DESIGNING AN INTEGRATED, ACCESSIBLE INFORMATION MANAGEMENT SYSTEM FOR THE CHESAPEAKE BAY REGION

A Report of the Chesapeake Bay Program Data Center Workgroup's
Chesapeake Information Management System Workshop:
United States Geological Survey
Reston, Virginia — June 19 & 20, 1996

Submitted to:

Chesapeake Bay Program Office
U.S. Environmental Protection Agency
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Submitted:

Under Contract 68-C4-0072
Work Assignment EC-1-8

September 30, 1996

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

EXECUTIVE SUMMARY

INTRODUCTION

A workshop to help design an integrated, accessible information management system for the Chesapeake Bay Program was held on June 19 and 20, 1996, at the United States Geological Survey headquarters in Reston, Virginia. The workshop was sponsored by the Data Center Workgroup of the Chesapeake Bay Program. Over 80 individuals from Federal, State and local government agencies, academic institutions, nonprofit organizations, citizens groups, and Chesapeake Bay Program consultants and contractors participated in the workshop. This report provides background information on the process leading up to the workshop, discusses proceedings of the workshop itself, and presents recommendations developed from the workshop for implementing a Chesapeake Information Management System.

The workshop was part of a strategic planning initiative that was launched by the Data Center Workgroup in January 1996 with the intent of accomplishing the following three primary goals:

- Evaluate the current approach to data and information management within the Chesapeake Bay Program.
- Define the data and information needs of the diverse array of partners and participants in the Chesapeake Bay Program.
- Develop recommendations and an implementation approach for an improved Chesapeake Information Management System.

A variety of investigations were conducted to achieve the goals of the strategic planning initiative, including small-group and one-on-one investigations, and a thorough review of all existing background materials and documentation. These investigations culminated in the *Chesapeake Bay Program Information Management Requirements and Recommendations* report (Chesapeake Bay Program, 1996).

The two-day Chesapeake Information Management System Workshop was structured to facilitate the refinement of recommendations made in the report, and the creation of new recommendations where needed. The workshop consisted of plenary sessions where participants gave presentations and debated topics in an open forum, and breakout sessions where smaller groups of participants discussed specific topics with the goal of drafting recommendations. Concurrent breakout sessions in the following four topical areas were conducted:

- **User Interface for the Chesapeake Information Management System**: Breakout group participants discussed appropriate interface technology options (e.g., browser technology) and evaluated sample user interface screens. Screen selection options, as well as screen layout and features, were also discussed. The basis for a prototype was developed.

- **Information Management System Functions:** This group focused on the approaches that must be used, and/or actions that must be performed on data, to achieve the functions specified in the user interface. This section addressed some of the following questions: What must the information system accomplish? What functions must it perform? What approaches/actions must be used to perform these functions?

- **Accessing and Sharing Data:** Breakout group participants focused on determining options/best ways (e.g., establishing standards) for the wide range of Chesapeake Bay Program data users to access and share data and communicate effectively. Participants also addressed data and information input and output designs.

- **Data Base Design:** Breakout group participants determined which data base design model is the most appropriate design to effectively achieve the goals of the Chesapeake Information Management System. The models under consideration were hierarchical, network, relational, object, object-relational, and multidimensional. The breakout group also determined the appropriate architecture for the database management system (i.e., client server, distributed, or centralized).

The remainder of this Executive Summary first presents the overall conclusions and recommendations reached at the workshop, and then those that were specific to each of the four breakout groups described above. The body of the report presents background information for the Data Center Workgroup's strategic planning

initiative and research efforts leading to the workshop, and then presents the proceedings and findings of the workshop.

MAJOR CONCLUSIONS AND RECOMMENDATIONS

Participants in the Chesapeake Information Management System Workshop reached a number of conclusions and made a number of recommendations during the two days of deliberations. Conclusions and recommendations developed in the smaller breakout group sessions were all discussed in the plenary sessions to ensure that a majority of workshop participants were in agreement.

The conclusions reached and recommendations made during the workshop are presented below. Presented first are the overall conclusions and recommendations, which are so classified because they were generated either in a plenary session, or by more than one breakout group. Presented next are conclusions and recommendations that were generated in specific breakout groups regarding specific issues.

Overall Conclusions and Recommendations

Conclusions

- The current system of information and data management at the Chesapeake Bay Program is in need of improvement and modification to better meet the needs of the Bay Program partners and to fulfill the Bay Program's changing directions in research and communications. Some of the following problems typify concerns with the current approach to information management:

  - Many users or potential users find it difficult or impossible to access the Chesapeake Bay Program data they need;
  - Documentation of data is often insufficient and, therefore, often increases the amount of time needed to perform even routine procedures;
  - The desired types of data are not available and/or up-to-date;
  - Available data are often too geographically broad;
  - Environmental indicators, an important priority for the Chesapeake Bay Program, are difficult and inefficient to produce with the current system;

- There is an important unmet need for integration of different types of data in order to undertake complex analyses;
- There are important changing directions in Chesapeake Bay Program investigations that will require new types of data and new ways of managing data; and
- There is difficulty summarizing and analyzing data spatially (e.g., loadings related to sources in a specific geographic location).

Recommendations

- **Promote an Executive Council Directive on data and information management**. An Executive Council Directive is necessary due to the importance and difficulty of establishing uniform standards among a group of entities as large and diverse as the Chesapeake Bay Program. In essence, the success of the Data Center Workgroup's strategic planning effort is dependent on convincing everyone involved in the Chesapeake Bay Program to use software, hardware, and data procedures and protocols that are uniform enough so that all data providers and users can effectively access and share all appropriate data.

- **All Chesapeake Bay Program data must be accompanied by standardized metadata.** Metadata, or information about data (e.g., where the data came from, what format is it in) is critical to creating data menus and catalogs for the purpose of automatic computer searches. Also, researchers need metadata in order to evaluate one data set against another. Metadata standards should be established by the Chesapeake Bay Program as soon as possible, adherence to them should be required of all researchers funded by the Chesapeake Bay Program, and the standards should be promoted among all other data providers as actively as possible.

- **Reach an agreement among data providers on developing data standards.** Standards should be sought on all aspects of the system: acquisition, processing, documentation, reporting formats, data base standards, data dictionary, and data set maintenance and update procedures. The broad outline of this agreement could be specified in an Executive Council Directive.

- **As much Chesapeake Bay Program information and data as is feasible should be available over the Internet.** An ever increasing number of Chesapeake Bay Program data users have access to computers with Internet capabilities. This would be the single easiest step to take to provide rapid data access to a large number of data users that currently lack rapid or easy access.

- **The overall Chesapeake Information Management System should be appropriate for the needs of a diverse user group and diverse data formats.** While Internet access will be the preferred method for many users, some Bay Program partners do not have the necessary technology or knowledge to use the Internet, and some data types (e.g., very large data sets) are difficult to transfer over the Internet. A variety of other ways to access and share data and information (e.g., through telephone requests, on floppy disks, at ftp sites, highly summarized information on paper printouts, etc.) should also be built into the Chesapeake Information Management System.

- **The Chesapeake Information Management System must allow users to query information that is both spatial/geographic and textual.** There is a high demand for electronic maps of various aspects of the Chesapeake Bay and its surrounding watershed and airshed. Many Chesapeake Bay Program data users and workshop participants voiced strong support for the establishment of mechanisms that would allow them to query information from the Chesapeake Information Management System by specifying geographic locations, perhaps by "pointing and clicking" on a part of an electronic map, and then selecting the desired information from a pick list. Similarly, users would like text search capabilities whereby key words could be entered onto a search screen and all relevant records automatically retrieved for review.

- **It is imperative that the Chesapeake Information Management System be easy to use for the full-range of Chesapeake Bay Program data users.** Some Chesapeake Bay Program data users are scientists with sophisticated knowledge of the subject matter. Others may be volunteers in a citizen group who have limited expertise. Both types of users and everyone in between should be able to access and use the information and data they need, with the least difficulty feasible.

- **The Chesapeake Information Management System should be a hybrid system that is part centralized, part decentralized.** It is appropriate for some data to be stored and managed by the Chesapeake Bay Program Data Center in Annapolis, but some other data and information should be stored and managed by the many other Bay Program partner agencies and stakeholders distributed throughout the watershed. All Chesapeake Bay Program data and information should be directly and easily accessible through the Chesapeake Bay Program Data Center (e.g., through links in the Chesapeake Bay Program homepage on the Internet), but not all has to be physically present at, or managed by, the Chesapeake Bay Program Data Center.

Breakout Group-Specific Conclusions and Recommendations

User Interface

The Interface for the Chesapeake Information Management System Breakout Group came to consensus on the following conclusions and recommendations:

System Design Characteristics

- The interface should provide for platform independence (i.e., the user should be unaware if their local platform and the Chesapeake Information Management System platform are different).
- The data base should be based on standards, allowing it to be readily transferrable from one software package to another, should the Chesapeake Bay Program decide that this is necessary.
- The interface should support multiple browsers simultaneously.

Visual Design Characteristics

- The interface should be easy to read. Special attention should be paid to the selection of fonts, colors, and text height.
- The interface should be aesthetically pleasing. Uniformity in screen layout, dates, etc. would be the first step in achieving this.
- The interface should be well organized to allow users to easily and quickly gain access to the information they are looking for.

Access Characteristics

- The general public should have access only to appropriate data (e.g., fully documented, quality assured, non-copyrighted data).
- The Chesapeake Bay Program partner agencies and institutions should have access to a wider range of data (e.g., primary data) than the general public (with the goal to eventually make most of these data available to the general public once data are finalized).

- Once the results of data analyses are made available through the Chesapeake Information Management System, the supporting, or primary, data should also be made available.

Navigation Characteristics

- Graphical menus as well as text menus should be used in the user interface where appropriate.
- Users should have the ability to locate exactly where they are within the interface (i.e., direct access to a "roadmap" on the interface screen) at all times.
- The interface should provide direct links to other internal and external sites of interest to the user.

Output Characteristics

- Users should have a limited number of standard formats for data output available to them.
- Users should be able to output data to an on-line/interactive view.
- Users should be able to output data to a hardcopy format.
- Users should be able to output data to an e-mail message.
- Users should be able to output data to an ftp transfer.
- Users should be able to output data to a CD-ROM.
- Users should be able to submit a batch job to obtain the return output (e.g., e-mail, ftp, online).

Implementation

- The interface should maximize access to data with minimal investment.
- The implementation of the interface should be phased, with higher priority functions being implemented first.

Information Management System Functions

The Information Management System Functions Breakout Group reached the following conclusions and recommendations:

Metadata

- Metadata will be provided for primary data, Chesapeake Bay Program data, analyzed data and environmental indicators. Metadata will not be provided for raw data, which is not normally submitted to the Chesapeake Bay Program by data collection agencies.

- Metadata should include elements from the data dictionary but should be broader in scope and also contain the following:

  - The data source;
  - Whether the data are measured or estimated;
  - Any assumptions made in generating the data;
  - The detection limits, instruments, and methods used in collecting the data;
  - The purpose for which the data were collected;
  - Any caveats, comments, or qualifications that should be taken into consideration in using the data;
  - Elements from the data dictionary (e.g. format, field names);
  - Specific definition of the pollutant type (e.g. NOx particulate);
  - The spatial and temporal extent of the data; and
  - The revision history of the data, if they have been revised.

