-friendly.
- Can load data in PSAMP format, and output data in PSAMP or ODES format
- In use for several years at 3 different state agencies in Washington
- Flexible data retrievals
- Stores almost all desired data types.

Disadvantages include:

- No mapping capabilities
- Sometimes slow in retrievals and formatting
- Very limited custom report capability
- Uses dBase IV software, considered by some to be out-of-date.

3.2.6 Recommended System

The PSAMP and COMPAS database systems could both be used successfully by the Bi-State Program. The COMPAS system will require more effort to set-up (@$20-30,000), but will provide excellent easy-to-use data retrieval and mapping capabilities. The PSAMP system will have lower set-up costs (@$5,000-$10,000) but will provide limited analyses and no mapping capabilities. The PSAMP system is currently available from the Puget Sound Water Quality Authority, while COMPAS will not be available for use until March of 1994.

3.3 EVALUATION OF OTHER ORGANIZATION-MANAGED DATABASE SYSTEMS

In order to meet Bi-state Program needs, a database system managed by another organization must have required, preferred and technical elements (features) as defined in the Needs Assessment Report (Tetra Tech 1993) and modified through meetings with Bi-State Program staff:

- Required Elements are fundamental to the utility of the proposed data management system to the Bi-State Program

- Preferred Elements are those that enhance, but are not fundamental to, the utility of the proposed data management system

- Technical Elements are those elements that, though not always seen by a user, affect system performance, flexibility and life-span. The Technical Elements were added as a means of factoring the professional judgement of the system evaluators (the consultants) into the scores.

The technical appendix (Appendix A) contains detailed definitions of the Required, Preferred and Technical elements.

3.3.1 Evaluation Methods

A number of database systems that are managed by other organizations were evaluated for use by the Bi-State Program:

- EPA's STORET

- EPA's ODES (Ocean Data Evaluation System)

- Idaho's EDMS (Environmental Data Management System)

- Northwest Power Planning Council's CRCIS (Columbia River Coordinated Information System).

The evaluation was not based on judgment of the overall "quality" of a particular system, but on whether it meets the required, preferred, and technical elements. Each system was rated according the following scale for each element:

3 - Meets user requirements for this element (as is, or with minor changes)
2 - Nearly meets user requirements for this element
1 - Barely meets user requirements for this element
0 - Does not meet user requirements for this element

The technical appendix contains a detailed discussion of the ways that different systems evaluated have or don't have the Required, Preferred and Technical elements. Tables 7-10 show the scores for each element, and an overall unweighted score for each system.

There was no single system that had all the required elements; each system evaluated had pros and cons. System approaches to meeting the required elements differed widely, so that comparisons were often difficult to make. Each system evaluated is discussed briefly below.

3.3.2 STORET

STORET is EPA's major water quality database, and has been in existence for at least 15 years. STORET is the most widely used water quality database in the U.S.. It is located on an EPA mainframe computer in North Carolina.

STORET would not be able to handle all of the data types and all of the information the Bi-State Program needs. STORET was not originally designed to handle sediment, bioassay, or tissue data and does so in a clumsy way, that does not capture minimum essential information. The quality assurance information that can be stored for any data type is very limited.

Because STORET is so long-lived, it uses out-dated computer technology and its capabilities for working with data are very limited. STORET does not use consistent sets of parameter or method codes, so that retrieving and combining data from different datasets is difficult. Most people who use Storet tend to down-load data to personal computers, and use spreadsheets, statistical packages and databases to analyze it. STORET's mapping capabilities are extremely limited.

TABLE 7. DATABASE EVALUATION - MEDIUM-TERM NEEDS: OTHER ORGANIZATION-MANAGED DATABASE SYSTEMS

Required Elements	STORET	ODES	Idaho EDMS	CRCIS
Existing Database	3[a]	3	2	3
Stores and Reports Data on:	2	3	1	1
Fish health/mortality	N	Y	N	N
Contaminants in biota	N	Y	N	N
Wildlife and fish communities	N	Y	N	Y
Invertebrate communities	N	Y	N	N
Sediment contaminants	Y	Y	N	N
Radionuclides in sediment	Y	Y	N	N
Water column contaminants	Y	Y	Y	N
Water quality	Y	Y	Y	N
Point sources	N	N/Y	Y	N
Pollutant loadings	N	N/Y	N	N
Stores/Reports Geographic Data	0	1.5	2	2
Inexpensive to Modify	0	1	1	2
Inexpensive to Maintain	3	2	2	3
Consistent with Standards	NA	NA	NA	2
Data Quality Assurance	1	3	2	1
Import/Export Features	2	3	1	1
Communicates with DOS Computers	3	3	3	3
TOTAL UNWEIGHTED SCORE	14/27	19.5/27	14/27	18/27

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of required elements may be found in Appendix A.

TABLE 8. DATABASE EVALUATION - MEDIUM-TERM NEEDS:
OTHER ORGANIZATION-MANAGED DATABASE SYSTEMS

Preferred Elements	STORET	ODES	Idaho EDMS	CRCIS
Provide Results in Various Forms	1[a]	3	2	2
Generate Detailed and Summary Lists	2	3	2	1
Easy Access for Non-Experts	2	3	2	3
Mapping Capabilities	1	2	2	2
Stores and Reports Data on:	1	0	0	1.5
Dredging and disposal activities	N	N	N	Y/N
Hydrologic data	Y	N	N	Y
Changes in wetland resources	N	N	N	N
Changes in nearshore habitats	N	N	N	N
Monitoring Data can be Updated	3	2	1	2
Contains Most Recent Resource Data	0	0	0	3
TOTAL UNWEIGHTED SCORE	10/21	13/21	9/21	14.5/21

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of preferred elements may be found in Appendix A.

