

Gulf of Maine Data and Information Systems

Workshop Proceedings

3-5 November, 1993
The University of New Hampshire
Durham, New Hampshire 03824

convened by

**Regional Marine Research Program for the Gulf of Maine
Regional Association for Research on the Gulf of Maine
Gulf of Maine Council on the Marine Environment**

edited by the Workshop Steering Committee

**Donald K. Phelps, U.S. Environmental Protection Agency, Chair
Wendell Brown, University of New Hampshire
Marilyn Buchholtz ten Brink, U. S. Geological Survey
John Paul, U.S. Environmental Protection Agency
Jeffrey S. Rosen, American Management Systems, Inc.**

produced by

**Eugenia Braasch, Associate Director
Regional Association for Research on the Gulf of Maine
Dartmouth College
Hanover, NH 03755**

PHE
08335

Acknowledgements

Wendell Brown and Mariellen Carpenter Lee, at the University of New Hampshire, made the conference arrangements which were characterized by pleasant surroundings and a well-run meeting. The meeting was supported by the Regional Marine Research Program, preparation of the proceedings was provided by the Regional Association for Research on the Gulf of Maine, printing and distribution of the report was sponsored by the Gulf of Maine Council on the Marine Environment.

To be referenced as:

Gulf of Maine Data and Information Systems: Workshop Proceedings, RARGOM Report number 93-1, Phelps, D.K. et al., Editors

Table of Contents

Forward.....	1
I. Executive Summary / Recommendations.....	2
II. White Papers.....	4
III. Background Information / Plenary Papers.....	20
Academic Research Systems.....	28
Regional Systems.....	70
Agency Systems.....	80
IV. Working Group Reports	
Regional Marine Research Program IMS.....	170
Gulf of Maine IMS.....	177
Appendix A: Workshop Agenda.....	182
Appendix B: Participants.....	185
Appendix C: Posters.....	191

Forward

The long-term research goal for the Regional Marine Research Program (RMRP) of the Gulf of Maine (GoM) is to produce predictive models for the Gulf which incorporate physical circulation, chemical fate and transport, and biological (ecosystem) effects. Development and interpretation of processes in the Gulf of Maine is also of interest. The pivotal role of data and information management within the broad Marine Research Community of the Gulf of Maine emerged as a key issue during the initial Regional Marine Research Board (RMRB) meetings.

RMRB agreed, that in order to achieve the long-term goal, it was essential to convene a workshop including the major research interests/institutions in the Gulf of Maine. The purposes of the Workshop are: 1). to address institutional roles/responsibilities, and 2). to determine the levels of interest/commitment in the establishment of a Regional Data and Information Management System. Three working groups were identified composed of a core of individuals from 1). Researchers currently and/or potentially funded directly by the RMRB; 2). members of the Regional Association for Research on the Gulf of Maine (RARGOM) and the Gulf of Maine Council on the Marine Environment (CME); and 3). Representatives from the other Regional Marine Research Program Boards.

The Regional Association for Research on the Gulf of Maine and the Gulf of Maine Council on the Marine Environment joined the Workshop effort as Co-conveners, reflecting their mutual interest in the subject. The Workshop was convened on November 3rd through 5th, 1993 at the New England Center on the University of New Hampshire campus. The workshop objectives were addressed through presentations of current data and information systems during the plenary session followed by discussion held in the working group sessions. Each group was charged with producing recommendations focused on identified issues. Due to the fact that only three Regional Marine Board representatives were able to attend the Workshop, working group three was dispersed among the other two groups.

This is the record of the meeting. It includes the an overall executive summary and recommendations, the working groups recommendations, plus all of the pre and post workshop support materials.

I. Executive Summary / Recommendations

Executive Summary and Recommendations

The IMS Workshop was composed of two working groups; the Regional Marine Research Program IMS Working Group (#1), chaired by Daniel Lynch with Glenn Flierl as Rapporteur; and the Gulf of Maine IMS Working Group (#2), chaired by Paul Boudreau with John Sowles as Rapporteur. The Steering Committee/Editors wish to express their appreciation for the excellent job done in the conduct of the workshop by these individuals.

Each of the working groups were given charges and presented with a series of Priority and Secondary Issues about which discussions were to be focused (see the body of these Proceedings for details). While there was considerable redundancy among the issues presented to each working group, priorities were ordered differently for each.

Recommendations were made by each of the working groups. Several generic recommendations emerged:

1. The Gulf of Maine research and management community is committed to supporting an Information Management System.

RECOMMENDATION: A regional Gulf of Maine Information Management System should be developed.

2. The Gulf of Maine Information Management System will be accessible to all and will be formulated as a distributed, networked system.

RECOMMENDATION: A distributed system utilizing INTERNET, as opposed to a centralized system, should be implemented.

3. The Gulf of Maine Information Management System will build on existing systems; it will generate and distribute a directory of these systems, and it will strive for compatibility with important specific systems.

RECOMMENDATION: The Gulf of Maine Information Management System should not re-invent any wheels. A vehicle should be put in place to generate and distribute a directory of the component systems and assure compatibility with important specific systems.

4. An oversight mechanism will be emplaced to insure that community needs are being addressed.

RECOMMENDATION: A Gulf of Maine IMS Oversight Group should be formed and charged with ensuring that the needs of various groups are incorporated in the planning process.

5. Connectivity, compatibility and data and information exchange procedures will be the joint responsibility of data generators, users, and a GOM implementation team.

RECOMMENDATION: A Technical Start-up Committee should be established.

6. The Gulf of Maine Information Management System will have a vision statement and identified goals.

RECOMMENDATION: Charge the Technical Start-up Committee with proposing short and long-term goals for the Gulf of Maine Information Management System and the Oversight Group with developing a vision statement.

Editors' Notes

Due to Working Group time limitations, the Working Groups were unable to discuss and reach consensus on a number of the identified issues. One such issue, of interest to the Steering Committee, was the issue of Quality Control. While Working Group notes indicate that this issue was discussed, no specific recommendation was forthcoming. It is the sense of the Steering Committee that the Working Groups deferred resolution and guidance on this issue (as well as other unresolved issues) to the purview of the Technical Start-up Committee and the Oversight Committee.

Generic Comment. As regards a strategy for implementation of the Workshop recommendations, Roles and Responsibilities must be clearly defined within the Gulf of Maine Marine Research community. Who recommends? Who approves? Who funds? What vehicle should/will be used for funding? Who monitors?

While the suggested Oversight Group should reflect the broad interests of the Research Community as represented by RARGOM, the CME, and RMRP, there must be a mechanism for dealing with the short-term information management needs of the RMRP. We recommend that the Technical Start-up Committee serve that purpose. We further recommend that the RMRP should be the focal point for the initial developmental stages of the IMS. This is based on the assumption that the RMRP will fund the initial effort. Toward this end we recommend that the RMRB:

1. Appoint the Technical Start-up Committee, taking into consideration the discussions contained in the Working Group meeting notes and recommendations.
2. Solicit nominations from the Research Community at large (particularly RARGOM and the CME) for membership in the Oversight Group.
3. Prepare a solicitation for Proposals, the purpose of which will be to initiate the development of an IMS Central based on the Workshop Recommendations in which the scope of work, outputs, and a target range of funding are clearly presented. The Technical Start-up Committee could provide guidance in the development of the scope of work and outputs (unless there is a perceived conflict of interest).

II. White Papers

10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

Framework for Environmental Information Management Systems

Jeffrey S. Rosen
Gulf of Maine Data and Information Management Systems Workshop
Steering Committee
American Management Systems, Inc.
One Bowdoin Square
Boston, MA 02114-2919
617-544-3085

Introduction

The purpose of this document is to establish a framework for the discussions at the Gulf of Maine Data and Information Management Systems workshop being held at the New England Center, November 3-5, 1993. Two matrices are presented below which define the levels of data and users (figure 1) the types of data and corresponding users. These two figures are presented as a concise well-defined domain of environmental information management.

Environmental research, monitoring and management have a clearly definable set of steps that are followed to a varying degree in all data collection efforts. These steps can be summarized by an Enterprise Model that includes:

- Identify Issues which will be addressed
- Define a design and an approach to measure or monitor the processes of interest
- Implement the sampling plan
- Collect the required data
- Organize, process and analyze data
- Interpret data
- Formulate conclusions and make decisions if appropriate
- Develop new hypotheses or Implement decisions
- Enforce decisions
- Defend decisions
- Iterate the process

At all steps in this process, data are a key resource. The efficient access and use of the data require well-designed accessible data systems. A conceptual framework for the users of environmental data is summarized in figures one and two below.