- Metadata should be directly linked to the data so that both are automatically downloaded in response to data requests.

- The level of detail of metadata should correspond to the level of data that it documents.

- Metadata standards need to be established for the Chesapeake Information Management System.

Primary Data

- Some Chesapeake Bay Program primary data will reside at the Chesapeake Bay Program Data Center in Annapolis.

- Other primary Chesapeake Bay Program data to be accessible through the Chesapeake Information Management System will be maintained by the data source or data owner.

- The Chesapeake Information Management System will provide links for downloading of data housed outside of the Chesapeake Bay Program Data Center.

- The Chesapeake Information Management System will provide communications links to data providers maintaining their own data.

- The Chesapeake Information Management System will maintain only contact information for those data housed outside of the Chesapeake Bay Program Data Center.

Accessing and Retrieving Data Spatially

- The Chesapeake Information Management System should include a simple method of retrieving data spatially (e.g., put boundary coordinates on metadata and query metadata).

- No on-line GIS engine need be provided through the Chesapeake Information Management System for external users. However, the Chesapeake Bay Program partners will provide the products of in-house geographic analyses on-line.

Data Library

- The on-line data library will contain the following kinds of information:

  - Unmodified copies of primary data in their native format;
  - Links and/or pointers to agencies that wish to provide data themselves;
  - A list of Chesapeake Bay Program reports and relevant grey literature, and copies of frequently requested reports; and
  - Selected tools, such as QA/QC software and analysis algorithms used by Chesapeake Bay Program data analysts.

- The Chesapeake Bay Program should consider developing a search engine that can identify relevant reports or portions thereof. This would require indexing all documents to be searched.

Linking Stressor and Response Data

- Common variables must be created in data sets to enable links to be established.

Data Dictionary

- A data dictionary is needed. The Chesapeake Bay Program's 1993 data dictionary may still be useful but will need to be updated. The EPA Great Lakes National Program Office's data dictionary (which built from the Chesapeake Bay Program's 1993 data dictionary) presented at the conference may be better, but it will need to be expanded to accommodate all the Bay Program needs.

Data Librarian

- A data librarian will always be needed to assure that metadata are created and maintained, to maintain links to other data sources, to assist users in identifying data sets if the on-line assistance is insufficient, and to maintain institutional memory of data holdings and revisions.

Facilitation of Univariate and Multivariate Analyses

- The highest priority for facilitating univariate and multivariate analyses is to provide quality assured, well-documented data sets. Data format standards need to be established (e.g., designate missing values differently from zero values).

- The Chesapeake Information Management System will not provide statistical packages but will provide results of analyses completed by Bay Program partners and the analytical algorithms used to obtain these results. The Chesapeake Bay Program may also provide guidance on useful statistical packages through the Chesapeake Information Management System.

Model Output

- The Chesapeake Information Management System should provide the following model-related information:

  - Summarized model output in graphical and tabular forms;
  - An online index to final model scenario runs;
  - Detailed model results on request (on magneto-optical media);

- Brief descriptions of each of the Chesapeake Bay models and how they are related;
- A list of model documentation and how to obtain it; and
- Links to other model sites (with their consent).

Accessing and Sharing Data

The Data Access and Sharing Breakout Group reached the following conclusions and recommendations:

Data Access Techniques

- The following data formats/mediums should have a place in a Chesapeake Information Management System, and are listed in order of preference:

  - World Wide Web server and file transfer protocol (ftp) server;
  - e-mail and e-mail distribution lists;
  - CD-ROM;
  - Magnetic media; and
  - Paper printouts.

- Webmasters at different Chesapeake Bay Program partner organizations should work together to ensure the same or a similar look and "feel" to all Web pages.

- The system must include access to "grey literature" (i.e., in-house technical reports—not peer reviewed journal publications—from all Chesapeake Bay Program participants).

- The Chesapeake Information Management System should make it easy to find and access primary data associated with a report or highly summarized data.

- The data access solution must accommodate "fire walls" (barriers to data access built into Web site-accessible networks).

Data Format

- Tabular data should be in delimited ASCII format, as it is the most common and easiest to share format. All data should be documented precisely.

- Document formats should be Hyper Text Markup Language (HTML) for small documents that will be viewed on-screen; and Adobe Portable Document Format (PDF) for large documents or documents that will be printed.

- Chesapeake Bay Program standard formats for GIS and graphical data must be developed. The Bay Program's GIS data is in ArcInfo export file format (.E00), but there are other GIS program types (Intergraph, Atlas, and MapInfo). Each of these programs has its own file format best suited to handle its own data.

Metadata

- Standardized metadata is important, with content appropriate to data level. The recommended format for metadata is regular ASCII if bundled with the data, and in a searchable format for database queries.

- The Chesapeake Bay Program should develop and distribute metadata tools (to minimize expenses of metadata).

- The expense of documenting data may be offset by a reduction in duplicate research; however, it should be acknowledged that generating metadata is expensive.

Data Librarian

- A data librarian(s) will always be needed. The primary role of the data librarian(s) should be to fill data requests. A secondary role should be to perform minor data analysis to help outside entities (e.g., local governments) answer multidisciplinary questions. The data library must develop and maintain continuity of institutional knowledge by ensuring that valuable knowledge is not lost when a data librarian leaves the Chesapeake Bay Program.

Data Dictionary

- A data dictionary is needed. The Chesapeake Bay Program's 1993 data dictionary may still be useful but will need to be updated. The EPA Great Lakes National Program Office's data dictionary (which built from the Chesapeake Bay Program's 1993 data dictionary) presented at the conference may be better, but it will need to be expanded to accommodate all the Bay Program needs.

General

- It is critical that creation of a Chesapeake Information Management System prepare for and, to the extent feasible, overcome problems associated with data ownership.

- Both distributed and centralized data management formats have a purpose. Neither is appropriate all the time.

- The National Spatial Data Infrastructure (NSDI) should be investigated for applicability to Chesapeake Bay Program needs.

- The Government Information Locator (GIL) may also be useful to the Chesapeake Bay Program, and should be investigated.

Data Base Design

The Data Base Design Breakout Group reached the following conclusions and recommendations:

Data Base Model

- A relational data base should be established initially.

- Over time, the Chesapeake Bay Program should migrate towards an object-relational data base that will allow the storage of more complex data types.

Data Base Architecture

- Initially, the data base should function as a central repository of data, eventually moving towards a more distributed data base.

- The Chesapeake Bay Program Data Center should function as an archival center.

- A mirror/replicated system should be implemented which would make copies of distributed data available through the Chesapeake Bay Program Data Center, as well as through the data source.

- System capabilities and implementation should be demonstrated with the use of prototypes.

Data Requirements

- The Chesapeake Bay Program's data dictionary should be expanded/updated.
- Common data standards should be developed.
- Common tools to facilitate data sharing should be developed.
- The history of data (e.g., originating agency/individual) should always be included.
- Lessons learned from other efforts (e.g., STORET Modernization, NWIS) should be considered

Data Standards Implementation

- Subject experts should develop the subject-specific standards and key fields.
- Chesapeake Information Management System standards should be created from the subject-specific standards.
- Minimum requirements for data should be established.
- Standards should be kept basic and simple (i.e., don't try to specify everything).

INTRODUCTION

Managing the Chesapeake Bay Program's extensive, diverse, and diffuse information[1] resources is a challenging task. Each year, Chesapeake Bay Program partners generate and use enormous quantities of information to accomplish a wide range of goals, including:

- Identifying and targeting priorities;
- Tracking implementation progress;
- Evaluating the health of the Chesapeake Bay ecosystem; and
- Educating decision-makers and the general public.

An elaborate environmental monitoring, scientific research, and information gathering network provides the foundation for developing and implementing Chesapeake Bay Program restoration and protection goals. The resultant data and information are managed in the following three principal ways:

- Much of the data are collected and maintained by the Chesapeake Bay Program Data Center. These data are actively solicited, reviewed, and often reformatted for storage in Chesapeake Bay Program Data Center computer systems.

- Other data are maintained by the agency or group from which the data originated; occasionally these data are integrated with the overall Chesapeake Bay information management framework. This data management category comprises two primary types of data providers: (1) participants directly involved with the Chesapeake Bay Program (e.g., state academic institutions and state agencies), and (2) participants who generate data for other purposes that are also useful to the Chesapeake Bay Program (e.g., comprehensive Federal agency data collection efforts such as the Toxics Release Inventory).

[1]The word information, as used in this report, encompasses all types of data generated and used by the Chesapeake Bay Program (ranging from primary data submitted "as is" to the Program by various data providers, to highly reviewed and summarized data that are prepared for reports). *The Chesapeake Bay Program Information Management Requirements and Recommendations* report (Chesapeake Bay Program, 1996) describes Chesapeake Bay Program data in the context of an information pyramid; all levels of the pyramid are encompassed in the word "information". Information management refers to the procedures and infrastructure used to handle these disparate data types.

- Another category of data are collected on an ad hoc, as needed basis, and are not maintained as part of an overall data management framework.

Keeping track of these vast information resources and making them available to a wide-range of users in an efficient, cost-effective, and user-friendly way is an important goal of the Chesapeake Bay Program, albeit a difficult one to achieve.

The Chesapeake Bay Program Data Center Workgroup was established to oversee the collection, maintenance, and dissemination of data useful to the Chesapeake Bay Program and its constituents. The Data Center Workgroup initiated a study in January 1996 to evaluate the current information management approach of the Chesapeake Bay Program and to make recommendations for the future. The Data Center Workgroup identified the following three goals for its strategic planning initiative:

1. Evaluate the current approach to data and information management at the Chesapeake Bay Program.
2. Define the data and information management needs of the diverse array of participants in the Chesapeake Bay Program.
3. Develop recommendations and an implementation approach for an improved Chesapeake Information Management System.

This report presents the findings and recommendations that emerged from the Chesapeake Information Management Workshop that was held on June 19 and 20, 1996 in order to address the issues presented above. The Data Center Workgroup invested considerable time during the first quarter of 1996 collecting and analyzing existing background materials, conducting one-on-one and small group inquiries, giving presentations to Subcommittees and Workgroups, and reviewing information management approaches at non-Chesapeake Bay Program agencies. The information

gathered from those efforts formed the basis for a preliminary report[2] that outlined the current Chesapeake Bay Program information management system, and presented recommendations for modifications to the current system. That report served as a starting point for the workshop, which focused on refining the preliminary recommendations made in the report, as well as developing new recommendations as needed.

The workshop included a representative group of Chesapeake Bay Program data users and providers and technical information management experts. Using a facilitated decision-making process, participants evaluated report recommendations in four different Breakout Groups (User Interface, System Functions, Accessing and Sharing Data, and Data Base Design) designed to represent the four major components of the overall information management system. The results of the workshop are presented in this report.

The remainder of this report is arranged in the following sections:

- Section II - Background to the Chesapeake Information Management System Workshop
- Section III - Breakout Session Summaries
- Section IV - Overall Conclusions and Recommendations
- Appendices.

[2]Chesapeake Bay Program. 1996. *Chesapeake Bay Program Information Management Requirements and Recommendations*. Report prepared for the Data Center Workgroup

BACKGROUND

IMPETUS FOR THE CHESAPEAKE INFORMATION MANAGEMENT SYSTEM WORKSHOP

The following two primary factors contributed to the Data Center Workgroup's decision to conduct a strategic planning initiative:

- Growing user frustration with the current system.
- Changing directions in Chesapeake Bay Program research, scientific, and management analyses.