TABLE 9. DATABASE EVALUATION - MEDIUM-TERM NEEDS:
OTHER ORGANIZATION-MANAGED DATABASE SYSTEMS

Technical Elements	STORET	ODES	Idaho EDMS	CRCIS
State-of-the-Art Software	1[a]	1.5	3	2
Logical/Flexible Data Model	1	2	2	1
Adequate Performance	2	3	2	2
Necessary Functions:	2	2	1	2
Add/edit	N	N	N	Y
Automated Checking	Y	Y	N	Y
Transfer in Any Standard Format	N	N	Y	N
Reports	Y	Y	Y	Y
Analyses	Y	Y	N	Y/N
Flexible Data Selection Criteria	1	2	1.5	1
Accessibility	2	2	1	1
TOTAL TECHNICAL SCORE	9/18	12.5/18	10.5/18	9/18

[a] 3 = Meets user requirements.
2 = Nearly meets user requirements.
1 = Barely meets user requirements.
0 = Does not meet user requirements.
NA = No information available.

Detailed definitions of technical elements may be found in Appendix A.

TABLE 10. OVERALL SCORES - MEDIUM-TERM NEEDS: OTHER ORGANIZATION-MANAGED DATABASE SYSTEMS				
	Required (Total = 27)	Preferred (Total = 21)	Technical (Total = 18)	Overall (Total = 72)
ODES	19.5	13	12.5	45 (68%)
CRCIS	18	14.5	9	41.5 (63%)
IDAHO EDMS	14	9	10.5	33.5 (51%)
STORET	14	10	9	33 (50%)

EPA is currently proceeding with a major project to update and redesign STORET. The intent is not to rebuild STORET as a complete data management and analysis tool, but to build a more user-friendly data archive, with extensive capabilities for transferring information to user computers for further manipulation and analysis (King, R., 31 March 1993, personnal communication).

Costs of using STORET would include acquiring hardware (e.g., computers, modems) and software for communicating with the remote database, (@$2,500), and Bi-State Program staf training ($5,000). Ongoing costs would include costs of any communications lines or long-distance connections (< $2,000/year), and costs to add datasets to the database system (@$1,500-$2,000/monitoring study). There would be additional costs to working with the database, since data must be down-loaded and reformatted before being analyzed.

3.3.3 ODES

ODES (Ocean Data Evaluation System) is a mainframe-based database developed in the early 1980's by Tetra Tech, Inc. and AMS Inc. for the EPA Office of Marine and Estuarine Programs, Washington DC. Originally designed to store monitoring data from the 301(h) ocean discharge program, ODES is currently used to store a variety of marine and estuarine data.

ODES was designed as a data archive and a data analysis system. It can store most of the data types required by the Bi-State Program. ODES can store quality assurance information including data from lab blanks, matrix spikes, duplicate analyses etc. Before a data set can be archived in the ODES database, it goes through a series of automated data checks, and technical review to assure data quality.

ODES tools allow data selected by area, and date to be listed and analyzed in a number of ways . Users can get simple data listings, bar and line graphs, and conduct t-tests, analysis of variance, cluster analysis and other statistical procedures on selected data. ODES has an enhanced option that allows customized selections of data to be downloaded to personal computers for additional analyses.

Costs of using ODES would include acquiring hardware (e.g., computers, modems) and software for communicating with the remote database, (@$2,500), and Bi-State Program staff training ($5,000). Ongoing costs may include costs of any communications lines or long-distance connections (< $2,000/

year), although access to ODES may be achieved through local connections to the Regional EPA computer or through an 800 phone number. Other ongoing costs would be to add datasets to the database system (@$1,000/data type; 3-6 data types per monitoring study). Data can be worked with in ODES so the time required to do analyses would be much less than for STORET, where data need to be donwloaded and reformatted for use.

The ODES system had the highest overall score for the database systems managed by other programs. The advantages of using ODES are:

- System is accessible via modem to any user with a valid user ID
- ODES stores and allows analysis of most of the data types of interest to the Bi-State Program
- ODES provides on-line data retrieval, graphing, and analysis capabilities.

The disadvantages of using ODES are:

- Organizational arrangements must be worked out with EPA's Office of Water in Washington DC, a group that may not have any stake in allowing the management of Columbia River data on ODES
- Costs to enter data into ODES are relatively high (about $5,000) per monitoring survey, due to ODES stringent quality control requirements.
- The ODES system may not be in existence for more than 2-3 years; it may be replaced by a modernized STORET. However, if this occurs, the data in ODES will be converted into the modernized STORET.
- ODES mapping capabilities are limited.

3.3.4 Idaho EDMS

The Idaho Environmental Data Management System (EDMS) is a groundwater quality and well data database developed in 1989-1991 by the Idaho State Department of Water Resources. It is used as a repository of groundwater quality data from all over the state. In the future, the state of Idaho plans on adding air quality and surface water quality modules to the database. The main task of EDMS is to provide this data to users in reports, forms, charts, and maps. In the near future, they plan on adding a query generator so that specialized forms and tables can be created.