Figure one is a conceptual model of the levels of data and of the general categories of data and information users. On the left side of the figure are the levels of data. The lowest level is identified as Field Information Systems. It is given the designation D1 to facilitate discussion. Similarly, each level above D1 identifies a specific aggregation of the data. Each level is distinct from the other levels in the content and nature of the data contained in it. Yet each level is also related to the level below it and above it. In a properly designed information system the data can be relationally joined between adjacent levels.

The data levels decrease in width going up the model because the aggregation tends to decrease the volume of data. It is important to note however, that as the volume of data decreases the volume of required documentation will increase dramatically.

In most cases the process of analyzing interpreting and aggregating the data is far too involved and technically demanding to justify regenerating the results each time a user needs the data turned into information. This model suggests that the data should be stored in different but related information systems.

Levels of Data Aggregations and Users

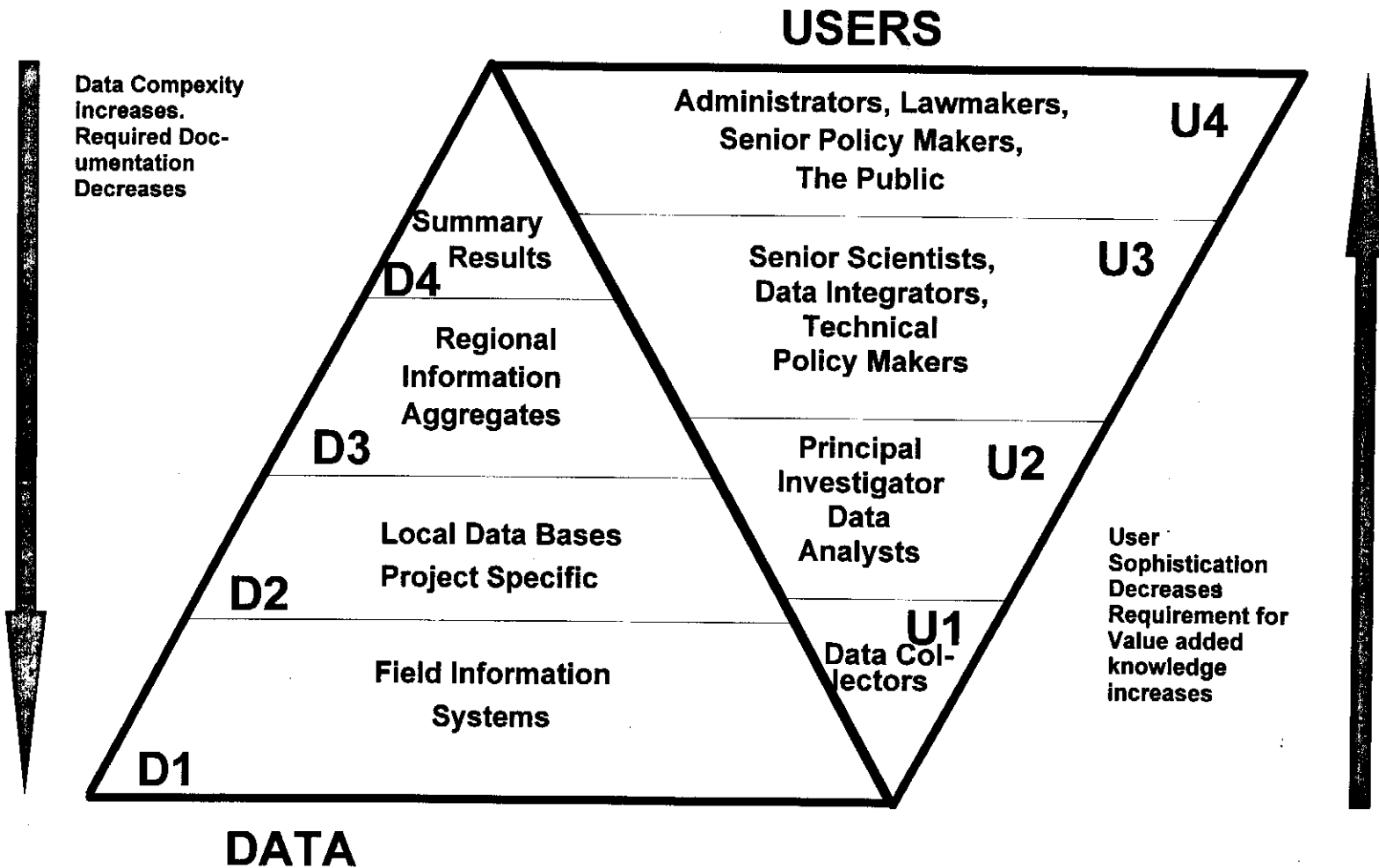


Figure 1 - Conceptual model of the levels of data aggregation and the levels of users

The right side of figure one represents the different users of the information. Clearly this is not meant to be an exhaustive list or characterization of the users but rather a conceptualization. The users at each level have very specific requirements of the data in the adjacent data compartments. The model is not meant to suggest that users will only access data in the data level immediately adjacent to their user level. Users will probably access data at a number of data levels but will use the data level adjacent to their user level most often. The user compartments increase in width going up because the number of potential users at each level increases.

The line between the users and data represents the user interfaces to the data. This is a critical component of the model. If the user interfaces are not appropriate to the users and the data, then the information system will not succeed.

Figure two represents a functional characterization of the types of data and the disciplines of the users. The disciplines are listed across the tops of the columns. Again this is not meant to be an exhaustive characterization but rather a conceptual model. Each discipline is likely to have users from a variety of the user levels described in figure 1. These users will interact with many different types of data to complete their mission. The types of data are listed at the left of each row. Each cell in the matrix represents the interactions between a discipline and a class of data. The content of each cell indicates the interaction between the users in the discipline in the columns and the types of data in the rows.

G = The users in this discipline are involved in generating this kind of data

R = The users in this discipline are likely to Read (access) data

M = The users in this discipline are involved in maintaining this kind data

D = The users in this discipline may be involved in deleting data of this type from data bases.

(The contents of the cells in figure two are meant as an example to encourage further discussion)

Users Type Data Type	Administrators	Researchers	Conservation groups	Public and others	ADP Staff professionals
Historical	GRMD	GRMD	GRMD	R	RMD
Characterization	GRMD	GRMD	RM	GRM	GRMD
Monitoring	GRMD	GRMD	GRMD	RM	GRMD
Remote Sensing	GRMD	R	R	R	RM
Research	R	GRMD	R	R	GRMD

Figure 2 -- An example of the Conceptual Matrix of users and data types.

For each cell in figure two a set of additional issues is listed below. These issues should be considered to help determine the processes and controls necessary for the information system to function properly.

The issues that should be considered for each compartment in the matrix include (but are not limited to):

- Does the group have use for the data?
- For how long must the data remain proprietary from the data generator's perspective ?
- What access should the group have to the data (Generate, Read, Maintain, Delete)?
- What are the uses of the data in each compartment, e.g., research, compliance, control, education, background information, litigation, etc.
- What standards are required in the compartment?
- What is the required QA/QC for the data?
- What tools are needed to access and use the data (software)?
- What is the level of expertise needed to access and correctly use the data?
- What infrastructure is required to access the data (hardware, communications, storage)?
- What is the annual budget for information processing in the cell?

- What is the priority for consolidating the data and making it available to the different levels of users (as defined in Figure 1)?
- Does the Discipline have a role in:
 - generating the data
 - processing the data
 - interpreting the data
 - aggregating the data
 - documenting the data
 - maintaining the data (backups, storage, data base storage, etc.)
 - distributing the data

Discussion of these issues will define many of the requirements of the system and will also force the discussion of some difficult issues that must be resolved if an effective information management system is to be developed for an environmental research and monitoring program.

Information Management for the RMRP

Daniel Lynch, Dartmouth College
Glenn Flierl, MIT

22 October 1993

What is the RMRP?

The Regional Marine Research Program was created in 1990 by act of Congress. Overall, the program is responsible for

- a) setting priorities for regional marine and coastal research in support of efforts to safeguard the water quality and ecosystem health of each of 9 identified regions; and
- b) carrying out research through grants and improved coordination.

The Gulf of Maine is one of the nine identified regions; its program is administered through an executive office at the University of Maine.

The 10-year goal of the Gulf Program is to work toward a suite of models that collectively simulate how the Gulf of Maine ecosystem, and its interacting components, function naturally and under stress. To date, approximately \$3M has been granted to 13 different projects, which are beginning in 1993(see appendix); and an additional RFP will be issued in early 1994, for roughly \$1M in new projects.

What are its information needs?