Concerns With the Current System

There were a number of concerns with the current information management system that led the Chesapeake Bay Program's Data Center Workgroup to initiate its strategic planning effort. These issues were highlighted and expanded through the investigations, especially interviews, conducted to complete the strategic planning initiative. The concerns identified during those investigations are summarized below:

- **Many users or potential users find it difficult or impossible to access the Chesapeake Bay Program data they need.**
  - The current data management system is diffuse, with many different Chesapeake Bay Program participants obtaining and managing data for different purposes, in different formats, and on different platforms.
  - The disparity in the hardware and software used by different participants and the lack of commonly implemented data and documentation standards often makes the transfer of text and data files difficult.
  - There are no clearly defined procedures for obtaining information, and most Chesapeake Bay Program Office staff as well as non-Chesapeake Bay Program Office users are unable to access data directly.

- **Documentation of data is often insufficient and therefore often increases the amount of time needed to perform even routine procedures.**

  - Data coming into the Chesapeake Bay Program Data Center often lack documentation, especially on methods and data quality.

- **The desired types of data are not available and/or up-to-date.**

  - The data acquisition process is inefficient; data are subject to lengthy, iterative (i.e., many exchanges between data reviewers and providers who are often in different locations), QA/QC procedures before being loaded into the current Chesapeake Bay Program Data Center data management system.
  - Data requests are not effectively coordinated, resulting in duplication of effort.
  - Many data requests are ad hoc in nature, making it difficult to prioritize data management efforts.
  - Key types of data are not currently in the system (e.g., comprehensive tributary water quality, environmental threshold), either because they have not had high enough priority to be processed and documented, are currently not integrated or linked using Internet technology to the Chesapeake Bay Program Data Center, or because they are located in boxes and file cabinets in individual offices without others' knowledge.

- **Available data are often too geographically broad.**

  - Information is not readily available at local and sub-watershed levels. Tributary data are not effectively integrated into the current data management system.

- **Environmental indicators, an important priority for the Chesapeake Bay Program, are difficult and inefficient to produce with the current data management system.**

  - Obtaining the data to create environmental indicators is time consuming; often, individuals/organizations must be "tracked down" every time an indicator update is needed.
  - Standard procedures are not in place; data often arrive in a non-summarized format, requiring time-consuming analysis and/or conversion to the appropriate software platform.
  - Data often arrive without sufficient documentation to understand the methods employed to generate the original data or to conduct QA checks.

- **There is an important unmet need for integration of different types of data in order to undertake complex analyses.**
  - Many problems facing the Chesapeake Bay involve numerous elements, and the data representing these elements are often recorded and evaluated using different types of software and hardware. At present, it is often difficult or impossible to combine distinct data sets in order to answer complex questions, due to the incompatibility of the different software and hardware platforms involved. Thus, many important questions go unanswered.

These concerns were further illuminated at the workshop when participants described the current information management system as focusing the most resources (i.e., in terms of funding, staffing, and planning/meeting) at the base level of data collection and processing and the top level of communications, with relatively fewer resources being utilized for the in-between levels of data base construction, maintenance, QA/QC, documentation, interface design, and data base access. As a result, participants felt that the amount of data generated from the current Chesapeake Bay Program Information Pyramid was shortchanged and/or underdeveloped at the higher levels of indicators and analyzed data, because those types of highly summarized data rely on an easy to access and analyze data base containing high quality and reliable data with thorough QA/QC and documentation.

Changing Directions in Chesapeake Bay Program Investigations

As the Chesapeake Bay Program matures, it requires more and more elaborate and complex information and an analytical base to make informed management decisions and otherwise achieve Chesapeake Bay Program goals and objectives. One important goal for a proposed Chesapeake Information Management System is improved data integration. The Chesapeake Bay Program sponsored a Program Integration Survey in October 1995, which underscored the importance of program integration in the context of ecosystem management.

The ideas of program integration are embodied by several Chesapeake Bay Program activities, including the following:

- Two extensive geographical targeting initiatives—Phase II Habitat Restoration Program Integration and Targeting described in *Chesapeake Bay Habitat Restoration: A Framework for Action*, and the *Chesapeake Bay Chemical Contaminant Geographical Targeting Protocol*;
- Chesapeake Bay Basinwide Toxics Loading and Release Inventory; and
- Chesapeake Bay Environmental Indicators Program.

Each of these efforts share a common approach of acquiring, synthesizing, and analyzing significant amounts of disparate data from a wide variety of sources to answer specific questions and/or tell a story about the health of the Chesapeake Bay. These initiatives provide excellent examples of the integrated analysis the Chesapeake Bay Program desires to achieve with its information management system. Almost all will require new sources of data and/or methods of analysis, and these needs will have to be incorporated into the Chesapeake Information Management System.

ACTIONS LEADING TO THE CHESAPEAKE INFORMATION MANAGEMENT SYSTEM WORKSHOP

Evaluation of Existing Information

The Data Center Workgroup invested considerable time from January to May 1996 collecting and analyzing existing background materials, conducting one-on-one and small group inquiries, giving presentations to Subcommittees and Workgroups, and reviewing information management approaches at non-Chesapeake Bay Program agencies. The information gathered from those efforts formed the basis for the initial report, *Chesapeake Bay Program Information Management Requirements and Recommendations* (Chesapeake Bay Program, 1996), and also provided a framework of goals and tentative recommendations for the workshop.

The Data Center Workgroup completed several activities in order to acquire the background information needed to develop a comprehensive and effective data management system. The following activities were under taken to develop an understanding of current and future Chesapeake Bay Program data management:

- Evaluated existing materials,
- Evaluated current data management system,
- Conducted Subcommittee and Workgroup presentations/inquiries, and
- Conducted discussion sessions with user groups.

Small Group and Individual Program Investigations

The Data Center Workgroup made an effort to investigate a broad cross section of the Chesapeake Bay Program, ranging from program managers to scientists. Investigations followed the organization of the Chesapeake Bay Program and were conducted for Subcommittees and Workgroups involved with data management. Individual and small group discussions were conducted, and supplemented by mailing (electronic and regular) a questionnaire to other interested parties. Where possible, the Data Center Workgroup went beyond the immediate Chesapeake Bay Program to include other potential users of Chesapeake Bay Program data, including citizens, nongovernmental and not-for-profit organizations, and local governments.

The following Subcommittees and/or Workgroups of the Chesapeake Bay Program were studied in order to determine the Chesapeake Bay Program's program-specific data and information needs:

- Air Quality Coordination Group
- Communications Subcommittee
- Forestry Workgroup
- Land, Growth, and Stewardship Subcommittee
- Living Resources Subcommittee
- Modeling Subcommittee
- Monitoring Subcommittee

- Nutrients Subcommittee
- Toxics Subcommittee.

The Chesapeake Bay Environmental Indicators Program was also evaluated. In addition, analysis of cross-cutting data used by multiple programs (e.g., GIS, point sources, and historical data) was undertaken, and the needs of data users outside of the Chesapeake Bay Program were assessed.

Evaluation of other, comparable data management programs was undertaken, including:

- The Environmental Monitoring and Assessment Program (EMAP) Phase II
- The STORET (information STOrage and RETrieval) Modernization Effort
- The National Water Information System (NWIS) II
- The San Diego Bay Interagency Water Quality Panel Data Base
- The EPA Systems Development Center which maintains the Environmental Protection Agency (EPA) Spatial Data Library System (ESDLS) ENVIROFACTS (EF), and the Demographic Database (DEMOG).

A variety of questions guided these background investigations. The following broad topical areas were covered:

- Inventory of software, hardware, and telecommunications
- Inventory of products and/or output generated:
  - Impetus for data management efforts (e.g., mission statement, directive, work plan).
  - Common data requests and/or analyses.
  - Final products or output generated from Chesapeake Bay Program data/data management.
- Future activities:
  - Impetus for future activities.
  - Summary of future activities.

  - Data types and associated software, hardware, and telecommunication capabilities required to meet future activities.

- Data description:

  - Types and sources of data used.
  - Means of accessing these data.

- Data processing:

  - Format and processing steps for data that are received.
  - Format and processing steps for data that are generated.

- Vision or wish list for data management.

CHESAPEAKE INFORMATION MANAGEMENT SYSTEM GOALS AND OBJECTIVES

Several preliminary goals and objectives for a Chesapeake Information Management System were identified through the evaluation and planning process conducted for the Data Center Workgroup's strategic planning initiative. The following goals were identified:

- **Improve overall efficiency and cost-effectiveness of Chesapeake Bay Program data and information management.**

- **Improve data access.**

  - Better advertise available data.
  - Improve documentation and prepare detailed metadata.
  - Improve access of data to users of the Chesapeake Bay Program Data Center data base.
  - Allow data and information to be shared efficiently between all Chesapeake Bay Program partners.
  - Appeal to all user types.

- **Improve data availability.**

  - Provide access to different data types: raw, unprocessed, semi-processed (e.g., QA/QC performed), and highly summarized.

- Expand the amount and type of data that are available through the Chesapeake Information Management System. Expand availability and access to a broader array of data types (e.g., tributary data, subwatershed land use and population data).
- Provide access to large Federal data bases (e.g., TRI, PCS).
- Catalog, "clean up, and make useable" historical data.
- Ensure that the most current data are available and that all data are up-to-date.
- Organize available data geographically on a variety of geographic scales, from watershed-wide to local and subwatershed.

- **Improve data integration to enable more complex analyses**.

  - Develop a system that can access and integrate various types of data to address ecosystem linkages and perform complex geographical targeting. Priority data sets for integration include living resources, water quality, toxics, point sources, tributary loads, land use, and demographic information.
  - Provide ability to perform multiple functions against multiple data sets: browse/download/export; user-defined and standard queries; access summarized results.
  - Plan for cross-platform compatibility.
  - Provide a means to identify and access available data for a particular geographic region. Provide a geographic visual and query tool (e.g., clicking on a particular region in the watershed will provide an inventory of available data).

WORKSHOP APPROACH

Workshop Goals and Objectives

The primary goal of the workshop was to present the findings from the analysis of Chesapeake Bay Program data needs that took place during the months leading up to the workshop, and gauge whether the resulting recommendations were consistent with the interests of the full range of Chesapeake Bay Program data users. The workshop was therefore a kind of "reality check" on the direction the Data Center Workgroup was recommending the Chesapeake Bay Program begin to move in terms of data and information management. Four specific goals were set to guide the workshop:

1. Present preliminary recommendations for a basin-wide information management system to a diverse group of potential system users.

2. Solicit comments and suggestions on preliminary recommendations.

3. Refine preliminary recommendations and conduct reality checks against evaluation criteria.

4. Further refine recommendations for an information management system prototype.

Workshop Agenda

Each day of the workshop was a combination of plenary sessions and small breakout groups. The first day's activities included stating the workshop goals, presenting major findings from the background research leading up to the workshop, introducing the conceptual framework for the proposed Chesapeake Information Management System, and working in breakout groups to address the four primary questions of the workshop:

1. What kind of **user interface** should the Chesapeake Information Management System have?

2. What kind of **functions** should the data base perform?

3. What are the most efficient ways to ensure optimal **accessing and sharing** of Chesapeake Bay-related data?

4. What is the optimal **data base design** for the Chesapeake Information Management System?

Each breakout group was tasked with evaluating the proposed recommendations in their subject area, refining (or eliminating) those recommendations, and proposing new recommendations. Specifically, the breakout groups were asked to do the following six things:

- Assess the current situation.
- Describe the proposed approach.
- Discuss alternatives as needed.
- (Ideally) Reach consensus on an approach.
- Develop standards (minimally invasive; not over specified).
- Perform reality checks on the approach.