EDMS cannot store all of the data types desired by the Bi-State Program; in fact there are no plans to add sediment data to the system. In addition, EDMS is not set up to accomodate remote users or users from other states; the organizational arrangements required, and the costs are unknown.

3.3.5 CRCIS

CRCIS is a system of databases maintained by the Northwest Power Planning Council, the Bonneville Power Administration (BPA), and state environmental, and resource agencies. CRCIS is a sub-set of the Northwest Environmental Database (NED), which was originally developed to provide information on potential hydroelectric sites on rivers in the Northwest. The Northwest Environmental Database consists of:

- Local menu-driven databases of resource information maintained by a designated state environmental agency. These databases contain information on the presence/absence of important resources (fish, spawning areas, waterfalls, cultural resources) along thousands of river segments.

- A centralized copy of information integrated from all of the local databases, maintained by the Northwest Power Planning Council.

- Mapping and geographic analysis capabilities, provided on an informal basis by the Bonneville Power Administration.

The Northwest Environmental Database is not appropriate for Bi-State Program needs, because it does not store detailed monitoring information in a form useful to the Bi-State Program. The evaluation of

CRCIS assumes that Program data would be sent to the Northwest Power Planning Council or BPA and kept in special files separate from the rest of the database. The Bi-State Program could get maps or analyses of the data by calling one of these organizations. This is a substantially different arrangement from using STORET or ODES where Program staff could dial the database and conduct their own retrievals or analyses. Bi-State Program staff would need to make arrangements with the Northwest Power Planning Council or the BPA every time a simple report, chart, or analyses was needed or requested. For users other than the Bi-State Program staff, data would only be available on a floppy disc or other medium in a standard format. However, use of CRCIS would provide access to mapping capabilities that are not available for the other alternatives.

Costs to the Bi-state Program to use CRCIS are unclear. The Bi-State Program would need to select a data format and reformat existing data or reenter data ($1,500-$2,000 per dataset). Data would then be sent to NWPPC or BPA. If analyses or maps are required, Bi-State Program staff may need to pay BPA staff for data retrievals and analyses. A cooperative agreement between the Bi-State Program and NWPPC/BPA would be required.

3.3.6 Recommended System

As with the "Program-managed" approach, there was no single system that had all the required elements for this approach; each system evaluated had pros and cons. System approaches to meeting the required elements differed widely, so that comparisons were often difficult to make.

A main-frame based Federal database system would be the best approach if the Bi-State Program wanted to minimize local responsibility for and control over the data, and minimize costs, while providing on-line data analysis capabilities. The ODES system had the highest overall score for the database systems managed by other programs. In selecting a main-frame database, ODES would provide extensive capabilities for data use and analysis, high data quality assurance, and greater access to data.

While STORET is the oldest, most widely used database system for environmental monitoring data, its ability to meet Bi-State Program needs is severely limited. Use of STORET would really be use of an existing data archiving system, not use of a functional database system to meet program needs. On the other hand, STORET, and particularly the modernized STORET, would provide a more well-known, long-lived, and lower cost alternative.

If STORET is to be used for the Bi-State Program, the Bi-State Program needs to work with data collectors to agree on standards for data collection, quality assurance, and reporting that are more stringent than those required for entry into STORET. Otherwise data will not be comparable. However, with the planned STORET modernization, many of these existing problems may be accounted for. If this occurs, the modernized STORET may look much more attractive than the existing STROET version.

If the Bi-State Program wants to work with a local database, and does not need on-line analysis capabilities, then working out an arrangement with the Northwest Power Planning Council and the BPA for being part of CRCIS, would be an attractive alternative. While CRCIS would not be a viable alternative for a database management system, CRCIS would be an attractive hybrid approach between maintaining data in a data archive, and using an on-line database management system. If Bi-State Program staff needed occasional reports, analyses or maps, being able to call on BPA expertise using data archived in CRCIS could be a satisfactory approach. Access to GIS through BPA would give the Bi-State Program access to mapping capabilities not found in any of the other systems evaluated.

3.4 OVERALL MEDIUM-TERM RECOMMENDATIONS

A medium-term system cannot be selected until the key questions raised in Section 3.1 are answered:

- How frequently do Bi-State Program staff need to access and use Bi-State Program data?
- Does the Bi-State Program want to and is the Bi-State Program able to manage its own database system?

Once these questions are answered, an approach, and a data management system can be selected (Figure 2).

3.4.1 Influence on Short-term Approach

If a database system can be selected for the medium-term before the short-term approach is implemented, then the approaches should be compatible. For example, if ODES is selected as the medium-term approach, ODES data submittal format should be required in Bi-State Program contracts and cooperative

agreements. This will eliminate the need for reformatting data as they move from the short-term archive to the medium-term system.

4.0 LONG TERM NEEDS/APPROACHES

In the long-term (greater than 5 years), the data management objectives of the program are to ensure cooperative sharing of all available information on the lower Columbia River, in order to improve environmental decision-making.

The Bi-State Program may not continue to exist into the long-term time frame. However, the needs for evaluation and management of the lower Columbia River will remain. With that in mind, one of the goals of the Bi-State Program is to make recommendations about a long-term institutional framework for monitoring and management. Currently, a steering committee through the office of the governors of both states is considering the issues. The committee should at least consider taking steps to address issues of data sharing.