The complexity of the research planned, its scope and duration, the need for real-time interaction among disciplines, institutions, and PI's all demand a carefully designed approach to data management and information flow. In particular it will be critical that data collection, processing, and distribution be carefully planned and coordinated to ensure

- timely and easy distribution of both new and archived data
- the highest standards of quality assurance
- feedback to the field program activities
- permanent archiving of a large volume of new RMRP data and information

We take as a starting point that the goals of the RMRP are ambitious and that a complete model-based description of the GoM ecosystems is likely to continue to elude scientists at the end of the expressed 10-year horizon. It is therefore important that the research community commit to a framework for efficient exchange and archival of an *evolving* information base.

We hypothesize an **RMRP Information Management System (RMRP-IMS)**. Such a system will necessarily include diverse data types:

- historical data
- new observational data
- real-time observations
- models and model products
- bibliographic information

obtained from a diverse set of data providers, including:

- federal archives
- real-time data from observational networks
- individual research project archives

and provide several modes of information service, including:

- browsing, querying, searching
- data delivery/transfer across network
- graphical display and visualization
- statistical and other data processing

A well-designed RMRP-IMS will serve both to facilitate research now and in the future, and to convey research results to the management community in the long run. It is vitally important, however, that the system be first and foremost one which amplifies, coordinates, and preserves the efforts of individual scientists. A long-term adherence to this principle is crucial if we are to build a reliable information base for scientific resource management which is organically related to contemporary understanding of the Gulf.

Accordingly, we focus the RMRP working group on the needs of scientific researchers. In the short run, scientists will be both data users and data suppliers. Their needs include:

easy and timely archival of new data with uniform standards of quality assurance and data "ownership", attribution, etc.

easy access to archival and contemporary data; ideally an RMRP-IMS would provide a regional focal point for the diverse data sources maintained by various federal and regional agencies, as well as to data currently being obtained through research projects.

continuity (stability) of information over time and among institutions -- research is a distributed process in search of a unified central understanding of nature

relief from the burden of providing ongoing, ad-hoc data retrieval for future researchers -- i.e. simple, one-time transfer of archived data to a single central repository

an overarching, integrating framework that simplifies data browsing, query, and retrieval from diverse data sources and formats.

Facts of Life (talking points)

Do we hold these truths to be self-evident?

- 1a) Researchers will proceed to build useful data sets and analysis tools within whatever data management environment is available/appropriate to the task.
- 1b) Each component of this distributed process reflects the research priorities and intrinsic needs of working scientists and is therefore inherently strong.
- 1c) Centralized coordination is needed to enhance and preserve the value of these efforts to the larger community.

- 1d) Therefore a RMRP-IMS should be fundamentally a distributed system, with centralized guidance, support and standards.
- 2a) Any system constructed from useful, distributed parts will involve a collection of incompatible equipment, software, and local institutional constraints. It will not be possible to dictate detailed technical choices for either data providers or data users.
- 2b) Any useful IMS must be designed to outlive the technical platform on which it was first built.
- 2c) Individual data archives will migrate over time from one technical platform to another; a distributed system must accommodate this without widespread dislocation.
- 2d) A central focus of RMRP-IMS development should be in defining and creating standard software interfaces among diverse data suppliers and diverse data users which are machine- and system-independent.
- 3a) Internet is the evolving scientific communication medium of choice.
- 3b) Unix is the evolving Operating System of choice.
- 3c) X-windows is the emerging graphics standard.
- 3d) It will be difficult to take full advantage of these standards in the near term
- 4) An RMRP-IMS should be common among several regional research programs. Conspicuous in this regard is the GLOBEC Georges Bank program which is just beginning.

Thoughts on the design of an RMRP-IMS

We advocate the development of a distributed information system which provides useful exchange among

- a diverse group of **data providers**, from agencies such as NMFS to individual scientists, and
- a wide group of scientific **data users** who have widely different needs.

Certain design considerations must be respected:

- a) It must be possible and relatively simple to connect existing archives into the system without data reformatting and without changing the existing procedures for adding/editing data at the individual lab/archive.
- b) Data products will be developed by users; however, the system should have a means for easily feeding those products back into the system and making them available to all. One cannot predict these in advance or expect a single IMS group to develop them.

- c) Data servers must be able to respond to location queries in a timely fashion. Some directory and inventory management will be required.
- d) The data system must carefully separate the IMS functions from the user interface. Different user groups will require different user interfaces (built with different languages) with different functionalities. For example, some people will want a "point and click" system which will lead one to the desired data, while others know pretty much what they want and wish to incorporate it directly into a program (e.g., a data assimilation activity).
- e) Application programs for data retrieval, processing and display will be developed by users; again, a sharing mechanism is important. This implies that the data system needs a set of standards so that one knows exactly what calls an application can make to obtain information.
- f) "Filters" or data transformation/ combination procedures should also be relatively simple for users to develop and share.

A Strawman RMRP-IMS Project

Overview

We outline a project devoted to the creation of a distributed IMS for the RMRP. The attached Figure illustrates the concept. This IMS would evolve over a 5-8 year period under the project supervision. In the long run its operation, maintenance and updating would be turned over to two permanent operating agencies:

- at the regional level, the Gulf of Maine Council on the Marine Environment (CME) or its designated subcontractor;
- at the national and international level, NODC.

The technology transfer would occur continuously, as portions of the system mature. The project itself would be a collaboration among several research institutions, with an RMRP Data Management Office (DMO) at one of the sites providing centralized project coordination. The DMO would be a temporary (3-5 years) creation; the long-run outcome would be the IMS itself, with CME the primary owner/manager and NODC a primary distribution mechanism for non-regional users.

Data

The scope of the information system would be quite comprehensive. Naturally, this includes conventional data types (atmospheric, physical, biological, and chemical oceanographic data). It also includes computational data in the form of model output and other data processing products. Computer programs for data processing and/or graphical display are yet another form of data of concern. Finally, there are important real-time data streams which are not archived. A comprehensive RMRP information system will need to include all of these data products.

System

a) Network: An Internet-based system is envisioned, with Unix workstations the primary clients and servers. Network services and system management in this arena is available almost universally within the scope of the RMRP. Needed additions/extensions include

- i) the transmission of satellite imagery (potentially too data-intensive for large-scale internet use);
- ii) networking to ships to provide real-time and archival information during field work;
- iii) networking to moored or drifting instruments for capture or real-time data.

b) Software: We expect to utilize the results of previous and on-going efforts developing distributed data systems, such as the JGOFS Data Management project (See Flierl et al.) and the Distributed Ocean Data System (Cornillon, et al.). These are Internet-based client/server systems, with ability to handle diverse all data sets. Source code is available. We expect one of these will be a sufficient platform to begin with; for our user community, however, we may wish to extend the system with

- networking to DOS and Macintosh;
- user utilities for DOS and Macintosh
- user graphics
- data services ("Methods") for new data types

The DMO

The Data Management Office will be the central focus of services to both data suppliers and data users. Among its service functions are

- maintaining a master data directory
- monitoring data submission
- quality control for new data
- data storage (only in cases where storage is not available at the originating institution)
- user support and workshops
- transmission to/liaison with permanent archive

In addition, the DMO would coordinate of software development and distributed data entry/maintenance. These "producer" functions themselves would be carried out at participating RMRP research sites. Each site would be responsible for software generation, the flow of data products and related programs into the system, and local data base maintenance, with special attention to the scientific users and data sources represented. The DMO would be physically located at one of these.

Resources

It is expected that during startup, 1/2 FTE skilled technician will be required at each of the software development sites, plus an additional 1/2 FTE for administrative and support services at the DMO site. Regional travel sufficient for 2 coordination meetings per year would be required, plus 1 workshop per year. Finally, a developmental workstation will be required at each site.

APPENDIX

Gulf of Maine Research Projects Funded June 1993:

Sources, Transport, and Nutrient Environment of Toxic "Red Tide" Populations in the Western Gulf of Maine. Donald M. Anderson, Woods Hole Oceanographic Institution; Wayne R. Geyer, Applied Ocean Physics & Engineering, Woods Hole Oceanographic Institution; Bradford Butman and Richard Signell, U.S. Geological Survey, Woods Hole; Peter J.S. Franks, Scripps Institution of Oceanography; Theodore C. Loder, III, University of New Hampshire; \$600,000; 36 months.

The Relation of Slope Water Inflow to the Eastern Maine Coastal Current and to the Jordan Basin Gyre. Neal R. Pettigrew, University of Maine; \$600,000; 36 months.

An Observational/Modeling Study of Western Gulf of Maine Circulation. Wendell S. Brown, University of New Hampshire and Daniel R. Lynch, Dartmouth College; \$400,000; 36 months.