Breakout groups consisted of approximately 12-25 participants, and were supported by a team of two facilitators and two recorders. At the end of the day, each breakout group reported their proceedings at a plenary session for comment from other breakout groups in order to ensure cross-group compatibility of recommendations.

After the end of the first day, facilitators and recorders from each breakout group convened in order to clarify the conclusions and recommendations that emerged during the first day's deliberations. These were presented during the morning plenary session on the second day in order to frame the day's discussion and inform breakout groups of findings with cross-group importance. The breakout groups then re-convened to discuss the proposed conclusions and recommendations, and to conduct "reality checks" on whether the proposed approaches would satisfy a list of pre-formulated Chesapeake Information Management System needs. The workshop then concluded with a plenary discussion and summary of recommended Chesapeake Information Management System design features. A copy of the workshop agenda is contained in Appendix A.

Participants

Participants in the Chesapeake Information Management System workshop were drawn from the full range of Chesapeake Bay information users. There were 87 representatives from Federal, State and Local government agencies, including a large number of individuals from the Chesapeake Bay Program Data Center in

Annapolis, as well as representatives from non-profits, citizens groups, academic institutions, and corporations serving the Chesapeake Bay Program in a contractual role (Appendix B).

BREAKOUT SESSION SUMMARIES

The following pages contain summaries of the proceedings of the four breakout groups that met concurrently during the workshop.

USER INTERFACE BREAKOUT GROUP SESSION SUMMARY

Overview

The User Interface Breakout Group participants focused on design characteristics for the interface that will allow users to access the data and information contained within the Chesapeake Information Management system. The basis for the breakout group's deliberations were the findings contained in the Chesapeake Bay Program Data Center's report *Chesapeake Information Management Requirements and Recommendations* (Chesapeake Bay Program, 1996). With that report as a starting point, the breakout group set the following goals for the workshop:

- To define the system design characteristics of the user interface.
- To define the visual design characteristics of the user interface.
- To define the access characteristics of the user interface.
- To define the navigation characteristics of the user interface.
- To define the output characteristics of the user interface.

Highlights and Issues

Through discussions of a diverse range of opinions and ideas, the breakout group listed out preferred characteristics for the user interface.

- **System Design**. The breakout group recognized the importance of a well thought out and appropriately designed data base. It is important that the users be able to quickly and easily access the data that they are interested in. Ideally, the data base would integrate with existing technologies as well as with projected future capabilities. The breakout group also decided that the data

base should not be tied to specific software, as this could prove limiting in the future.

- **Visual Design**. The first aspect of the Chesapeake Information Management System that will be noticed by the users is the visual design of the interface. The breakout group stressed the necessity of having an aesthetically pleasing and well organized interface. The interface should be both visually appealing and easy to use.

- **Access**. The breakout group spent time discussing access to data through the interface. They realized that different levels of access would need to be set up: from full access to all data to restricted access to very specific data sets. It was not considered appropriate to open all data (e.g., partially quality assured data, copyrighted data) to the public. Access to different data could be opened on an as-needed basis. The interface will allow the users access to multiple levels of data, including:

  - Primary data: information submitted and exchanged "as is" by the data provider;
  - Chesapeake Bay Program data: information reviewed and processed into Chesapeake Bay Program standard specifications;
  - Analyzed data: summary results from data analyses; and
  - Indicators data: highly summarized data designed to tell the story about the health of the bay.

  It was noted that if higher level summarized data are made available, the data that went into these analyses (i.e., lower level data) should be made available to users as well.

- **Navigation**. The breakout group recognized that navigation, or the ability to move around within the user interface, is a key element of the user interface design. Many options were discussed by the group. Graphical menus would be useful for certain data and could be made available at multiple scales; for example, a graphical menu could include a map of the Chesapeake Bay with delineated regions that would allow users to indicate the geographical region for which they want data. A text menu, however, would allow faster access to certain data and allow users who are very familiar with the data and the interface to quickly jump to another location in the interface. The user's ability to determine their exact location within the Chesapeake Information Management System interface (i.e., a "roadmap" to the interface) is critical. It would also be beneficial for the user to have the ability to identify and locate related data and information.

- **Output**. The breakout group discussed a number of ways that data could be output for the user. Multiple formats of output are considered important for the Chesapeake Information Management System; if users cannot work with the format in which they receive data, the data are useless. It is important that a wide variety of users, with a wide variety of skills, be able to use the data being provided by the Chesapeake Bay Program. Available formats must be limited to a reasonable number of standard options to limit the complexity of the system and to minimize the costs incurred.

- **Implementation**. Finally, the breakout group discussed the implementation of the interface. It is important that the development of the interface move forward quickly to help maintain the momentum of the Chesapeake Information Management System effort. The first phase of implementation will have to show a useable product with limited funding.

Conclusions, Recommendations and Thoughts for Future Discussion

The Interface Breakout group came to consensus on the following conclusions and recommendations:

System Design Characteristics

- The data base should be based on standards, allowing it to be transferrable from one software package to another, should the Chesapeake Bay Program decide that this is necessary.

- The interface should provide for platform independence (i.e., the user should be unaware if their local platform and the Chesapeake Information Management System platform are different).

- The interface should support multiple browsers simultaneously.

Visual Design characteristics

- The interface should be easy to read. Special attention should be paid to the selection of fonts, colors, and text height.

- The interface should be aesthetically pleasing. Uniformity in screen layout and dates would be the first step in achieving this.

- The interface should be well organized to allow users to easily and quickly gain access to the information they are looking for.

Access Characteristics

- The general public should have access only to appropriate data (e.g., quality assured data, non-copyrighted data).
- The participants in the Chesapeake Bay Program should have a higher degree of access to data than the public.
- Once data analyses are made available through Chesapeake Information Management System, the supporting, lower levels of data should also be made available.

Navigation Characteristics

- Graphical menus as well as text menus should be made available where appropriate.
- Users should have the ability to locate exactly where they are within the interface (i.e., a "roadmap") at all times.
- The interface should provide links to other internal and external sites of interest to the user.

Output Characteristics

- Users should be able to submit a batch job to obtain the return output (e.g., e-mail, ftp, online).
- Users should have a limited number of standard formats for data output available to them.
- Users should be able to output data to an on-line/interactive view.
- Users should be able to output data to a hardcopy format.
- Users should be able to output data to an e-mail message.

- Users should be able to output data to an ftp transfer.
- Users should be able to output data to a CD-ROM.

Implementation

- The interface should maximize access to data with minimal investment.
- The implementation of the interface should be phased, with prioritized functions being implemented first.

FUNCTIONS BREAKOUT GROUP SESSION SUMMARY

Overview

Participants in the Information Management System Functions Breakout Group focused on determining and prioritizing desired functions to be provided by the Chesapeake Information Management System in order to enable the wide range of Chesapeake Bay Program data users to access and download data and reports. Participants began by discussing the overall goals of the Chesapeake Bay Program with the intent of determining the functions necessary to meet these goals.

The basis for the Breakout Group's deliberations were the findings contained in the Chesapeake Bay Program Data Center Workgroup's report *Chesapeake Information Management Requirements and Recommendations* (Chesapeake Bay Program, 1996). With that report as a starting point, the Breakout Group set the following goals and objectives for the Workshop:

- A goal of Chesapeake Bay Program data collection effort is to answer the question "What is the status of the Bay." The Chesapeake Information Management System must be able to help answer this question.
- A goal of the Chesapeake Bay Program is to communicate information about the Bay to policy-makers, managers, researchers, scientists, local governments, and stakeholders. The Chesapeake Information Management System must facilitate communication with this broad range of users.

Objectives

- To provide access to Chesapeake Bay Program primary, analyzed, and indicators data and their associated metadata as the highest priority.
- To establish data documentation or metadata standards and begin implementing them immediately.
- Initially to only develop those functions necessary to fulfill Chesapeake Bay Program policies, directives, strategies, and plans. These should be implemented using a phased approach. Two high priority functions are:
  - The ability to access and retrieve data spatially, and
  - The development of an on-line data library that will guide the user to available data, metadata and tools.

The Breakout Group discussed a number of issues related to the functions that should be provided by the Chesapeake Information Management System. The following sections highlight some of the important discussions that took place, and provide a list of recommendations. All recommendations were reached through consensus.

Highlights and Issues

An overall recommendation of the group was that the Chesapeake Bay Program should be ambitious in identifying desired functions, but that a phased approach should be taken in implementing these functions. In this way, simpler functions can be made available in a timely manner, while sufficient time is provided to refine and test more complicated functions prior to their implementation. The Breakout Group also agreed that the primary function of the Chesapeake Information Management System is to help answer the question "What is the status of the Bay" as posed by a broad range of users and stakeholders, including local governments, managers, policy-makers, researchers and scientists. These users' needs differ tremendously but their interests can be grouped into three broad categories of information:

- Primary data (researchers, scientists),
- Regional or geographical data (stakeholders, local governments, managers), and
- Highly summarized and interpreted data, such as indicators (stakeholders, policy-makers).

The Breakout Group acknowledged that each of these levels of data requires some degree of quality assurance and quality control.

The Functions Breakout Group also discussed the following topics:

- **Primary Data**. The Breakout Group suggested that primary data (data as they are received from the data generator) should be provided by the Chesapeake Bay Program only if:
  - The data are well documented,
  - The data are in high demand, and
  - The data generator or source agency is unable to provide the data on a large scale.

The Chesapeake Bay Program should provide links for the download of available data and should provide communications links to those data generators who provide their data themselves. The Chesapeake Information Management System will provide only contact information for those data sets maintained by the data generators.

- **On-Line Library**. The breakout group agreed that the Chesapeake Information Management System should include an on-line library that will contain the following kinds of information:
  - Unmodified copies of primary data in their native format;
  - Links and/or pointers to agencies that wish to provide data themselves (it is possible to record the number of data requests received);
  - A list of Chesapeake Bay Program reports and relevant grey literature, and copies of frequently requested reports;
  - A table of contents or catalog of all of the information and tools available through the system;

- A search engine for grey literature and reports;
- Selected tools, such as QA/QC software and analysis algorithms used by Chesapeake Bay Program data analysts; and
- A browsing capability.

The Breakout Group acknowledged that the Chesapeake Bay Program must receive approval from data generators prior to providing their data on-line.

- **Distribution Technology**. The Breakout Group discussed currently available technology for the distribution of information.

  - Hardcopy. The group agreed that it will remain necessary to provide hard copies of Chesapeake Bay Program data and documentation to those users who cannot access the Chesapeake Information Management System electronically, but that it is not the responsibility of the Chesapeake Bay Program Data Center to provide hard copies of all Chesapeake Bay Program reports.

  - Magneto-optical media. The group agreed that the distribution of electronic data using tapes and floppy disks will remain necessary until all Chesapeake Bay Program participants have electronic access to the Chesapeake Information Management System. The group also discussed the viability of optical media (CD ROM) as a means of distribution, but agreed that this medium is only economically viable when done in high volume.