The ability to access and share data is one functional aspect that must be addressed in developing this framework. The programmatic needs assessment revealed that there is a need for better information, better access to and sharing of existing information, and better use of information in order to manage the river's resources.

The areas in which information sharing needs to be improved include:

- Information about information: what data are available, what projects are going on, etc.
- Permit information: who has permits in a given reach, what are loading and permit limits
- Resource information: where are resources located? will they be impacted by a proposed project?

Some approaches that should be considered for meeting these long-term needs include:

- Development of a newsletter, on-line bulletin board or other ways to keep both inter- and intra-state staff aware of agency activities in areas that affect the lower river

- Ways to provide immediate or short-term (few-day) access to information by permit managers, proposal reviewers, etc.

- Agreement on minimum requirements for quality assurance in data acquisition, spatial positioning, sample analysis

- Agreement on a standard format for transfer of data of various types

- Ability to integrate lower Columbia River data into and get data out of existing agency databases including:

  - Permit databases (e.g., Ecology WPLCS)

  - Dredging databases (e.g., PSDDA database, Portland Corps of Engineers database)

  - Resource databases (e.g., National Wetlands Inventory, Northwest Environmental Database, Oregon and Washington Rivers Information system.

  - State and Federal land use databases

- Periodic publication of an Atlas or "Big Map" of the lower river summarizing resources, activities and conditions.

4.1 CURRENT DATA INTEGRATION EFFORTS

4.1.1 National Water Information System (NWIS II)

The U.S. Geological Survey (USGS) is currently completing Phase I of the development of the National Water Information System II (NWIS II). This system is a successor to the USGS WATSTOR database, which is about 20 years old, and outmoded. USGS information management staff and scientists have spent considerable time designing NWIS II, which will manage groundwater and water column data.

NWIS II will operate with nodes (UNIX workstations running INGRES database) in regional USGS offices. Phase I of NWIS II is scheduled to be completed by April of 1993, with Phase II scheduled for October. There may be some funding problems.

Ecology is currently following the development of this USGS system. It is possible that USGS will allow Ecology to use the database application either by directly accessing USGS regional office computers running NWIS II or by providing the agency with a copy of the application for their use, with the data eventually transferred to USGS or with Ecology becoming an additional data center.

NWIS II could be an attractive option for the Bi-State Program, especially if Oregon were also looking at using it for water quality data management. In addition, EPA is also looking at NWIS II's data architecture for use in their major 5-year STORET revision.

While Ecology is very interested in this system, discussions about its use are still very preliminary. Costs are unknown. Meanwhile, several Ecology programs that are working on data standardization are looking at adopting the NWIS II data model.

Since NWIS II is designed to handle environmental monitoring data, at best, it could only be used for storing some of the data needed for long-term lower Columbia River management.

4.1.2 ORACLE 7/SQL Databases

Current database technology is moving towards database systems that all use the same Structured Query Language (SQL) to manage and report data. In the future, a programmer will be able to write programs

in Oracle, dBase or Paradox to access the same information - the choice of tool will be based on the desired functions, and desired look of the final result.

Oregon DEQ and Ecology are both using Oracle as their "high-end" high-capability database. Oracle Version 7, due in late 1993, will allow distributed data processing across both local and wide-area networks. Oracle could become the backbone of a multi-state integrated database system.

4.1.3 ARC/INFO Geographic Information System

Both Oregon and Washington are using Arc/Info to manage geographic information on conditions and resources. As time passes, these project-based systems may be integrated into agency-wide GIS systems, using databases such as Oracle, or dBase. Such an approach would allow integration of lower Columbia River monitoring and resource data and integration with other project and program data.

4.2 THE IDEAL LONG-TERM APPROACH

If any single database application such as NWIS II is selected for long-term management of lower Columbia River data, it will probably not meet long-term program needs. If one agency system stores water quality data, another will likely be used to manage permit information, sediment data, or dredging information. Wetlands and near-shore habitat information will be in a geographic information system.

Figure 6 shows an ideal long-term approach to cooperative data sharing among agencies and organizations. Instead of giving up control over their data, or moving data on one geographic area into a special database, each agency maintains one or more databases on its own equipment. Such databases are linked together through hardware "bridges" and standard software query languages. Menu-driven applications are written for a specific purpose (e.g., to calculate sediment standards, or update a Columbia River Atlas), but the applications access a shared database. Each user would see a system that looked like it was configured just for her, but would appear so due to the menus; the underlying system structure would be flexible and allow many different "views" of the same data.