Microplankton Bloom Dynamics in the Gulf of Maine; Nutrients, Hydrography and Community Structure: A Program to Investigate a Nutrient Trap in the Western Gulf of Maine. Maureen D. Keller, Chris Garside and Michael E. Sieracki, Bigelow Laboratory for Ocean Sciences; \$61,276; 12 months.

A Program to Investigate Long-term Seasonal and Interannual Variability in the Optical Properties and Optically-Active Constituents in the Gulf of Maine Waters: Database Development and Historical Data Analysis. Steven G. Ackleson, Office of Naval Research and Charles S. Yentsch, Bigelow Laboratory for Ocean Sciences; \$65,091; 12 months.

Gulf of Maine contaminated-Sediment Database: Compilation, Desk-Top Access and Analysis of Existing Data for Scientific and Policy Needs. Marilyn Buchholtz ten Brink, Frank T. Manheim, John Hathaway, U.S. Geological Survey; Stephen H. Jones, Larry Ward, University of New Hampshire; Peter Larsen, Bigelow Laboratory for Ocean Sciences; Bruce W. Tripp, John W. Farrington, Woods Hole Oceanographic Institution; and Gordon T. Wallace, University of Massachusetts, Boston; \$200,010; 24 months.

Preparation of a Review of Uptake and Depuration Kinetics of Algal Toxins by Filter-feeding Molluscan Shellfish. Sandra E. Shumway, Bigelow Laboratory for Ocean Sciences, and V. Monica Bricelj, State University of New York at Stony Brook; \$90,596; 24 months.

Publication of a Phytoplankton Manual for the Gulf of Maine. Robert A. Andersen, Bigelow Laboratory for Ocean Sciences; \$61,474; 12 months.

Research Projects Submitted for Funding Starting in Fall 1993:

An Observational Study of Gulf of Maine Circulation Variability: An Augmentation Request to the Regional Marine Research Program for the Gulf of Maine. Wendell S. Brown, Theodore D. Loder, III, University of New Hampshire; Donald Anderson, W. Rockwell Geyer, Woods Hole Oceanographic Institution; Neal R. Pettigrew, University of Maine; \$246,000; 36 months.

Spatial and Temporal Variability of Satellite-Derived Sea Surface Temperature and Pigment Fields in the Northern Gulf of Maine. James J. Bisagni, NOAA/NMFS Narragansett, RI, and Peter Cornillon, University of Rhode Island; \$99,870; 36 months.

Construction of Satellite-Based Wave Climatology for the Gulf of Maine. Vijay G. Panchang, University of Maine; \$43,957; 12 months.

Atmospheric Nitrogen Inputs to the Gulf of Maine: The Relative Importance of Inorganic and Organic Forms. Robert W. Talbot and Byard W. Mosher, University of New Hampshire; \$233,136; 36 months.

Cycling of Particle-Reactive Trace Metals and Organic Pollutants in the Gulf of Maine. Philip M. Gschwend, Massachusetts Institute of Technology; Ken O. Buesseler and S. Bradley Moran, Woods Hole Oceanographic Institution; \$427,140; 36 months.

Regional Data and Information System Requirements for the Gulf of Maine

Paul Boudreau
Department of Fisheries and Oceans Canada
P.O. Box 1006
Dartmouth, Nova Scotia.
B2Y 4A2
Canada

Introduction:

The Gulf of Maine that borders on three US States and two Canadian Provinces, is a semi-enclosed body of water of interest to a large number of agencies. The variety of activities in the Gulf and the proximity of a large and diverse community of people has placed great pressures on the system. Activities from shellfish harvesting and aquaculture to urban and industrial waste disposal are becoming more and more demanding on the limited resources in the Gulf.

The limited resources and increased demands are giving rise to a need for greater understanding of the Gulf ecosystem. This interest in developing our understanding has been recognized in the establishment of the Gulf of Maine Council on the Marine Environment (CME) and the more recent Gulf of Maine Regional Marine Research Board (RMRB) and Regional Association for Research on the Gulf of Maine (RARGOM). All of these entities involve agencies which are keenly interested in promoting and developing management systems that will enhance the rational use of the coastal and marine resources associated with the Gulf.

As in all management systems, there is a certain amount of information that is required to ensure proper planning and implementation in the aim of meeting an objective. This is also the case here for the GOM. The purpose of this workshop, this working group and this paper is to work towards a consensus on an approach for the development of an information management system (IMS) which would organize the data, generate useful information and provide controlled access to a large community of interested agencies.

Definitions:

The basic premise of this paper is that data represents detailed pieces of knowledge that, through the application of tools and methodologies, can be used to generate information. Information being more summary and general in nature. Data is the purview of scientists and technicians who spend significant amounts of their time and effort developing methods for its collection and storage. Information on the other hand is of more general interest and use to scientists, managers and in some cases the public at large.

An information management system (IMS), is a system to organize all aspects of data collection, information generation and dissemination. In its broadest modern sense it is comprised of software, hardware and the people involved. Much is often said about the electronic components of an IMS, i.e., operating platforms, disc space, electronic communication packages etc., however the more important component to any successful IMS is the people or agencies, i.e., their capabilities, needs and interest.

In an IMS there are several types of participants. For the moment, it is convenient to identify two types, the data provider and the client or information user. Typically a single

agency would represent both types. In the past the data provider, most often a scientist, concerns himself with all aspects of information generation including the collection and quality control of the data, design of collection tools and methods of handling the data. All of these activities are driven by his purpose. Commonly this is where the data and information flow starts and ends. The data provider, being the sole information user, knows his needs and sets out to achieve them. The feedback loop is very short and direct.

In more recent times, for a variety of reasons, there is a growing recognition that research systems with single objectives are insufficient to meet present day management challenges. One reason behind this recognition is the limited financial resources available for scientific research. Equally strong arguments for coordinated research and information generation can be made from the point of view of the need to respond more quickly to questions dealing with more complex issues such as cumulative effects and ecosystem stability, as opposed to single event crisis and single fish stock management. The increased complexity of issues which must be addressed usually requires a broad spectrum of information shared among a broad community. An IMS must be able to provide remotely sensed real-time data from NOAA, side by side with historical records of point data collected by community groups. Such a system based on the broadest information base and diverse user input ultimately allows for the best opinion into the data analysis/information generation. It also requires increased functionality over single provider/user system.

At this point we must identify two alternative ways of achieving the necessary functionality: 1) centralized and 2) distributed. A centralized system is a classical one in which a central agency or organization maintains total control of the IMS. This is very analogous to the single provider/client situation. The organization in control identifies its need and addresses them. If it has the mandate of providing the information to a community, it generally stipulates the method of information access and develops all of the necessary audit trails, security etc. to ensure that the system operates properly. Alternatively, the distributed system is made up of a loose federation of the participants. The individual participants allow controlled access to their data directly. In this way the data provider, who is most familiar with the data, is placed in the responsible role of maintaining the data. This allows the individual participants to address their particular needs, while still playing a role in the information community.

In the case of a single provider/client system or a centralized system there are many roles which are successfully carried out implicitly: format standards, access, responsibility, security etc. Experience suggests that one of the problems associated with centralized systems is their inertia and reluctance to be sufficiently outward looking to embrace change. As a result they tend to be very secure systems without the necessary response times often required to address new and urgent problems. In cases where they are organized separately from the broader user community, they are very inefficient at meeting the client users needs. The benefits of the short feedback loop in a single provider/client system is difficult to achieve in centralized system. As a result they often provide information and address questions which are no longer relevant.

A distributed IMS has both advantages and challenges relative to the centralized system. Its main challenge is the need to develop a consensus view among a wide variety of participants. The advantages of such systems are many. The main one is the active participation of both the data provider and the information user in a way that mimics the single provider/user situation. The ability to directly access the data/information, streamlines the process of information dissemination and usage. This provides the quickest access to the best data. Having the information user involved in a distributed system ensures that his needs will be satisfied. His input into the type of information needed will ensure that the system has the necessary functionality, very often this will require the necessary analytical tools to convert the data to useful information. A truly distributed system will minimize useless redundancy by providing users access to the single responsible data source and will guarantee that all agencies are using

the best and most appropriate data source. Overall a distributed system has more advantages and should provide a more robust IMS to meet the needs of research and management.

INTERNET as a Model for the Gulf of Maine IMS:

Once we move away from a classical single provider/user system there are many issues which must be considered explicitly and these issues must be assigned a single responsible agency. In general there are some easily identified roles that can be assigned to both the provider and the user. The provider would most naturally look after quality control and procedural documentation. The user would most naturally look after data products and system development funding, particularly for special dedicated functions. There are several additional functions, which could be assigned to either the provider, the user or a new role having to do with the data and information exchange. Things like data format standards, system access options, communications standards and data dictionary. Such "architecture" is necessary to ensure easy communication among participants. It would provide the basic infrastructure upon which individual agencies could build. These functions should be the focus of this working group.