  - World Wide Web. The group identified the World Wide Web as the most promising means of distribution, but recognized that many local governments may not have access to the Web for several years. The Web is an effective means of distribution for short reports, links to data providers and the download of data files and tools. The group stressed that it is the responsibility of both parties to maintain Web links.

- **Facilitation of Univariate and Multivariate Analyses**. The Breakout Group stated that many software and analytical packages are currently available and that the primary focus of the Chesapeake Information Management System should be to make the data available, along with the results of standard "in-house" analyses. The group agreed that no on-line analytical capability should be provided at present, but that some tools may be made available for download, such as analysis algorithms, and the Chesapeake Bay Program may provide some guidance on other analytical tools. The group stated that the

Chesapeake Information Management System should provide data in a form that will facilitate analysis by the users.

- **Model Output**. The Breakout Group agreed that there is no need to provide general access to requesting model runs. The Chesapeake Information Management System should, however, provide the following model-related information:

  - Summarized model output in graphical and tabular forms;
  - An online index to final model scenario runs;
  - Detailed model results on request (on magneto-optical media);
  - Brief descriptions of each of the Chesapeake Bay models and how they are related;
  - A list of model documentation and how to obtain it; and
  - Links to other model sites (with their consent).

- **Accessing and Retrieving Data Spatially**. The Breakout Group recognized that the Chesapeake Information Management System must provide the ability to query data spatially; however, the group did not feel that it was the responsibility of the Chesapeake Bay Program to provide an on-line analytical tool, such as a GIS-engine. The group suggested that data be thoroughly documented and fitted into a coarse grid in order to help users to narrow their requests. Data that are not currently spatially located (using a tributary code, latitude and longitude coordinates, a segment code or a station code) should be provided in response to all queries. A query engine that operates on the metadata would be useful. A graphical or map interface should be provided for spatial queries. Standards must be established for locating and relating data spatially.

- **Simple Graphical Displays and Interpretation**. Providing summary information is an important function of the Chesapeake Information Management System. The system should provide standard, continuously updated graphics that are prepared by the Chesapeake Bay Program, not interactively by the user.

- **Linking Stressor and Response Data**. The Breakout Group recognized that it may be difficult to establish links with data that are maintained outside of the Chesapeake Bay Program and that there is a need to create common variables in data sets to enable links to be established. The group also agreed that it is easier to establish links between highly summarized data (e.g. environmental indicators) than between primary data. The group felt that it is the responsibility of the Chesapeake Bay Program to establish links between stressor and response data in order to meet several of its directives, and that information gained from these links must be made available to the public. Ideally, the Chesapeake Information System would allow the user to click on a

geographic region and access data on such varied topics as water quality, land use, point source loads and living resources.

- **Metadata**. The Breakout Group agreed that the establishment of documentation standards and the generation of metadata for all data levels except raw data should be a top priority. Information that should be contained in the metadata includes:

  - The data source;
  - Whether the data are measured or estimated;
  - Any assumptions made in generating the data;
  - The detection limits, instruments and methods used in collecting the data;
  - The purpose for which the data were collected;
  - Any caveats, comments or qualifications that should be taken into consideration in using the data;
  - Elements from the data dictionary (e.g. format, field names);
  - Specific definition of the pollutant type (e.g. NOx particulate);
  - The spatial and temporal extent of the data; and
  - The revision history of the data, if they have been revised.

  The level of detail of the metadata should correspond to the level of the data. The system should automatically provide metadata with requested data and should force updates to the metadata when changes are made to the data themselves.

Conclusions and Recommendations

The Information Management System Functions Breakout Group reached the following conclusions and recommendations:

Metadata

- Metadata will be provided for primary data, Chesapeake Bay Program data, analyzed data and environmental indicators. Metadata will not be provided for raw data.

- Metadata should include elements from the data dictionary but should be broader in scope and also contain the following:

  - The data source;
  - Whether the data are measured or estimated;
  - Any assumptions made in generating the data;
  - The detection limits, instruments and methods used in collecting the data;

  - The purpose for which the data were collected;
  - Any caveats, comments or qualifications that should be taken into consideration in using the data;
  - Elements from the data dictionary (e.g. format, field names);
  - Specific definition of the pollutant type (e.g. NOx particulate);
  - The spatial and temporal extent of the data; and
  - The revision history of the data, if they have been revised.

- Metadata should be linked to the data so that both are downloaded in response to data requests.

- The level of detail of metadata should correspond to the level of data that it documents.

- Metadata standards need to be established.

Implementation Priority: Highest priority. Metadata need to be created and made available to users in order to make the data themselves useful.

Primary Data

- Some primary data will reside at the Chesapeake Bay Program Data Center.

- Other primary data will be maintained by the data source or owner.

- The Chesapeake Information Management System will provide links for the downloading of data housed outside of the Chesapeake Bay Program Data Center.

- The Chesapeake Information Management System will provide communications links to data providers maintaining their own data.

- The Chesapeake Information Management System will maintain only contact information for those data housed outside of the Chesapeake Bay Program Data Center.

Implementation Priority: Highest priority. Access to the data must be provided before any other information management system functions can be used.

Accessing and Retrieving Data Spatially

- The Chesapeake Bay Program should develop a simple method of retrieving data spatially (e.g. put boundary coordinates on metadata and query metadata).

- No on-line GIS engine will be provided. The Chesapeake Bay Program will provide the products of in-house geographic analyses on-line.

Implementation Priority: Near-term goal. Interest in querying data spatially is high, but spatial data standards must be established and data sets must be spatially located before these queries can be made.

Data Library

- The on-line data library will contain the following kinds of information:
  - Unmodified copies of primary data in their native format,
  - Links and/or pointers to agencies that wish to provide data themselves,
  - A list of Chesapeake Bay Program reports and relevant grey literature, and copies of frequently requested reports, and
  - Selected tools, such as QA/QC software and analysis algorithms used by Chesapeake Bay Program data analysts.

- The Chesapeake Bay Program should consider developing a search engine that can identify relevant reports or portions thereof. This would require indexing all documents to be searched.

Implementation Priority: Near-term goal. A well structured data library will inform users of what data and tools are available and will facilitate use of the data base and should be made available as the Chesapeake Information Management System becomes operational.

Linking Stressor and Response Data

- Common variables must be created in data sets to enable links to be established.

Implementation Priority: Medium-term goal. This measure will be labor intensive but will allow more thorough exploration of the data sets.

Data Dictionary

- A data dictionary is needed. The Chesapeake Bay Program's 1993 data dictionary may still be useful but will need to be updated. The EPA Great Lakes National Program Office's data dictionary (which built from the Chesapeake Bay Program's 1993 data dictionary) presented at the conference may be better, but it will need to be expanded to accommodate all the Bay Program needs.

Implementation Priority: Medium-term goal. A data dictionary will compliment the metadata by providing an overview of data set characteristics.

Data Librarian

- A data librarian will always be needed to assure that metadata are created and maintained, to maintain links to other data sources, to assist users in identifying data sets if the on-line assistance is insufficient, and to maintain institutional memory of data holdings and revisions.

Implementation Priority: Long-term goal. This should be part of the Chesapeake Information Management System when it is fully operational.

Facilitation of Univariate and Multivariate Analyses

- The highest priority for facilitating univariate and multivariate analyses is to provide quality assured, well-documented data sets. Data format standards need to be established (e.g., designate missing values differently from zero values).

- The Chesapeake Bay Program will not provide statistical packages but will provide results of analyses completed in-house and the analytical algorithms used to obtain these results. The Chesapeake Bay Program may also provide guidance on useful statistical packages.

Implementation Priority: Long-term goal. This should be part of The Chesapeake Information Management System when it is fully operational.

Model Output

- The Chesapeake Information Management System should provide the following model-related information:

  - Summarized model output in graphical and tabular forms;
  - An online index to final model scenario runs;
  - Detailed model results on request (on magneto-optical media);
  - Brief descriptions of each of the Chesapeake Bay models and how they are related;
  - A list of model documentation and how to obtain it; and
  - Links to other model sites (with their consent).

Implementation Priority: Long-term goal. This should be part of The Chesapeake Information Management System when it is fully operational.

DATA ACCESS AND SHARING BREAKOUT GROUP SESSION SUMMARY

Overview

Participants in the Data Access and Sharing Breakout Group focused on determining options (e.g., establishing standards) for the wide range of Chesapeake Bay Program data users to access and share data and communicate effectively. Participants also addressed data and information input and output designs. The basis for the Breakout Group's deliberations were the findings contained in the Chesapeake Bay Program Data Center Workgroup's report *Chesapeake Information Management Requirements and Recommendations* (Chesapeake Bay Program, 1996). With that report as a starting point, the Breakout Group set the following data access and sharing goal and objectives as a guide for its deliberations.

Goal

- Make all Chesapeake Bay Program data available and accessible.

Objectives

- Improve access to data.
- Improve the shareability of data.
- Improve the finding of data.
- Improve the documenting of data (metadata).

The Breakout Group discussed a number of issues related to accessing and sharing Chesapeake Bay Program-related information. The following sections highlight some of the important discussions that took place, and provide recommendations. All recommendations were reached through consensus of the members of the breakout group.

Highlights and Issues

In order to communicate effectively, the Breakout Group found it useful to establish definitions of words that were open to multiple meanings. The box on this page lists definitions of terms agreed upon by the Data Access and Sharing Breakout Group.

DEFINITIONS AGREED UPON BY THE DATA ACCESS AND SHARING BREAKOUT GROUPS

- Chesapeake Bay Program data = all Chesapeake Bay Program information, including raw data.
- Chesapeake Bay Program participant = any entity or individual interested in obtaining Chesapeake Bay Program data.
- Data accessibility = the ability to obtain data (information) easily.
- Data shareability = the usability of accessible data to interested Chesapeake Bay Program participants. If data is accessible, but in a format that is not usable by the entity obtaining the data, it is not sharable.

The Data Access and Sharing Breakout Group also discussed the following topics:

- **Data ownership**. The Breakout Group agreed that one of the most difficult issues the Chesapeake Information Management System will have to face is data ownership. It is clear that not all participating Chesapeake Bay Program entities are equally enthusiastic about making the data they manage available to outsiders, even to other Chesapeake Bay Program participants. Similarly, there are sometimes individuals or departments/groups that do share data and those that do not share data within the same entity. The degree to which valuable data is withheld by one entity from other Chesapeake Bay Program participants will be an important factor in the success or failure of the Chesapeake Information Management System.

- **Data access and storage techniques**. The Breakout Group spent much of its time discussing ways to make Chesapeake Bay Program data easily accessible and in a storage format that would be compatible with those of a majority of Chesapeake Bay Program data users. The following data access options were discussed: a Local Area Network (LAN) file server, a database server, a terminal access bulletin board system, an Internet file transfer protocol server, and a World Wide Web server. The data storage formats discussed were: paper, read/write magnetic media, and CD-ROMs. All of these options have strengths and weaknesses relative to the specific needs of a Chesapeake

Information Management System and the resources available to fund new systems. After discussing all options, the group ranked them in order of preference. The only option completely ruled out was a terminal access bulletin board system, because participants felt that the technology would soon be obsolete. Some options were classified as potentially feasible only in the future.

- **File formats and sharing**. Also discussed in detail were potential file formats of Chesapeake Bay Program data and methods to facilitate sharing of data between distinct entities. File formats discussed included ASCII delimited text, Geographic Information System (GIS) data types such as ArcInfo export files (.E00 files), formats for metadata, Hypertext Markup Language (HTML), and Adobe Portable Document Format (PDF) format. As with data access and storage techniques, the group ranked all options in order of preference.