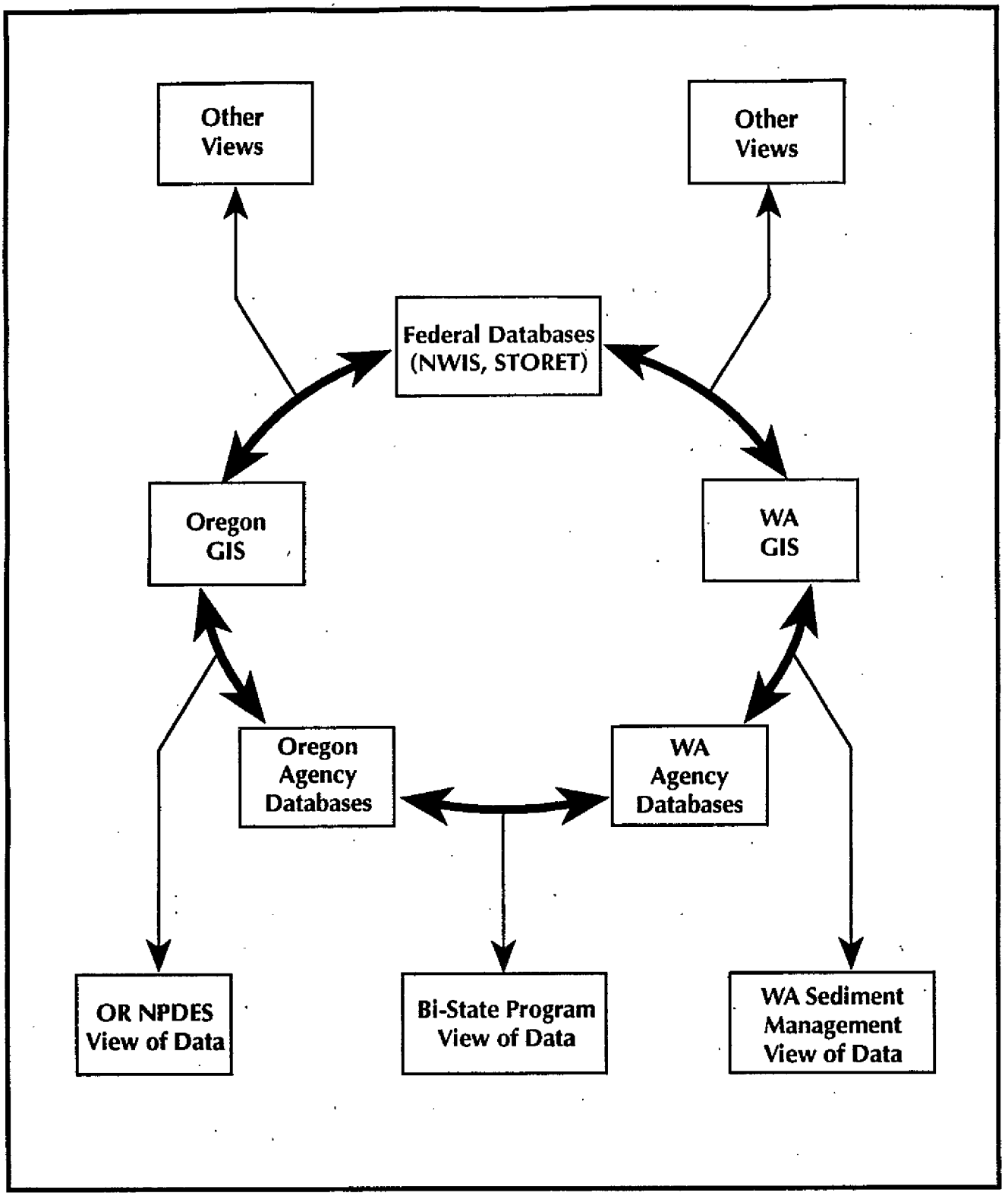


Figure 6. Schematic Diagram of Ideal Long-Term System.

Such an approach can link together different programs in a state agency, different state agencies, agencies from different states, Federal, local, and tribal governments, businesses, and individuals. It would have the following advantages for the Bi-State Program:

- Information on the Columbia River from different agencies and on different topics could be integrated and analyzed

- The Bi-State Program would not have to reenter or reformat data of interest for their own particular database, but would be able to tap into and provide data to a growing pool of information

- The Bi-State Program would not have to maintain or control a database of information, but could still have access to the information

- The agencies (e.g., Oregon Department of Environmental Quality) could provide a much broader set of analysis tools for staff in various agency programs.

The ideal approach described above represents a growing consensus among information management professionals in various agencies about future direction. However, movement towards the ideal is slow, due to lack of funding, lack of attention and priority paid to environmental information systems within agencies, lack of standardization, and the difficulties in beginning to work through the complex of organizational arrangements that will be required.

4.3 RECOMMENDATIONS

No crystal ball is available to determine exactly what databases, software and networks will link environmental agencies by the year 2000. Since technology is changing so quickly, it would be a mistake to plot out a long-term course at present. However, such a course is sure to involve data integration.

The Bi-State Program cannot embark single-handedly on a long-term quest for better data integration. However, the Bi-State Program can participate in long-term information planning efforts (such as those

of the Washington Geographic Information Council and the Northwest Land Information Systems Network) and advocate approaches that will best serve the data coordination needs of the lower Columbia River. These approaches are as follows:

- The Bi-State Program should encourage the discussion of data sharing and data integration as part of the Steering Committee's development of a long-term institutional framework for Columbia River monitoring and management.

- A working group of data experts from the states and interested Federal agencies should meet to explore issues, and make recommendations to the Steering Committee for a long-term approach.

- The Bi-State Program should seek to take advantage of and encourage efforts to improve access to existing agency data needed to make management decisions.

5.0 REFERENCES

King, R. 31 March 1993. Personal Communication (discussion with Ms. Roberta Feins). U.S. Environmental Protection Agency, Washington, DC.

Oregon State Map Advisory Council. 1990. Digital Spatial Standards ad Procedures. Prepared by the GIS Standards and Procedures Working Group, Oregon State Map Advisory Council. March 22, 1990. 13 pp.

Tetra Tech. 1992. Data management work plan. Final report. Prepared by Tetra Tech, Inc., Bellevue, Washington. Prepared for the Lower Columbia River Bi-State Water Quality Program. 14pp.