One of the most useful existing information networks that can be accessed by a large research community is the Internet computer network (see Krol, 1992). This network has both strengths and weaknesses, yet provides a real example of an IMS architecture for consideration in the Gulf of Maine. First the weaknesses. Its greatest limitation at present is its user friendliness. In a computer climate of Macs and Windows based applications, it is difficult to accept the traditional Internet interfaces that are based on text commands. Besides the difficulties in remembering written commands, the variety of keyboards in use, PCs, VT100, etc. make it difficult to develop standard interfaces that would be understood by all users. To some extent this difficulty is being addressed by the development of Gopher and other menu driven interfaces. Applications are also being developed for XWindows, which allows Internet to interpret mouse actions. Such developments would be highly desirable for an IMS that would serve a wide community.

A second difficulty of the Internet is the partitioning of the two main functions between Telnet and the FTP applications. Telnet is the application that allows the user to logon to a remote computer. That is, if the user knows the Internet address, appropriate passwords and the instructions to run the remote computer, he can logon and perform whatever applications have been installed on the remote computer. FTP is a separate application that is used to transfer files from one computer to another. Similar to Telnet, if the user knows the address, password and has proper authorization, he can receive or send files between the two computers. For work on local computers, these two functions are built in without any great amount of switching. The partitioning of these two functions over Internet makes it awkward to carry out more than one function at a time. This would happen if a user wished to log on to a remote computer, view some data, select some data and then receive the data to their local computer. Again, this is not impossible, just difficult. Developments to make such functionality more transparent to the user would ensure that a broad community takes advantage of the system.

In terms of the advantages of an Internet architecture they are many. It is in many ways self regulating. Additional computers are easily added onto the network without a great deal of difficulty. The individual computers can be tailored to the local needs, while fitting into a worldwide network. In many cases, where a dedicated connection to Internet would be prohibitively expensive, there are private connections that will provide Internet services for a small monthly fee. This ability to easily tie into a large existing network would greatly benefit the efforts of the GOM research community. The preeminence of Internet for present communication needs guarantees that any developments/improvements in electronic networking

will provide a mechanism for converting from the existing system. This will ensure the long term viability of any initiative based on the Internet system.

Identification of Responsibilities:

The most difficult task in initiating a distributed system is the identification of responsible agencies for each of the functions. Some of the functions will require extensive development, e.g., user friendly Internet front end, while some have already been carried out, e.g., the Data and Information Committee (DIMC) Directory of Data Bases.

a) Data provider - The data provider would be required to commit to maintaining his data in an agreed format with some form of access. This might include provisions for an Internet connection, routine downloading of data to a central distribution point, or circulation of a data disc at regular intervals. For each high priority data requirement, a single responsible agency should be identified that would ensure continued quality of data. This class of participant must also commit to ensuring that any entry in a data dictionary is up to date. Together with the data exchange facility, the data provider may have to consider security procedures to guard against misuse.

b) Data Exchange - As stated above this function can be performed by the data provider, the user or by a third interested party. It would make recommendations as to the actual architecture in use, i.e., Internet, and establish that this party commit to maintaining the data directory so that all parties can identify existing data. It may also provide the role of centralized access for data that is not/cannot be maintained on line. Functioning as a repository, important data that usually resides on a stand alone PC may be available from this party. As the link between the provider and the user the data exchange center could play an important role in developing and / or disseminating analytical tools to the community at large.

c) Data User - The data user must commit to be responsible in their use of the system and the data in it. They must respect the ownership of the data and through memorandums of understanding (MOU's) should attempt to acknowledge the efforts of the data provider. Users would have a responsibility to share any analytical tools they develop with the IMS community.

Where to go from here:

It is highly recommended that the establishment of an Information Management System for the Gulf of Maine be developed from existing components. At present there are numerous researchers and managers who are collecting and using data from the Gulf of Maine. In doing so they have established methods of acquiring, analyzing and possibly sharing data. It is advisable to identify a number of existing data providers and users, characterize their present operations, in terms of data requirements, technology used for analysis and communication, so as to attempt to integrate existing functions. The Data and Information Committee has made some important steps in the production of the Data Directory and the Users Needs Analysis.

The next steps to be taken require a more detailed look at the participants in the IMS. The IMS community should attempt to identify priority databases that would serve a specific number of users needs. The connection between identified providers and users should be reviewed and tested to see if they meet the need. Actual exchange of data, on a trial basis and possibly on a routine basis, is essential to establish the practicality of the IMS. Such steps may be best taken between parties with a real interest and existing needs. Success at this level will secure additional partners and continued interest.

In the longer term, it would be necessary to broaden the IMS community. This effort would involve both the education of potential participants as well as providing assistance to those considering participation. The most important assistance may be the development of a strong data exchange function within the IMS. This would provide a central point for contact and support.

If the Gulf of Maine IMS can be seen as a "win-win" situation for its participants then it will persist in time. It appears that existing technology can, for the most part, provide the backbone of the system, the desire to cooperate for the benefit of all is perhaps a more important concern.

References:

Krol, E. 1992. *The Whole Internet: User's Guide & Catalog*. O'Reilly & Associates, Inc. Sebastopol, CA. 376 pp.

NATIONAL RMRP IMS

Robert Wall/David Townsend
Regional Marine Research Program Office
University of Maine

Charge: Consider the approach being use by the Gulf of Maine RMR Board as an information management model for other regional boards and consider the role of technology transfer among the boards.

This is a worthy task to begin discussions on. However, of necessity, it will have to be limited to rather general principles and considerations since the GOM model is yet to be defined. And, many of the issue discussions at this workshop will factor into what the GOM model ultimately looks like and what warrants transfer to other regional boards.

Despite the above limitation, there are several general questions which bear on a possible National RMRP IMS that can be productively raised and discussed at the present time. One is simply, what can the other regions gain from the learning exercise we are now going through with regard to information management? Although the answers are not clear, the question suggests that a section of the workshop report be devoted to each of the explicit issues raise, its resolution for our region, the reasoning underlying that resolution, and the extent to which the issue and its resolution might be applicable to other regions.

A second is to assess in terms of a potential National RMRP IMS what modest modifications might be made to our own regional IMS and components thereof. These modifications should be such that they do not compromise the value to us but do enhance their "model" value to the other regions.

A third is the important task of asking ourselves to consider some specific and significant reasons or activities for which a National RMRP IMS might be important and for which it might be an appropriate vehicle. Thoughts that come to mind include standardized data sets and access to them for (1) regional comparative studies of processes, distributions, and trends, (2) larger scale studies that cross RMRP boundaries, (3) augmenting and/or making use of national research and monitoring programs in the coastal regions.

All of the above assume that the National RMR Program will actually develop and flourish for an extended period of time. In addition, fruitful discussions of a National RMRP IMS presumes some significant level of coherence amongst the nine regions in terms of their research plans, objectives and goals. At this time, the long-term future of the RMRP is uncertain. And, based on the several research plans currently available, the level of coherence from region to region appears to be modest.

III. Background Information / Plenary Papers

The Gulf of Maine Regional Marine Research Program: History, Goals and Status

Bob Wall

Nine Regional Marine Research (RMR) Programs were established by Federal law in 1990 for the purposes of:

- (1) setting priorities for regional marine and coastal research in support of efforts to safeguard the water quality and ecosystem health in each region; and
- (2) carrying out such research through grants and improved coordination.

Their establishment recognized the value of the Nation's coastal marine waters, the mounting pressures that threaten their ecological integrity, and the need for regional research to help safeguard their quality and health.

The nine regions have been selected on ecological rather than political or jurisdictional grounds. Each regional program is governed by an 11-member board. Of these, three members are appointed by NOAA, two by EPA, and six by the governors of the states adjoining the region. One of the NOAA appointees is a Sea Grant Director in the region who serves as the Board chair.

Each of the regional programs, within its region, is to

- foster cooperation and coordination in the planning, conduct, and use of the research;
- Take into account the current and projected research needs and existing programs;
- make use of existing infrastructure;
- maintain a long-term perspective within which the research is planned and conducted; and
- ensure high technical quality through the use of peer review from outside of the region.

Each RMR Board is required to develop an RMR plan for its own region which provides :

- a description of the environmental conditions and trends;
- an inventory of the marine research being conducted and planned for the next four years;
- a description of the prioritized research needs over a four year period within a ten-year context;
- plans for coordinating the RMR program with existing research and management activities; and
- a schedule and approach for implementing the planned research.