- **Data librarian**. The Chesapeake Bay Program has tasked various staff members for varying periods of time over the years with organizing and cataloging Chesapeake Bay Program data. The Chesapeake Bay Program is now considering creating a permanent position(s) for a data librarian in order to better manage Chesapeake Bay Program data, and to serve as a resource to help people locate specific Chesapeake Bay Program data. The breakout group discussed the range of possible roles for a data librarian(s), and set priorities. A major concern raised in the breakout group was that entities outside of the Chesapeake Bay Program Data Center, some with limited hardware and software capabilities, might not be able to access Chesapeake Bay Program data. A data librarian or team of librarians would be one way for entities with lower technological capabilities to acquire data, as requests could be placed over the telephone. There were differences of opinion over the optimal scope of the librarian(s)'s duties. Some felt that the librarian(s) should undertake analyses of data in order to answer multidisciplinary questions, while others felt the scope should be limited to providing only existing data, leaving analysis up to the requesting entity. The group reached consensus that the primary role for a data librarian(s) should be to fill data requests. A secondary role should be to perform minor data analysis to help outside entities (e.g., local governments) answer multidisciplinary questions.

Conclusions and Recommendations

The Data Access and Sharing Breakout Group reached the following conclusions and recommendations:

Data Access Techniques

- The following data formats/mediums should have a place in a Chesapeake Information Management System, and are listed in order of preference:

  - **World Wide Web server and file transfer protocol (ftp) server**. These should be the preferred methods for accessing and transferring small to medium-size data sets.

  - **E-Mail and E-Mail distribution lists**. E-mail is good for outgoing communications such as correspondence, small data sets, Metadata, and catalogs. Also useful for communication with the data librarian(s).

  - **CD-ROM**. CD-ROMs will be the easiest format to provide large published and ad hoc data, and are more common on personal computers than magnetic tape readers. They can hold large amounts of data in a variety of formats. They could be used for long-term periodic updates (e.g., annual). Applications/front end viewing tools or querying tools should be built-in.

  - **Magnetic media**. Magnetic media includes floppy disks and magnetic tape. Floppy disks are useful for providing small to medium-sized data sets, and most entities have computers with floppy disk drives. Magnetic tape readers are less common, but large data sets can be stored on magnetic tape.

  - **Paper printouts**. Paper printouts of data should be discouraged, as they cannot be easily manipulated by computers. However, paper can play a role in published reports (which should also be available through an Internet ftp server or Internet World Wide Web server) and for presentation of highly summarized data. Also, paper printouts may be the only option for individuals or organizations without computers or with computers that lack necessary software.

  - It was agreed that a terminal access bulletin board system relies on technology that will soon be obsolete, and should therefore not be pursued.

Implementation Priority: The options listed above should be established as soon as possible, with emphasis placed on options with a higher rank.

- Webmasters at different Chesapeake Bay Program participating entities should work together to ensure the same or a similar look and "feel" to all Web pages.

Implementation Priority: Immediate; Chesapeake Bay Program Web Page Team should begin consulting with appropriate Webmasters as soon as possible.

- The system must include access to "grey literature" (i.e. in-house technical reports (not peer reviewed) from all Chesapeake Bay Program participants).

Implementation Priority: Long-term goal; this should be part of the Chesapeake Information Management System when it is fully operational.

- The system should make it easy to find and access data associated with a report or highly summarized data.

Implementation Priority: Mid-term goal; reports and highly summarized data should be made available through Chesapeake Information Management System as it becomes operational.

- The data access solution must accommodate "fire walls" (barriers to data access within Web site-accessible networks).

Implementation Priority: Incorporate into Chesapeake Information Management System from the start, if feasible.

Data Format

- Tabular data should be in delimited ASCII format, as it is the most common and easiest to share format. All data should be documented precisely.

Implementation Priority: Medium-term goal; begin recommending delimited ASCII now, with the understanding that considerable time will pass before everyone is using it.

- Document formats should be Hyper Text Markup Language (HTML) for small documents that will be viewed on-screen; and Adobe Portable Document Format (PDF) for large documents or documents that will be printed.

Implementation Priority: Medium-term goal; begin recommending now, with the understanding that considerable time will pass before these formats are fully accepted.

- Chesapeake Bay Program standard formats for GIS and graphical data must be developed. The Bay Program's GIS data is in ArcInfo export file format (.E00), but there are other GIS program types (Intergraph, Atlas, and MapInfo). Each of these programs has its own file format best suited to handle its own data.

Implementation Priority: High priority; the sooner standards are developed, the less data will end up in incompatible formats.

Metadata

- Standardized metadata is important, with content appropriate to data level. The recommended format for metadata is regular ASCII if bundled with the data, and in a searchable format for database queries.

Implementation Priority: High; establishment of good metadata should be accomplished as soon as possible.

- The Bay Program should develop and distribute metadata tools (to minimize the expense of generating metadata).

Implementation Priority: Medium-term goal; focus on establishing good metadata within the Chesapeake Bay Program first.

- The expense of documenting data may be offset by a reduction in duplicate research; however, it should be acknowledged that metadata is expensive to generate.

Implementation Priority: (not applicable).

Data Librarian

- A data librarian(s) will always be needed. The primary role of the data librarian(s) should be to fill data requests. A secondary role should be to perform minor data analysis to help outside entities (e.g., local governments) answer multidisciplinary questions. The data library must develop and maintain continuity of institutional knowledge by ensuring that valuable knowledge is not lost when a data librarian leaves the Chesapeake Bay Program.

Implementation Priority: High priority to create a position(s) for a data librarian(s).

Data Dictionary

- A data dictionary is needed. The 1993 Chesapeake Bay Program version may still be useful. The EPA Great Lakes version (which drew from the Chesapeake

Bay Program version) presented at the conference may be better, but it will need to be expanded to accommodate all Program needs.

Implementation Priority: High priority. Necessary precursor to achievement of metadata recommendation.

General

- It is critical that creation of a Chesapeake Information Management System prepare for and, to the extent feasible, overcome problems associated with data ownership.

Implementation Priority: High priority; this issue should be addressed right from the start and throughout the process.

- Both distributed and centralized data management formats have a purpose. Neither is appropriate all the time.

Implementation Priority: (not applicable).

- The National Spatial Data Infrastructure (NSDI) should be investigated for applicability to Chesapeake Bay Program needs.

Implementation Priority: High priority for initial analysis, medium priority if initial analysis shows that NSDI has less than high value.

- The Government Information Locator (GIL) may also be useful to the Chesapeake Bay Program, and should be investigated.

Implementation Priority: High priority for initial analysis, medium priority if initial analysis shows that GIL has less than high value.

DATA BASE DESIGN BREAKOUT GROUP SESSION SUMMARY

Overview

The Data Base Design Breakout Group focused on how to design the data base behind the Chesapeake Information Management System. Hierarchical, network, relational, object, object-relational, and multidimensional models were considered by the breakout group. The Breakout Group was also concerned with choosing the

appropriate architecture for the database (e.g., distributed, centralized) to efficiently meet the needs of the Chesapeake Bay Program.

The basis for the Breakout Group's deliberations was the findings contained in the Chesapeake Bay Program Data Center Workgroup's report *Chesapeake Information Management Requirements and Recommendations* (Chesapeake Bay Program, 1996). With that report as a starting point, the Breakout Group established the following goals for the session:

- To determine the appropriate model for the Chesapeake Bay Program data base.
- To determine the appropriate architecture for the Chesapeake Bay Program data base.
- To determine the necessary data requirements for Chesapeake Bay Program data.

The Breakout Group discussed a number of issues related to data base design for the Chesapeake Information Management System. The following sections highlight some of the discussions that took place, and provide a list of recommendations. All recommendations were reached through consensus.

Highlights and Issues

The Data Base Design Breakout Group discussed a variety of options for the design of the data base. All options were evaluated based on speed, ability to handle complex and diverse data, cost, and compatibility/ability to handle legacy data. The following topics were discussed:

- **Data base model**. There were six different data base models evaluated for implementation by the Breakout Group: flat file, hierarchical, network, relational, object oriented, and object-relational models. Benefits and drawbacks of each model were discussed. Flat files are a useful means of

transporting data, as many software packages are capable of handling data in this format. Flat files, hierarchical, and network data bases however, each have limited flexibility for data management. Relational, object oriented, and object-relational data bases each provide increased flexibility in managing data, however, they require the user of the system to have access to very specific, and often more expensive, software. Relational data bases (e.g., INGRES, ORACLE, SYBASE, INFORMIX) provide efficient management of well defined data. Object oriented data bases provide the ability to access and store more complex data types (e.g., image and sound data). Object-relational data bases (e.g., UNISQL, ILLUSTRA) merge the best features of object oriented and relational data bases, but are often slower.

- **Data base architecture**. There were two different architectures discussed at length by the Breakout Group: centralized and distributed. In a centralized system, all data are located in one physical location, as opposed to a distributed system, where data are spread over multiple platforms and/or geographic locations. A centralized system would allow the Chesapeake Bay Program to maintain tighter control of the data and its quality assurance, while a distributed system would allow the Chesapeake Bay Program to hand over portions of their data management responsibilities to data providers. The drawbacks to centralized and distributed systems were also discussed. A centralized system would be resource intensive for the Chesapeake Bay Program, requiring full time staff trained to manage and maintain the data base. Implementation of a distributed data base raised a number of questions, including:

  - What if an individual/agency is incapable of maintaining their data?
  - What happens if an individual/agency "disappears"?
  - How will you correct and archive the data?

- **Data requirements**. The Breakout Group spent considerable time discussing data requirements. The ability to access and use the data contained in the data base is vital to the Chesapeake Information Management System effort. This requires data consistency and standardization. It also requires that compatible formats be used, and that sufficient information exists to allow users to locate and understand the available data.

- **Data standards implementation**. The Breakout Group discussed the importance of the standards that would have to be implemented to make the data base function properly. It was recognized that the submittal and retrieval of data from the data base must be a simple process; this requires simple standards that will not prohibit users from accessing the data base. This is especially true for a distributed system. It was also recognized that the standards should be developed through a consensus building process involving the affected parties.

Conclusions and Recommendations

The Data Base Design Breakout Group reached the following conclusions and recommendations:

Data Base Model

- A relational data base should be established initially.
- Over time, the Chesapeake Bay Program should migrate towards an object-relational data base that will allow the storage of more complex data types.

Data Base Architecture

- Initially, the data base should function as a central repository of data, eventually moving towards a more distributed data base.
- The Chesapeake Bay Program Data Center should function as an archival center.
- A mirror/replicated system should be implemented which would make copies of distributed data available through the Chesapeake Bay Program Data Center, as well as through the data source.
- System capabilities and implementation should be demonstrated with the use of prototypes.

Data Requirements

- The Chesapeake Bay Program's data dictionary should be expanded/updated.
- Common data standards should be developed.
- Common tools to facilitate data sharing should be developed.
- The history of data (e.g., originating agency/individual) should always be included.
- Lessons learned from other efforts (e.g., STORET Modernization, NWIS) should be considered.

Data Standards Implementation

- Subject experts should develop the subject-specific standards and key fields.
- Chesapeake Information Management System standards should be created from the subject- specific standards.
- Minimum requirements for data should be established.
- Standards should be kept basic and simple (i.e., don't try to specify everything).