Tetra Tech. 1993. Data Management NEEDS Assessment Report. Final Report. Prepared by Tetra Tech, Inc., Redmond, Washington. Prepared for the Lower Columbia River Bi-State Water Quality Program. 25 pp. + appendices.

Webster, G.L. 1992. Bi-State Lower Columbia River Program programmatic needs assessment. Final report. ODEQ Contract No. 41-92. 27 pp. + appendices.

APPENDIX A

TECHNICAL APPENDIX

TECHNICAL APPENDIX

In this appendix, detailed definitions of the required, preferred and technical elements of a Bi-State Program data management system are provided. Examples are given to show the ways in which the different systems evaluated have or don't have these elements.

REQUIRED ELEMENTS

Required elements are those that are fundamental to the utility of the proposed data management system.

Existing Database: Is the system an existing database or database application? Of the systems evaluated, only COMPAS is not yet fully developed (expected to be completed 8/93; test version available at Oregon Dept of Lands at this time).

Stores and reports data on: Can the system store and report environmental monitoring data?:

- Fish health/mortality: results of fish pathology analyses, other conditions in fish
- Contaminants in biota: concentrations of chemicals in the tissues of various organisms of interest
- Wildlife and Fish communities: identification and enumeration of species sampled using various methods (transects, trawls)
- Invertebrate communities: identification and enumeration of marine benthos sieved from bottom sediments

- Sediment contaminants: concentrations of chemicals in bottom sediments

- Radionuclides in sediment: concentrations or counts of radionuclides in bottom sediments

- Water Column contaminants: concentrations of chemicals measured at various depths in the water column

- Water Quality: measurements of conventional water quality parameters (e.g. bacteria, total suspended solids, dissolved oxygen, ammonia, nutrients, total dissolved gases) measured at various depths in the water column

- Point sources: lists and descriptions of potential discrete sources of contamination - including NPDES permittees.

- Pollutant Loadings: flow, measurement and calculated loading data from point sources or river tributaries.

Not one of the systems evaluated stores all of these types of information; ODES and PSAMP came closest. DAIS and SEDQUAL only store information on sediment "triad" information; SEDQUAL also stores fish tissue data. COMPAS will be set up so that a user can store whatever information he or she wants in the database; the disadvantage of this approach is that someone must design the data files. CRCIS does not contain any of the data types.

Stores/reports Geographic Data: Does the system store location information for point resources (i.e. latitude/longitudes for outfalls). Can it delineate areas defined as wetlands, areas used for different land uses, or those containing certain resources? Can it report this data, not simply by printing a list of latitude/longitude points, but by producing maps and map overlays? Can it do any kinds of geographic analysis (e.g., show the intersection of two different types of data, or calculate acreage in different land uses)?

All of the systems store location information for points. However, onyl ODES, COMPAS, and CRCIS have any kind of mapping capabilities, and their abilities are limited. For example, ODES can display information from digitized maps, but has very limited capabilities to overlay information, and cannot do geographic analysis, COMPAS will be able to store geographic information, but only in a "grid-based" format. This means the system will be able to produce maps showing whether a given map cell (of a certain area) has a certain resource or does not have a certain resource. This is very different from a GIS system such as Arc/Info which allows exact mapping of the boundaries of a resource such as a wetland. However, Arc/Info can directly access data in the COMPAS database, Oracle, which means that geographic analysis could be conducted by GIS experts on data already in the COMPAS database.

CRCIS provides full Arc-Info capabilities but not in a way that is easily accessible to Program staff; a BPA expert must do the analyses and produce maps each time.

Currently, there is no comprehensive set of geographic data available for mapping in either system. The best sources of digital geographic information are the U.S. Geological Survey, and other Federal and state agencies who collect this information as part of their legal mandate (e.g., BLM collects land use information).

Inexpensive to modify: If the system were to be used by the Bi-State Program and were to need modification (additional programming, new data types), would this be a relatively expensive or inexpensive task?. The evaluation considered such factors as the programming language in which the system was developed, whether only one consultant had the knowledge to make changes to the system, the organizational constraints to modification.

Planning is currently underway for STORET modernization. However, it is unlikely that the Bi-State Program would be able to make any modifications to the STORET system to meet their medium-term needs. ODES is managed by the U.S. EPA Office of Water, and EPA may be willing to make modifications to ODES; the Bi-State Program would likely have to pay for these modifications. In any case, ODES modification would likely be quite expensive, since there a limited number of people with the ability to program in its software. Adding a new type of data to ODES could cost $5,000-$10,000.

Systems like SEDQUAL, DAIS, and PSAMP are all written in 4th generation commercial database software (FoxPro and dBase IV). There is a large base of programmers familiar with the software and programming is easier. For example, adding a new type of data to PSAMP would probably costs about $2,000. However, SEDQUAL is available in a compiled version only; the programs use to run it are not available in a programming language that can be modified. In the past, PTI Inc. has quoted a figure of $100,000 for purchasing the original programs.

COMPAS is written in C and Oracle, currently used by both the Washington Department of Ecology and the Oregon Department of Environmental Quality for managing large agency database.

Inexpensive to maintain: If the Bi-State Program were to use the system for data management, what costs would it incur? The evaluation considered such factors as costs to enter data into the system, and whether the Bi-State Program or another entity must provide and operate hardware and software.