Each region's plan is to be commented on by interested groups, modified if necessary, and submitted to NOAA and EPA for formal approval. Following this, research proposals based on the approved plan may be solicited, jointly reviewed by the RMR Board and NOAA, and submitted for funding on an annual basis.

These are challenging tasks. They are carried out by a governing body representing a wide variety of scientific and governmental interests. They must take into account the activities of diverse interest groups and organizations whose numbers are great and growing. And, they must plan and conduct their research in the face of uncertainty regarding program funding.

Within the Gulf of Maine (GOM), one of the nine regions, a great deal of regional coordination preceded and helped pave the way for the GOM RMR Program. These included the formation in 1989 of the three-state, two province Gulf of Maine Council on the Marine Environment. Its purpose is to coordinate regional efforts to preserve the environmental health of the Gulf. In addition, a major scientific workshop on the Gulf of Maine was held in early 1991 and involved environmental managers as well as the US and Canadian scientists interested in the Gulf. Finally, a federation of institutions, the Regional Association for Research on the Gulf of Maine (RARGOM), was organized in late 1991 to foster the coordination and development of research on the Gulf.

The GOM RMR Board was formed in early 1992. Its members, their affiliation and the appointing entity for each are shown below.

THE GOM RMR BOARD

11 Members

<u>Name</u>	<u>Affiliation/Title</u>	<u>Appointed by</u>
Ron Baird	VP/Dev., Worcester Polytec	NOAA
Dave Bartlett	Dep Dir. EOS, UNH	NH
Bill Brennan	Commissioner, ME DMR	ME
Matt Liebman	EPA, Boston	EPA
John Nelson	NH Fish and Game	NH
Jack Pearce	Dep Dir. NMFS/NFSC	NOAA
Judy Pederson	MA CZM Office	MA
Don Phelps	EPA Narragansett Lab, RI	EPA
Henri Rauschenbach	MA State Senator	MA
Jim Storer	Chair, ME Marine Res. Bd.	ME
Bob Wall	Dir., ME Sea Grant Program	NOAA

Collectively they are a diverse group with multi-level, multi-agency, multi-issue representation. Dr. Wall serves as the Board's chairperson. Dr. David Townsend of the University of Maine serves the Board as the Executive Director of the GOM RMR Program.

With major support from RARGOM, the Board prepared its Gulf of Maine Research Plan. The plan's development was guided by several overriding principles. The first recognized that our coastal ocean ecosystems are at base the resources we're trying to manage. Simply put, if there were no Gulf of Maine ecosystem, the management issues would disappear.

A second was the realization that the totality of needs for sound scientific information far outstripped any likely program resources. Thus there was a clear need to identify and maintain a program focus that was appropriate to the program purpose, broadly beneficial to the diversity of needs, and reasonably tractable.

A third was to take advantage of the unique opportunity presented by the enabling legislation. It allowed the design of a Gulf-wide, long-term program that (1) complemented the wealth of ongoing but relatively narrowly focused efforts; but that (2) did not assume ownership of their programmatic responsibilities.

In a more general sense the GOM RMR Board recognized that lots of groups are doing things in the Gulf that are both important and urgent, that there are lots of pressures to do things that are urgent but not broadly important, and that it should spend its limited resources on research that is important but not necessarily urgent.

In line with this, the ten-year goal of the Gulf of Maine Research Plan is to work toward a suite of models simulating the Gulf of Maine ecosystem. Ultimately these models would lead to predictive capability of how the Gulf ecosystem functions naturally and under stress.

The research priorities identified in the Plan related to (1) contaminants - where they are, how they got there, and their effects on living marine species of import - and (2) the factors controlling noxious and/or excessive phytoplankton phenomena.

The strategy followed in meeting these priorities while moving toward this goal was to (1) develop a working dynamic model of the Gulf of Maine's circulation and (2) interconnect this model with others to be developed on nutrient budgets and cycling, on food web dynamics, and on pollutant transport, cycling and effects.

The initial research being supported falls into two categories: observational/modeling and analysis/synthesis of existing data. Projects in the former category are:

- **The Relation of Slope Water Inflow to the Eastern Maine Coastal Current and to the Jordan Basin Gyre (Pettigrew, UMaine and Loder, UNH)**
- **An Observational/Modeling Study of Western Gulf of Maine Circulation (Brown, UNH and Lynch, Dartmouth)**
- **Sources, Transport, and Nutrient Environment of Toxic "Red Tide" Populations in the Western Gulf of Maine (Anderson and Geyer, WHOI; Butman and Signell, USGS; Loder, UNH; and Franks UCSD.)**
- **An Observational Study of Gulf of Maine Circulation Variability: An Augmentation Request to the Regional Marine Research Program for the Gulf of Maine (Brown and Loder, UNH; Anderson and Geyer, WHOI; and Pettigrew, UMaine)**
- **Spatial and Temporal Variability of Satellite - Derived Sea Surface Temperature and Pigment Fields in the Northern Gulf of Maine (Bisagni, NMFS; Cornillon, URI)**
- **Atmospheric Nitrogen Inputs to the Gulf of Maine: The Relative Importance of Inorganic and Organic Forms (Talbot and Mosher, UNH)**
- **Cycling of Particle-Reactive Trace Metals and Organic Pollutants in the Gulf of Maine (Gschwend, MIT; Buesseler, WHOI; and Moran, URI)**

From a collective, physical standpoint these projects will characterize the hydrographic and current structure (and their variability) of the Eastern and Western Gulf of Maine Coastal Currents and in the Jordan and Wilkinson Basins along with the forces controlling this structure and its variability. And, by integrating this information with the phytoplankton, nutrient, and pollutant studies, we will begin to quantify exchanges between the coastal regions and the

interior Gulf, the transit times and paths of entrained particles along the coast, and the manner in which these physical dynamics along with the growth and nutrient environment influence the development distribution and fate of toxic dinoflagellates in the western Gulf of Maine.

In the category of analysis/synthesis of existing data, the following projects are being conducted:

- **Microplankton Bloom Dynamics in the Gulf of Maine: Nutrients, Hydrography and Community Structure - A Program to Investigate a Nutrient Trap in the Western Gulf of Maine**, Maureen D. Keller, Chris Garside and Michael Sieracki, Bigelow Laboratory for Ocean Sciences.
- **A Program to Investigate Long-term Seasonal and Interannual Variability in the Optical Properties and Optically-Active Constituents in the Gulf of Maine Waters: Phase 1, Coastal Regions - Database Development and Historical Data Analysis**, Steven G. Ackleson and Charles S. Yentsch, Bigelow Laboratory for Ocean Sciences.
- **Gulf of Maine Contaminated-Sediment Database: Compilation, Desk-Top Access, and Analysis of Existing Data for Scientific and Policy Needs**, Marilyn Buchholtz-ten Brink, Frank T. Manheim, John Hathaway, USGS; Stephen H. Jones and Larry Ward, UNH; Peter Larsen, Bigelow Laboratory for Ocean Sciences; Bruce W. Tripp and John W. Farrington, WHOI; and Gordon T. Wallace, University of Massachusetts at Boston.
- **Preparation of a Review of Uptake and Depuration Kinetics of Algal Toxins by Filter-feeding Molluscan Shellfish**, Sandra E. Shumway, Bigelow Laboratory for Ocean Sciences, and V. Monica Bricelj, State University of New York.
- **Publication of a Phytoplankton Manual for the Gulf of Maine**, Robert A. Andersen, Bigelow Laboratory for Ocean Sciences.
- **Construction of Satellite-Based Wave Climatology for the Gulf of Maine**, V. Panchang, University of Maine.

These several projects when completed will provide the research and user communities with ready access to a great deal of scientific information on important elements of the Gulf of Maine ecosystem.

Collectively all of these projects will in the end contribute to our long-term goal of simulating major components of the Gulf of Maine ecosystem. Striving toward that goal will be scientifically and intellectually challenging and provide many potential benefits. The process itself helps build "community: by conceptually integrating research approaches and disciplines and providing a big picture context for small project science. To enhance success of this complex goal, we will have to come to grips with what are the most significant processes, species and habitats to focus our attention on. and we will have to work at matching a wide-variety of space- and time-scales. To these ends, planning workshops on habitats and ecosystem modeling will be organized for the coming year.

Finally, there is a fundamental need to provide for the effective access and use of important and high quality, scientific data and information concerning the Gulf of Maine. This need, of course, goes well beyond the RMRP program. The present workshop is an attempt to develop an overall context for meeting the need generally and with regard to the GOM RMR program in particular.