OVERALL CONCLUSIONS AND RECOMMENDATIONS

As noted earlier in this report, the goal of the Chesapeake Information Management Workshop was essentially two-fold: 1) to *refine* the recommendations contained in the report of the research effort leading up to the workshop; and 2) to propose *new* recommendations based on ideas and facts that emerged for the first time at the workshop. Thus, the conclusions and recommendations presented in this report are a mosaic of old (often refined) and new, and should not be considered a product exclusively of the workshop. Because of the volume of recommendations in the original Chesapeake Information Management System report, much of the effort at the workshop was directed toward refining recommendations that were imperfect and proposing new recommendations, rather than laboriously discussing and ratifying recommendations that seemed adequate as originally stated. The major conclusions and recommendations from the workshop were as follows; details are provided in each individual breakout session summary.

CONCLUSIONS

- The current system of information and data management at the Chesapeake Bay Program is in need of improvement and modification to better meet the needs of the Bay Program partners and to fulfill the Bay Program's changing directions in research and communications. Some of the following problems typify concerns with the current approach to information management:

  - Many users or potential users find it difficult or impossible to access the Chesapeake Bay Program data they need;
  - Documentation of data is often insufficient and, therefore, often increases the amount of time needed to perform even routine procedures;
  - The desired types of data are not available and/or up-to-date;
  - Available data are often too geographically broad;
  - Environmental indicators, an important priority for the Chesapeake Bay Program, are difficult and inefficient to produce with the current system;
  - There is an important unmet need for integration of different types of data in order to undertake complex analyses;

- There are important changing directions in Chesapeake Bay Program investigations that will require new types of data and new ways of managing data; and
- There is difficulty summarizing and analyzing data spatially (e.g., loadings related to sources in a specific geographic location).

RECOMMENDATIONS

- **Promote an Executive Council Directive on data and information management**. An Executive Council Directive is necessary due to the importance and difficulty of establishing uniform standards among a group of entities as large and diverse as the Chesapeake Bay Program. In essence, the success of the Data Center Workgroup's strategic planning effort is dependent on convincing everyone involved in the Chesapeake Bay Program to use software, hardware, and data procedures and protocols that are uniform enough so that all data providers and users can effectively access and share all appropriate data.

- **All Chesapeake Bay Program data must be accompanied by standardized metadata**. Metadata, or information about data (e.g., where the data came from, what format is it in) is critical to creating data menus and catalogs for the purpose of automatic computer searches. Also, researchers need metadata in order to evaluate one data set against another. Metadata standards should be established by the Chesapeake Bay Program as soon as possible, adherence to them should be required of all researchers funded by the Chesapeake Bay Program, and the standards should be promoted among all other data providers as actively as possible.

- **Reach an agreement among data providers on developing data standards.** Standards should be sought on all aspects of the system: acquisition, processing, documentation, reporting formats, data base standards, data dictionary, and data set maintenance and update procedures. The broad outline of this agreement could be specified in an Executive Council Directive.

- **As much Chesapeake Bay Program information and data as is feasible should be available over the Internet**. An ever increasing number of Chesapeake Bay Program data users have access to computers with Internet capabilities. This would be the single easiest step to take to provide rapid data access to a large number of data users that currently lack rapid or easy access.

- **The overall Chesapeake Information Management System should be ppropriate for the needs of a diverse user group and diverse data formats**. While Internet access will be the preferred method for many users, some Bay

Program partners do not have the necessary technology or knowledge to use the Internet, and some data types (e.g., very large data sets) are difficult to transfer over the Internet. A variety of other ways to access and share data and information (e.g., through telephone requests, on floppy disks, at ftp sites, highly summarized information on paper printouts, etc.) should also be built into the Chesapeake Information Management System. The Chesapeake Information Management System should provide access to all but the raw data levels of data on the Chesapeake Information Management Pyramid (i.e., primary data, Chesapeake Bay Program data, analyzed data, indicators). Raw data are defined as original, unmodified field and laboratory results that are maintained by the data originators and typically not submitted to the Chesapeake Bay Program or others.

- **The Chesapeake Information Management System must allow users to query information that is both spatial/geographic and textual**. There is a high demand for electronic maps of various aspects of the Chesapeake Bay and its surrounding watershed and airshed. Many Chesapeake Bay Program data users and workshop participants voiced strong support for the establishment of mechanisms that would allow them to query information from the Chesapeake Information Management System by specifying geographic locations, perhaps by "pointing and clicking" on a part of an electronic map, and then selecting the desired information from a pick list. Similarly, users would like text search capabilities whereby key words could be entered onto a search screen and all relevant records automatically retrieved for review.

- **It is imperative that the Chesapeake Information Management System be easy to use for the full-range of Chesapeake Bay Program data users**. Some Chesapeake Bay Program data users are scientists with sophisticated knowledge of the subject matter. Others may be volunteers in a citizen group who have limited expertise. Both types of users and everyone in between should be able to access and use the information and data they need, with the least difficulty feasible.

- **The Chesapeake Information Management System should be a hybrid system that is partly centralized, partly decentralized**. It is appropriate for some data to be stored and managed by the Chesapeake Bay Program Data Center in Annapolis, but some other data and information should be stored and managed by the many other Bay Program partner agencies and stakeholders distributed throughout the watershed. All Chesapeake Bay Program data and information should be directly and easily accessible through the Chesapeake Bay Program Data Center (e.g., through links in the Chesapeake Bay Program Homepage on the Internet), but not all has to be physically present at, or managed by, the Chesapeake Bay Program Data Center.

REFERENCES

Chesapeake Bay Program, *Chesapeake Bay Program Information Management Requirements and Recommendations.* Chesapeake Bay Program, Data Center Workgroup, Annapolis, MD. 1996.

APPENDIX A

AGENDA

CHESAPEAKE INFORMATION MANAGEMENT SYSTEM WORKSHOP

Designing an Integrated, Accessible Information Management System for the Chesapeake Bay Region
United States Geological Survey: Reston, Virginia
June 19-20, 1996

Workshop Goals

- *Present preliminary recommendations for a basinwide information management system to a diverse group of potential system users.*
- *Solicit comments and suggestions on preliminary recommendations.*
- *Refine preliminary recommendations and conduct reality check against evaluation criteria.*
- *Further refine recommendations for an information management system prototype.*

WEDNESDAY, JUNE 19, 1996

8:00 am - 8:30 am: Registration (coffee and donuts)

Plenary Session (Conference Room)

8:30 am - 8:45 am: Welcome and Introductions *[Lowell Bahner, Data Center Manager, NOAA; Roger Barlow, US Geological Survey]*

8:45 am - 9:00 am: Purpose of Workshop *[Dr. Robert E. Magnien, Data Center Workgroup Chair, MD Department of Natural Resources]*

9:00 am - 9:45 am: Keynote Address - New Ideas in Environmental Information Management *[Dr. Andrew P. Sage, First American Bank Professor, School of Information Technology and Engineering, George Mason University]*

9:45 am - 10:00 am: Chesapeake Bay Program Information Management Options -- Outlook for the Future *[Dr. Craig Miller, Vice President and Chief Scientist, Science Applications International Corporation (SAIC)]*

10:00 am - 10:30 am: Presentation of Major Findings from Background Investigations and User Needs Survey *[Cary Gaunt, Project Manager, SAIC]*

10:30 am - 10:45 am: Break (coffee and donuts)

10:45 am - 11:45 am: Presentation of Recommendations and Conceptual Framework for Proposed Chesapeake Information Management System. Introduction to breakout group roles and responsibilities. *[Cary Gaunt, Project Manager, SAIC]*

Lunch: 12:00 pm - 1:00 pm (Conference Room) (Brown Bag Lunches Catered by USGS)

Speaker: New Technologies and Applications in Information Management *[Dr. John Helly, San Diego Bay Program and San Diego Supercomputing Center, University of California]*

Concurrent Breakout Groups (Breakout Rooms)

1:00 pm - 2:30 pm: Concurrent breakout groups

Goals of Each Breakout Session: The main function of the breakout groups is to review the current status and evaluate proposed recommendations in their subject areas. Using a facilitated process, each breakout group will reach consensus on final information management design recommendations to be presented to the entire group in the concluding Plenary Session. Each of the four breakout groups will discuss the following:

- *Assess current situation*
- *Describe proposed approach*
- *Discuss alternatives as needed*
- *(Ideally) Reach consensus on approach*
- *Develop standards (minimally invasive; not over specified)*
- *Perform reality check on approach*

Attachment A provides an overview of each breakout session.

2:30 pm - 2:45 pm: Break (Coffee and cold drinks)

Breakout Groups Reconvene (Breakout Rooms)

2:45 pm - 5:00 pm: Concurrent breakout groups

1. *User Interface for Chesapeake Information Management System*
2. *Information Management System Functions*
3. *Accessing and Sharing Data*
4. *Data Base Design*

Evening Wrap-up Session (Location - TBD)

7:30 pm - 8:30 pm: Small group gathering

Facilitators, recorders, selected technical experts, and volunteers from each breakout group will convene to summarize and evaluate approaches, outcomes from the day's sessions. Main goal is to ensure that each breakout group is working in parallel towards common solutions. Highlights and key findings from each breakout group will be synthesized for presentation to the full group the following morning. Sharing of information between the groups will be fostered.

THURSDAY, JUNE 20, 1996

Plenary Session (Conference Room)

8:15 am - 8:45 am: Presentation by each breakout group on summary of findings and/or preliminary recommendations that require coordination between the four groups

Breakout Groups Conclude (Breakout Rooms)

8:45 am - 10:00 am: Concurrent breakout groups

1. *User Interface for Chesapeake Information Management System*
2. *Information Management System Functions*
3. *Accessing and Sharing Data*
4. *Data Base Design*

10:00 am - 10:15 am: Break (coffee and donuts)

10:15 am - 11:00 am: Concurrent breakout groups conclude

1. *User Interface for Chesapeake Information Management System*
2. *Information Management System Functions*

3. *Accessing and Sharing Data*
4. *Data Base Design*

Plenary Working Session (Conference Room)

11:15 am - 12:15 pm: Presentations by each breakout group of initial concept designs for a Chesapeake Information Management System:

1. *User Interface for Chesapeake Information Management System*
2. *Information Management System Functions*

Lunch: 12:15 pm - 1:00 pm (Conference Room) (Brown Bag Lunches Catered by USGS)

Speaker: Submerged Aquatic Vegetation Data Collection, Analysis, and Reporting *[Dr. Dave Wilcox, Virginia Institute of Marine Science]*

Plenary Working Session (Conference Room)

1:00 pm - 2:00 pm: Conclusion of presentations by each breakout group of initial concept designs for a Chesapeake Information Management System:

3. *Accessing and Sharing Data*
4. *Data Base Design*

2:00 pm - 3:00 pm: Summary of recommended Chesapeake Information Management System design; Comparison of workshop outcome to original conceptual framework and recommendations *[Lowell Bahner, Data Center Manager, NOAA, Cary Gaunt, Project Manager, SAIC]*

3:00 pm - 3:30 pm: Wrap-up, discussion of barriers and opportunities to implementation, and next steps *[Rich Batiuk, Associate Director of Science, EPA]*