All of the database systems that would require management by the Bi-State Program (SEDQUAL, DAIS, COMPAS, PSAMP) would be most expensive to maintain, since the Bi-State program would need to provide hardware and software, operate the database, load data into the database, provide security, and provide training. To keep such a system running, and the data up-to-date, could require between 15% and 20% of a staff-person's time.

The costs to maintain ODES and STORET are much lower, since most of the maintenance burden is borne by EPA. There may be connection charges to dial-in and use STORET or ODES. In addition, ODES costs to add data are high, due to ODES data checking and quality assurance process.

Costs become even higher if the Bi-State Program chooses a mapping system or GIS. The costs to turn map information into digital format, to a level of accuracy that meets program needs can be quite expensive. A 1990-1992 effort to use a GIS to update the Puget Sound Environmental Atlas cost over $200,000. About 20 different types of resource and monitoring data were updated at a scale of 1:100,000. Some of the data were already in digital format.

Consistent with standards: Is the system consistent or compatible with existing or developing state data management or geographic information system (GIS) standards? The two Federal systems, ODES and

STORET set their own data management standards. At the state levels, there are no state-wide data management standards for Oregon or Washington. SEDQUAL and DAIS have been developed in accordance with data standards for the Puget Sound Dredge Disposal Analysis Program (PSDDA); PSAMP system in accordance with those for the PSAMP program.

While Oregon has GIS standards, these were not applicable to any of the systems evaluated, with the possible exception of COMPAS. Not enough information on COMPAS is currently available to assess compatibility.

Data quality assurance: Are procedures in place to ensure data quality and integrity throughout the data "life cycle" - from collection in the field to computer entry to final use? Factors considered include: requirements for technical data review, automated data checking programs, use of well-designed standardized codes, protection against careless or unauthorized modification.

The STORET system has a bad reputation with respect to data quality for several reasons:

- There is very little checking of the incoming data for errors or omissions
- STORET uses many duplicate codes for the same chemical and biological variables. This makes it very difficult to extract a consistent data set from STORET, or compare data from different data sets.

ODES provides extensive data quality assurance checking. Each data set is run through automated data checks, and reviewed by a technical reviewer. The data set may then be returned to the submitter for correction and resubmittal. An abstract about each data set and its quality is available to each ODES user.

Immediate availability: Covered under existing database, above.

Import/export features: Does the system have the capacity to facilitate data exchange through import/export features? Can the system output data into ASCII text files that can be reformatted for use

in spreadsheets, word-processors and other databases? In addition, can the system output data into a standard archive format (e.g., STORET, PSAMP) for transfer to other databases?

All of the systems evaluated can output data in ASCII text files; there are download options in both STORET and ODES. In addition, the systems evaluated can read and write the following standard formats:

- STORET: READ: STORET format. WRITE: none
- ODES: READ: ODES format. WRITE: None.
- SEDQUAL: READ: PSAMP, PSDDA formats. WRITE: PSAMP, PSDDA formats.
- DAIS: READ: PSDDA format. WRITE: PSDDA format.
- PSAMP: READ: PSAMP format. WRITE: ODES format, PSAMP format, spreadsheet (comma-delimited) format.

Ability to communicate with DOS Computers: Does the system operate on a DOS-compatible micro-computer? Can it be run from such a computer, via a modem, or hook-up to a local area network? How straight-forward is such a set-up?

The two EPA systems operate on DOS computers via modems that connect to EPA's mainframes in North Carolina. Both require use of terminal emulation software that can limit, or slow down the types and amount of information displayed on the screen.

SEDQUAL, PSAMP and DAIS run directly on DOS computers and make full use of menus, pop-up windows, and colors. COMPAS is designed for the Apple MacIntosh, and so can be run from a DOS machine using a Local Area Network and protocol converters. Such an arrangement may be difficult to set-up and maintain.

PREFERRED ELEMENTS

Preferred elements are those that enhance, but are not fundamental to, the utility of the proposed data management system.

Provide results in various forms: Can the system provide output in various forms (listings, graphs, maps) for diverse audiences?

STORET, PSAMP, DAIS and SEDQUAL all tend to present data in tabular form. ODES has charting and mapping capabilities; COMPAS mapping capabilities. DAIS has the further limitation that it can only report data for a single sampling survey or a single dredge year - it cannot combine data across data sets.

SEDQUAL has particularly powerful data grouping and data retrieval capabilities.

Generate detailed and summary lists: Can the system store and report detailed information such as lists of benthic species found in a sample, individual bioassay results, or the measurement of each chemical in a sample? Can it store, calculate, or report summary information such as a benthic diversity index, a list of stations with lead concentrations over 20 ppm or mean amphipod mortality?

All systems except COMPAS have such capabilities for the data sets they store. COMPAS can do some basic statistics (count, average); special programs would need to be written to do more complex summaries.

Easy access for non-experts: When the user is communicating with the computer system, how "user-friendly" is it? How easy is it to determine what keys to type in order to get the desired result? Is help available?

All of the systems evaluated have some kinds of "Help" facilities, and all are menu-driven. STORET is still the most limited in user-friendly functions.

Mapping capabilities: Covered under Required Element: capacity to store and report geographic data, above.