The User Community in the Gulf of Maine Region

Maxine G. Schmidt

The Data and Information Management Committee (DIMC) of the Gulf of Maine Council on the Marine Environment was charged with developing an information management system (IMS) for environmental workers in the three states and two provinces bordering the Gulf. To determine the needs and capabilities of the potential user community, the DIMC conducted a user needs analysis. Surveys were distributed to over 700 people, representing each jurisdiction and level of government, as well as nonprofits and trade associations. One hundred twenty surveys were returned; of those 17 were selected for in-depth interviews to provide more detailed information on data use and management. The objectives of the surveys and interviews were (1) to characterize the user groups, (2) to ascertain the kinds of data and information users need, (3) to determine the functional and analytical capabilities of users and their systems, and (4) to discover what tools and applications users will need to make full use of the available data.

Data Users

With reference to the user matrix developed for this workshop by J. Rosen, the five categories of users in the region may be described as follows:

- Administrators - state/provincial agencies, local/regional agencies, federal agencies
- Researchers - mostly from academia and federal governments
- Conservation groups - NGOs and some government affiliates (e.g., public education committees)
- Public and others - private industry (environmental and engineering firms, builders and developers, fishing and aquaculture industries), education (mostly K-12), interested public
- ADP staff professionals - usually part of one of the other four groups

Data Used

Participants were asked to list the kinds of data they use on a regular basis. They were also asked about the data source and maintenance procedures for each data type. The results are summarized in the following table.

Administrators	Researchers	Conservation groups	Public & others	ADP staff
Physiographic and Physical data (GMRD)	Physiographic and Physical data (GMRD)	Physiographic and Physical data (GMR)	Data resources (R)	Physiographic and Physical data (MR)
Biological data (GMRD)	Biological data (GMRD)	Biological data (GMRD)	Habitat information (R)	Biological data (MR)
Cultural data (GMRD)	Data resources (R)	Habitat information (GRM)	Cultural data (GMR)	Cultural data (MR)

The table is the one described by J. Rosen earlier in this report. Each cell in the matrix represents the interactions between a discipline and a class of data. The content of each cell indicates the interaction between the users in the discipline in the columns and the types of data in the rows. G= users are involved in Generating data, R= users are likely to Read (access) data, M= users are involved in Maintaining data, D= users may be involved in Deleting data from data bases.

Data Needs

All of the participants in the survey emphasized the need for more data and information. Many people listed known data sources to which they would like access; others simply produced a "wish list," without knowing if such data exist. Data needs are summarized in the following table.

Administrators	Researchers	Conservation groups	Public & others	ADP staff
Physiographic and Physical data (including remote sensing)	Physiographic and Physical data (including remote sensing)	Physiographic and Physical data	Habitat information	Physiographic and Physical data
Biological data	Biological data	Biological data	Data resources	Biological data
Cultural data	Cultural data	Cultural data		Cultural data
Data resources	Data resources	Data resources		

Data Problems

When asked about problems with data they are currently using, participants cited several broad issues.

- Data access - Most frequent complaint. Many users know of existing data which they would like to use, but do not know how to obtain it.
- Data gaps - Users cited both spatial and temporal gaps in data coverage in the region.
- Data resolution - Because most marine environmental data is collected by the federal governments, they are the main data source for state and provincial environmental agencies. However, the federal governments operate on a regional scale, so the resolution of the data they provide is too coarse (both spatially and temporally) to be useful for workers dealing with a specific site.
- Data quality - Documentation is often missing from the data workers receive, so the utility is a concern.

System Functions

Respondents were also asked about the attributes they would like to see in a regional IMS. Ease of use and low cost were important to all users. Nearly everyone mentioned some

sort of analytical or visualization application. Most often cited were GIS and other imaging tools, as well as statistical and modelling packages and communications software. Many organizations are already using some sort of GIS, and expressed concern about exchange between systems. Participants approved of the concept of the Axes/MacLaren Plansearch application currently under development.

Another important concern articulated by the respondents was operating system compatibility. Many users in the public group operate with Macintoshes, while others use PCs running both DOS and Windows applications. A few agencies and many researchers have workstations running Unix or an equivalent, usually dedicated to a particular function, such as GIS or modelling.

Other Issues

Several additional issues for state and provincial agencies emerged from the user needs analysis, most having to do with financial restrictions. Financial problems are predicted for the long term, and may be the biggest obstacle to participating in a regional IMS for many potential users. Funding of state and provincial agencies has been severely reduced over the past several years, affecting both personnel and equipment. Most agencies surveyed are equipped with 386 PCs, and, if they have data communications capabilities at all, it is usually some kind of interoffice e-mail (though some offices can communicate with outside entities via modem). U.S. state agencies in the region do not have access to Internet at this time. Some Canadian provincial agencies have access through regional commercial systems.

Most organizations claimed some skills using fairly sophisticated applications such as GIS or relational databases, but most use spreadsheets and word-processing programs for most of their work. Some training will be needed to enable these users to exploit any dedicated applications and to navigate through unfamiliar systems.

Bottom Line

The potential user community in the Gulf of Maine has long recognized the need for data transfer and communication. Most users would be satisfied if they could merely obtain relevant data from other organizations. If this exchange is to be accomplished electronically, the IMS which serves this group will have to address these issues:

- Finances are a constraint on the acquisition of equipment and expertise. Smaller, PC-based systems should therefore be considered when designing the system.
- A variety of platforms exist in the region.
- Access to Internet is still a problem for many potential users. It may be necessary to provide access to the IMS for modems.
- This is a very diverse group. This diversity is reflected in all aspects of data management. In order to accommodate the strongest needs of such disparate users, a regional IMS may have to be confined, at least initially, to basic data transfer functions.

Summary

The User Needs Analysis conducted by the Data and Information Management Committee of the Gulf of Maine Council on the Marine Environment provided basic information about the data-related activities of many organizations throughout the region. This information enabled the DIMC to characterize functional user groups, to ascertain the kinds of data, information and applications users need to meet the goals of the Gulf of Maine Initiative, and to determine the functional and analytical capabilities of users and their systems.

The five functional groups of potential users in the Gulf of Maine region, though their activities are very different, cited many of the same concerns. The kinds of data and information needs are similar, and data functionality, in terms of resolution and transfer across platforms and applications, is a common issue. Financial considerations are important, and will be for some time to come, especially for state and provincial agencies. This affects both access and processing capabilities for those users.

All participants in the survey expressed a need for wider data access, and enthusiasm for a regional data directory. Many suggested datasets, both electronic and hardcopy, they would like to see included.

Academic Research Systems

A Distributed, Object-based Data Management System for JGOFS

Glenn Flierl¹, James Bishop², David Glover³, Satish Paranjpe²

¹MIT, ²LDGO, ³WHOI

Large oceanographic programs such as JGOFS (The Joint Global Ocean Flux Study) require data management systems which enable the exchange and synthesis of extremely diverse and widely spread data sets. We have developed a distributed, object-based data management system for multidisciplinary, multi-institutional programs. It provides the capability for all JGOFS scientists to work with the data without regard for the storage format or for the actual location where the data resides. The approach used yields a powerful and extensible system (in the sense that data manipulation operations are not predefined) for managing and working with data from large scale, on-going field experiments.

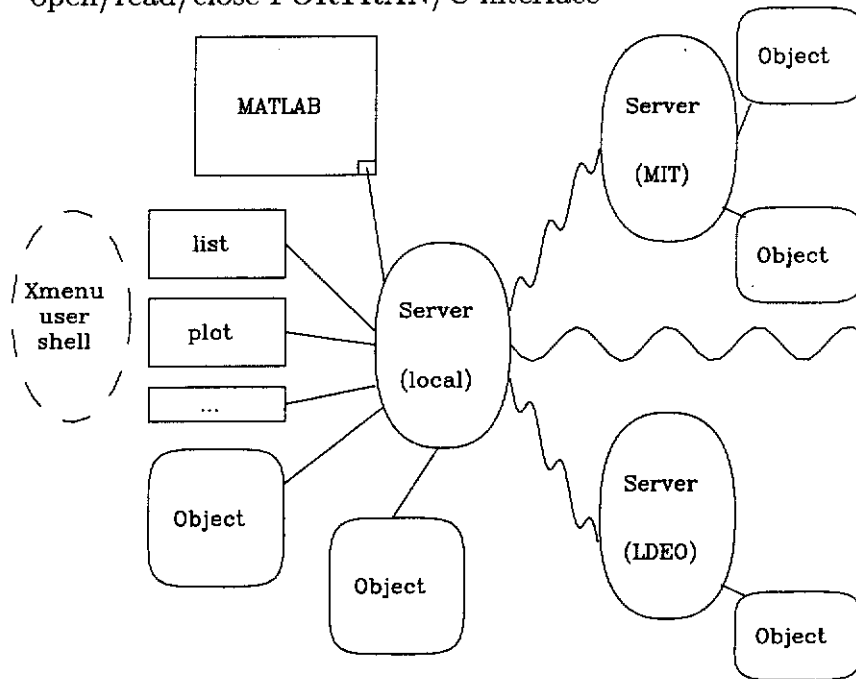
In the "object-based" system, user programs obtain data by communicating with a program (the "method") which can interpret the particular data base. Since the communication protocol is standard and can be passed over a network, user programs can obtain data from any data object anywhere in the system. Data base operations and data transformations are handled by methods which read from one or more data objects, process that information, and write to the user program.