OVERVIEW OF BREAKOUT GROUPS

1. User Interface for Chesapeake Information Management System: Breakout group participants will discuss appropriate interface technology options (e.g., browser technology) and evaluate sample user interface screens. Screen selection options, as well as screen layout and features, will also be discussed. The basis for a prototype will be developed. *[Example Topics: online data library or subject-matter directory, approach to compiling and accessing Metadata, example screens, lists of variables, range selection, common queries, ad hoc queries, spatial selection/display, character only selection, access to non-CBPO data bases, static text and maps, scrollable text and maps, graphical "hot spots" for point and click selection on maps, text input for queries, automatic links to other sites].*

   Facilitators: Rich Batiuk (CBP); Jim Parker (SAIC); Technical Expert: Wilson Hom (SAIC)
   Recorders: Brad McGuire (SAIC); David Kimball (Chesapeake Research Consortium)

2. **Information Management System Functions:** This group is closely related to the Interface session, but focuses more on the approaches that must be used, and/or actions that must be performed on the data, to achieve the functions specified in the user interface. The breakout group will use as a starting point a set of pre-established functions that serve as the basis for the interfaces proposed for Session #1. This section will address some of the following questions: What must the information system accomplish? What functions must it perform? What approaches/actions must be used to perform these functions? For each proposed function, breakout group participants will address the following questions with regards to achieving the function -- What are the steps needed to collect, assemble, and process the data to meet the function? What are the decision-rules needed to perform each function? *[Example Topics: download/export data to user, perform standard queries (e.g., view/print trend analysis results, environmental indicators), perform user-defined queries, identify and print maps, obtain a list of all available data by various geographic subcategories].*

   Facilitators: Bruce Michael, MD Department of Natural Resources for CBP; Craig Miller (SAIC)
   Recorders: Kate Bennett (SAIC); Carolyn Lieberman (CRC)

3. **Accessing and Sharing Data:** Breakout group participants will focus on determining options/best ways (e.g., establishing a standard) for the wide range of CBP data users to access and share data and communicate effectively. Participants will also address data and information input and output designs.

[Example Topics: user requirements such as Internet, remote ftp, telnet, modem, browser, age and type of platform; planning for future technologies; formats for input and output.]

Facilitators: Bob Edwards, Susquehanna River Basin Commission for CBP; Mark Jenkins (SAIC)
Recorders: Ken Pruitt (SAIC); Rob Nelson (CBP)

4. **Data Base Design:** Breakout group participants will determine which data base design model is the most appropriate design to effectively achieve the goals of the Chesapeake Information Management System. The models under consideration are hierarchical, network, relational, object, object-relational, and multidimensional. The breakout group will also determine the appropriate architecture for the database management system (i.e., client server, distributed, or centralized).

Facilitators: Peter Bergstrom, U.S. Fish and Wildlife Service for CBP; Diane Spaulding (SAIC); Technical Expert: Bill Samuels (SAIC)
Recorder: Christy Stoll (SAIC)

APPENDIX B

ATTENDANCE LIST

Marjorie Adkins
Alliance for the Chesapeake Bay
P.O. Box 1981
Richmond, VA 23216

Diana Alegre
MD Department of the Environment
2500 Broening Highway
Baltimore, MD 21224

Gary Anderson
Virginia Institute of Marine Science
College of William & Mary
P.O. Box 1346
Gloucester Point, VA 23062-1346

Lowell Bahner
NOAA Chesapeake Bay Office
Suite 107A
410 Severn Ave
Annapolis, MD 21403

Roger Barlow
US Geological Survey
USGS National Center/MS511
12201 Sunrise Valley Drive
Reston, VA 22092

Richard Batiuk
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Andrew Battin
US EPA/Waterside Mall
401 M Street, SW
Washington, DC 20460

Kate Bennett
Science Applications International Corporation
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Peter Bergstrom
US Fish & Wildlife Service
Chesapeake Bay Field Office
177 Admiral Cochrane Drive
Annapolis, MD 21401

Carin Bisland
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Candice Bostwick
US Geological Survey
437 National Center
12201 Sunrise Valley Drive
Reston, VA 22092

Claire Buchanan
Interstate Commission on the Potomac
River Basin
Suite 300
6110 Executive Boulevard
Rockville, MD 20852-3903

Brian Burch
USEPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Barbara Butler
Chesapeake Research Consortium
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Arthur Butt
VA Department of Environmental Quality
629 E. Main Street
Richmond, VA 23240-0009

Susan Clark
US Geological Survey
National Center
12201 Sun Rise Valley Drive
Reston, VA 22092

Rosalind E. Cohen
NOAA/NEDIS
SSMC3, 4th Floor, #4822
1315 East West Highway
Silver Spring, MD 20910-3282

Cory Dippel
Chesapeake Research Consortium
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Angelo Edge
Management Technology Incorporated
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Robert E. Edwards
Susquehanna River Basin Commission
1721 North Front Street
Harrisburg, PA 17102

Kelly Eisenman
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Gary Fisher
US Geological Survey
208 Carroll Building
8600 West LaSalle Drive
Towson, MD 21286

Tim Foresman
University of Maryland - Baltimore Campus
Catonsville, MD

Doug Forsell
US Fish and Wildlife Service
177 Admiral Cochrane Dr.
Annapolis, MD 21401

Pat Gant
US EPA
Suite 200
210 Defense Highway
Annapolis, MD 21401

Cary Gaunt
Science Applications International Corporation
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Scott Gibbons
Interstate Commission on the Potomac
River Basin
Suite 300
6110 Executive Blvd
Rockville, MD 20852

Steven Grass
PA Department of Environmental Protection
P.O. Box 2063
Harrisburg, PA 17105-2063

David Greene
US Geological Survey
National Center
12201 Sun Rise Valley Drive
Reston, VA 22092

Deena Grinbaum
US Geological Survey
National Center
12201 Sun Rise Valley Drive
Reston, VA 22092

Grace Hanrahan
US EPA/Region 3
401 M Street, SW
Washington, DC 20460

Carlton Haywood
Interstate Commission on the Potomac River
Basin
Suite 300
6110 Executive Blvd.
Rockville, MD 20852

Brian Hazelwood
Metropolitan Washington Council of
Governments
777 North Capitol Street, NE
Washington, DC 20002

John Helly
San Diego Supercomputer Center
P.O. Box 85608
San Diego, CA 92186-9784

Paula Hill-Jasinski
University of Maryland
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Phil Hissom
Management Technology Incorporated
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Scott Hoffman
US Geological Survey
840 Market Street
Lemoyne, PA 17043

Wilson Hom
Science Applications International Corporation
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Tally Huang
MD Department of Natural Resources
580 Taylor Avenue
Annapolis, MD 21401

Dan Jacobs
University of Maryland
Route 1
College Park, MD 20742-7640

Mark Jenkins
Science Applications International Corporation
200 Harry S. Truman Parkway
Suite 400
Annapolis, MD 21401

Jackie Johnson
Interstate Commission on the Potomac
River Basin
US EPA Chesapeake Bay Program
410 Severn Avenue, Suite 109
Annapolis, MD 21403

David Kimball
US EPA Chesapeake Bay Program Office
Chesapeake Research Consortium
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Mike Lane
Old Dominion University
Applied Marine Research Lab
1034 West 45th Street
Norfolk, VA 23529

Lucille Lanham-Ridley
VA Health Department
Division of Shellfish Sanitation
1500 East Main St., Suite 109
Richmond, VA 23219

Cathy Lascara
Old Dominion University
1043 35th Street
Annapolis, MD 21403

Peter Legg
Management Technology Incorporated
410 Severn Avenue, Suite 109
Annapolis, MD 21401

Connie Lewis
MD of Natural Resources
580 Taylor Ave.
Annapolis, MD 21401

Carolyn Liebermann
Chesapeake Research Consortium
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Lara Lutz
Alliance for the Chesapeake Bay
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Joe Macknis
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Russ Mader
Natural Resources Conservation Service
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Robert Magnien
MD Department of Natural Resources
Tawes Building
580 Taylor Avenue
Annapolis, MD 21401

Brad Maguire
Science Applications International Corporation
Tower I
1710 Goodridge Dr.
McLean, VA 22102

Ross Mandel
Interstate Commission on the Potomac River
Basin
Suite 300
6110 Executive Blvd
Rockville, MD 20852

Menchu Martinez
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Diane M. McCarthy
VA Department of Conservation & Recreation
203 Governor St., Suite 206
Richmond, VA 23219

Tawna Mertz
University of Maryland
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Bruce Michael
MD Department of Natural Resources
Tawes State Office Building
580 Taylor Avenue
Annapolis, MD 21401

Paul Miller
US Army Corps of Engineers/Waterway
Experiment Station
GGYH (Paul Miller)
3909 Halls Ferry Road
Vicksburg, MS 39180

Craig Miller
Science Applications International Corporation
Tower I
1710 Goodridge Dr.
McLean, VA 22102

Rob Nelson
MD Department of Natural Resources
Tawes Office Building
580 Taylor Avenue
Annapolis, MD 21401

Brand Niemann
National Biological Service
Rm MIB 3656
1849 C Street, NW
Washington, DC 20240

Marcia Olson
NOAA Chesapeake Bay Office
410 Severn Ave.
Suite 107A
Annapolis, MD 21403

Elissavet Pontikakis
Metropolitan Washington Council of
Governments
Suite 300
777 North Capitol Street, NE
Washington, DC 20002

John Posey
University of Maryland
Chesapeake Biological Laboratory
P.O. Box 38
Solomons, MD 20688-0038

Pranash Pranckevicius
US EPA/Great Lakes National Program
77 West Jackson Blvd
Chicago, IL 60604

Ken Pruitt
Science Applications International Corporation
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Mary Ranier
US EPA/Region 3
401 M Street, SW
Washington, DC 20460

Frances Rohland
University of MD Center for Estuarine and
Environmental Studies - Chesapeake
Biological Laboratory
P.O. Box 38
Solomons, MD 20688

George Sackett
MD Department of Natural Resources
B-2
580 Taylor Avenue
Annapolis, MD 21401

Bill Samuels
Science Applications International Corporation
Tower I
1710 Goodridge Dr.
McLean, VA 22102

Ron Santos
US Army of Corps of Engineers
Attn: CENAB-PL-EP
PO Box 1715
Baltimore, MD 21203

Debbi Skelly
MD Department of Natural Resources
580 Taylor Avenue
Annapolis, MD 21401

Gary Smith
MD Department of Natural
Resources/Cooperative Oxford Lab
904 S. Morris Street
Oxford, MD 21654-9724

Liz Smith
Old Dominion University
Applied Marine Research Lab
1034 West 45th Street
Norfolk, VA 23529

Diane Spalding
Science Applications International Corporation
Tower I
1710 Goodridge Dr.
McLean, VA 22102

Christy Stoll
Science Applications International Corporation
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Nita Sylvester
US EPA Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Hung Truong
(Affiliation unknown)

Howard Weinberg
University of Maryland
US EPA/Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

Allison Wiedeman
US EPA/Chesapeake Bay Program Office
410 Severn Avenue, Suite 109
Annapolis, MD 21403

David Wilcox
Virginia Institute of Marine Science
College of William & Mary
P.O. Box 1346
Gloucester Point, VA 23062-1346

Lois L. Woodward
NOAA Chesapeake Bay Office
410 Severn Ave., Ste 107A
Annapolis, MD 21403

Scott Wright
Management Technology Incorporated
410 Severn Avenue, Suite 109
Annapolis, MD 21403