Stores and reports data on:

- Dredging and Disposal Activities: Geographic data would include maps of sites that have been dredged or used for dredge disposal, regulatory information would include a record dredging activities

- Hydrologic Data: Monitoring information would include river and tributary flow, geographic data would include information on upstream/downstream relationships (e.g., river reaches)

- Changes in wetland resources over time: The system would have the ability to store geographic data about wetlands, and to maintain different "snap-shots" of wetlands at different times. Ideally the system would also have the capability to do geographic analysis - to overlay maps of wetlands at Time A and Time B, highlight and calculate the changes in acreage.

- Changes in nearshore habitats: The system would have the ability to store geographic data about nearshore habitats, and to maintain different "snap-shots" of wetlands at different times. Ideally the system would also have the capability to do geographic analysis - to overlay maps of wetlands at Time A and Time B, highlight and calculate the changes in acreage.

STORET stores hydrologic data, based on the EPA River Reach System. A user could compare changes in wetland resources over time by physically overlaying ODES wetland maps; no system evaluated can calculate changes in acreage or highlight areas of lost wetlands.

Monitoring data can be updated: Does the system have procedures for adding new information? Are organizational arrangements in place to get new data entered in a timely fashion from various agencies doing monitoring.

Many Federal and state agencies that do environmental monitoring have agreed to put certain portions of their data into the STORET system. STORET is certainly the first or second largest monitoring

database in the U.S. Data from EPA 301(h) program dischargers and some data from National Estuary Programs are regularly entered into ODES; unfortunately there are no 301(h) permittees in Washington or Oregon.

Staff at the Washington Department of Ecology Sediment Management Unit use SEDQUAL to manage sediment data from the Puget Sound area; Ecology, Fisheries and the Puget Sound Water Quality Authority use PSAMP to store PSAMP monitoring data.

There really are no assurances that database systems will contain the data needed for Bi-State Program analyses, unless the data are entered or transferred by the Bi-State Program itself.

Contains most recent resource data: Does the system have procedures for adding new geographic information? Are organizational arrangements in place to get new data entered in a timely fashion from various agencies collecting the resource information?

There are no ongoing organizational arrangements for the integration of geographic data in the medium-term. Both Washington State and Oregon have made attempts to encourage the development of integrated GIS databases for state agencies; one such attempt was made in updating the Puget Sound Environmental Atlas. There are ongoing problems with differences in scale, in selected base maps, in calibrating elevations, in data definitions. Again, if the Bi-State Program needs to analyze different types of resource information together, the Bi-State Program will have to compile and integrate these data.

TECHNICAL ELEMENTS

The following technical factors were also considered in the evaluation of databases for the Bi-State Program:

Written in state-of-the art software: Both Oregon and Washington use a variety of commercially available database packages to manage their environmental data. Currently, both states are using Oracle database management system as their preferred database software for large projects. Oracle is a "4th generation" computer language, so-called for its ability to translate simple user commands or menu

choices into complex procedures. Oracle is also Arc/Info compatible; one can purchase add-on products to Arc/Info that can take data directly from Oracle without reformatting.

Other commercially available software that is 4th generation includes microcomputer packages like dBase, FoxPro, and Paradox. These types of tools enable more rapid and efficient system development, and provide greater flexibility in system uses. Database software is currently moving towards a greater ability to exchange data automatically among 4th generation languages. Using such software for the Bi-State program in the medium term may later provide the best long-term options.

Adequate performance: Does the system have adequate capacity to handle data expected to be in the Bi-State Program's database? Does it perform efficiently, even with large quantities of information (i.e., the user does not have to sit and wait too long for output).

As main-frame systems STORET and ODES command substantial processing power. SEDQUAL, DAIS and PSAMP run well on powerful microcomputers (a 486 machine is recommended for PSAMP), but PSAMP data checking programs, calculations and reporting programs can run slowly on large datasets.

The COMPAS database was originally designed to manage data from large geographic areas. No tests have yet been done to determine performance when much finer-size grid cells are used, and much more detailed geographic information is entered.

Data model: The design of the files for storing data is logical, appropriate and provides sufficient flexibility for adding new data types or modifying what information is currently stored. In technical terms, a relational rather than flat-file approach is used.

Both STORET and ODES suffer from the use of out-dated flat-file technology. SEDQUAL and PSAMP have particularly robust data models.

Necessary functions:

- Allow entry and editing of data

- Automated data checking for: valid ranges, codes, formats, required information.
- Transfer data by reading certain standard formats, and writing data in these formats.
- Produce reports
- Provide analysis capabilities such as simple statistics, comparison of data to standards.
- Provide flexible data selection routines, allowing exact specification of the desired output.

Control: Who manages and controls the database system, and access to the system? How easy or difficult will it be to make Bi-State Program concerns known to system manager?

Accessibility: If an authorized user wants to use the system, how easy is it to dial up, start operation, or otherwise access the system? How easy is it to get output from the system on the screen, in a file or on the printer? How easy is it to get authorization to access the system?

STORET and ODES, as main-frame systems are accessible to anyone with an authorized user ID. Sometimes, these can be difficult or time-consuming to obtain. In addition, in order to get printed reports, output must be down-loaded in a file, then sent to a printer on the user's computer.

To the user sitting directly at the microcomputer with PSAMP, COMPAS, SEDQUAL or DAIS, the system seems eminently accessible. However, for those without a copy of the database, or not on a network, the system is not accessible at all. PSAMP is not designed to work on a network.

SEDQUAL, DAIS and PSAMP all have a system of security, involving user IDs. This controls who can see data, and who can change data. However, security may be a problem for a system on an agency network, since it is difficult to prevent data files from being copied to other computers.