Purpose:

- Permit scientists to use data without concern for storage technique, location, or format
- Networked interchange of data sets
- Access to most recent versions of data sets during experiments
- Handle multidimensional data
- Transmit metadata
- Extensible data manipulation routines
- Usable interactively or from programs

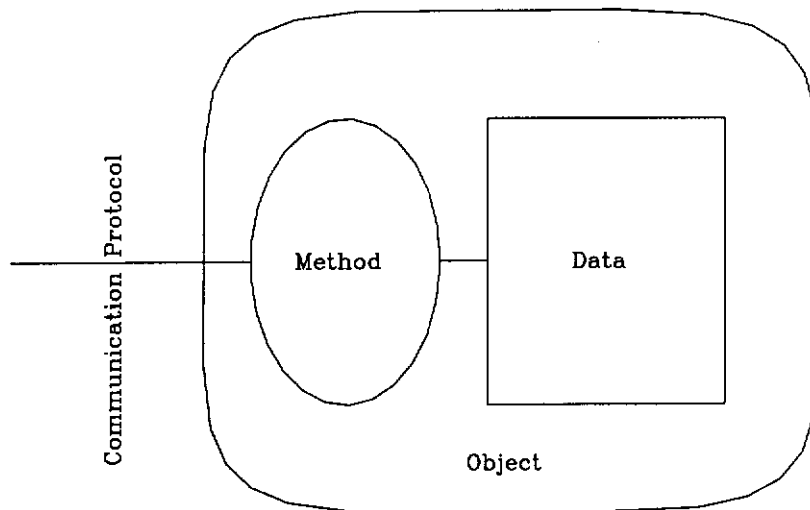
The basic elements of the system are:

- a server which routes requests
- data objects which receive requests and respond with data
- interfaces to other software:
 - data can be imported directly into packages such as MATLAB
 - simple listing and plotting programs (supplied)
 - open/read/close FORTRAN/C interface

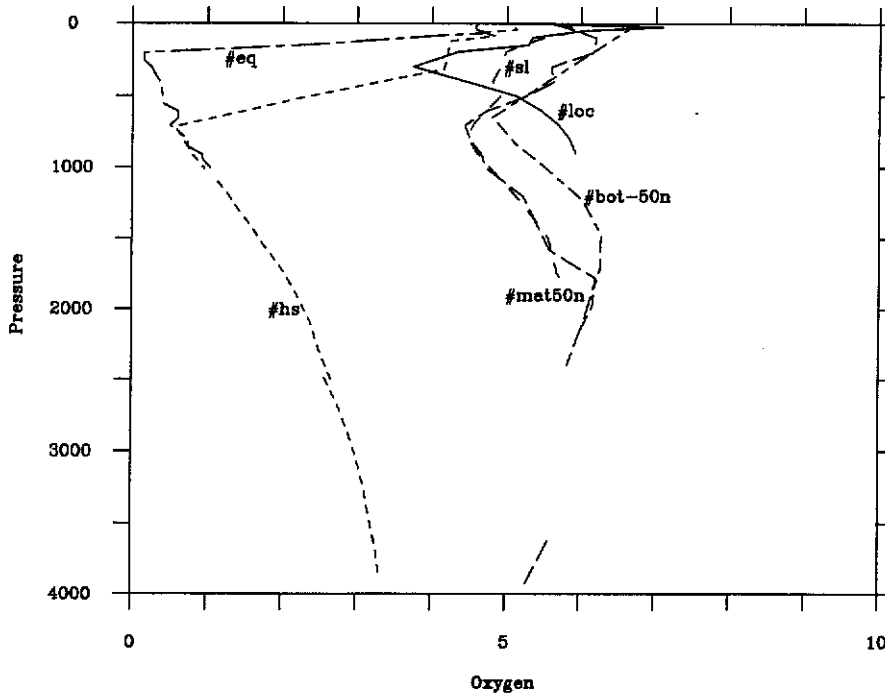


Data Objects package together a program (the method) and data. User programs never look at the data directly; rather, they communicate with the data object

Objects couple data & programs(methods)



- 1) Data Objects communicate with a common protocol
 → Programs can work with any data in the system
 As an example, we show a figure with plots from 6 different data objects



The commands which created the plot are

```

window 0 4000 10 0
axis x 2 Oxygen 5 dd
axis y 500 Pressure 1000 dddd
dash "#loc" o2 press -----
dash "#sl" o2 press ----.....
dash "#mat50n" o2 press -----.....
dash "#hs" o2 press --.....
dash "#bot-50n" o2 press -----.....
dash "#eq" o2 press -----.....

```

This shows a single program `dash` requesting "o2" and "press" data from each object and displaying it with the specified dash pattern. The objects are

object	location	machine type	storage method
#loc	MIT	Sun	flat ASCII file
#sl	RSMAS	Alpha	scaled binary integers
#mat50n	MIT	Sun	MATLAB binary floating point
#hs	WHOI	Sun	multiple files/directories
#bot-50n	MIT	Sun	multiple files
#eq	U. Chicago	IRIX	Single file with multiple stations

The communications protocol includes:

- Comments (text)
- Variable descriptions
 - Name
 - Dimensions for vectors/ matrices/ tensors
 - Attributes (e.g., units)
 - Hierarchical structure
- Data
 - Strings or numbers

The protocol allows hierarchical structuring

```
leg
year
month
  station
  lat
  lon
    press
    temp
    sal
    o2
    sigth
```

As an example, we show a partial listing of a data object

```
# Wunsch stations 3-10
# p<1000
=====
leg,   year,  month
.....
   1,   81,   6
=====
station, lat, lon
.....
   3, 38.28, -73.53
=====
press,  temp,  sal,   o2,  sigth
-----
 5.000, 18.334, 33.570, 5.970, 24.096
25.000, 12.848, 34.159, 6.990, 25.773
49.000, 11.070, 34.523, 6.060, 26.394
99.000, 11.093, 35.090, 5.340, 26.831
149.000, 11.906, 35.487, 5.020, 26.990
199.000, 10.819, 35.435, 4.210, 27.152
300.000,  8.293, 35.126, 3.730, 27.334
```

400.000, 6.363, 35.046, 4.640, 27.546
500.000, 5.724, 35.019, 4.980, 27.608
600.000, 5.031, 34.990, 5.460, 27.670
701.000, 4.633, 34.981, 5.680, 27.710

=====
station, lat, lon
.....
4, 38.19, -73.52

=====
press, temp, sal, o2, sigth

5.000, 17.516, 33.160, 5.840, 23.981
25.000, 12.315, 33.958, 7.090, 25.721
49.000, 9.612, 34.192, 6.020, 26.387
99.000, 12.095, 35.402, 5.340, 26.887
149.000, 12.407, 35.625, 5.290, 27.000
199.000, 11.287, 35.487, 4.340, 27.108
300.000, 8.059, 35.120, 3.750, 27.365
400.000, 6.606, 35.053, 4.440, 27.520
500.000, 5.467, 34.997, 5.130, 27.622
600.000, 5.003, 34.983, 5.460, 27.667
701.000, 4.680, 34.979, 5.700, 27.702

=====
station, lat, lon
.....
5, 38.16, -73.26

=====
press, temp, sal, o2, sigth

5.000, 18.382, 33.647, 5.770, 24.143
25.000, 12.040, 34.196, 6.660, 25.959
49.000, 11.951, 34.925, 5.510, 26.543
99.000, 11.914, 35.390, 5.100, 26.912
149.000, 12.045, 35.547, 5.070, 27.010
199.000, 11.976, 35.589, 4.940, 27.057
300.000, 9.425, 35.250, 3.620, 27.251
400.000, 7.003, 35.075, 4.210, 27.483
500.000, 5.827, 35.009, 4.910, 27.589
600.000, 5.252, 34.988, 5.300, 27.643
701.000, 4.845, 34.980, 5.610, 27.684

***** End of object ***

2) Data Objects handle

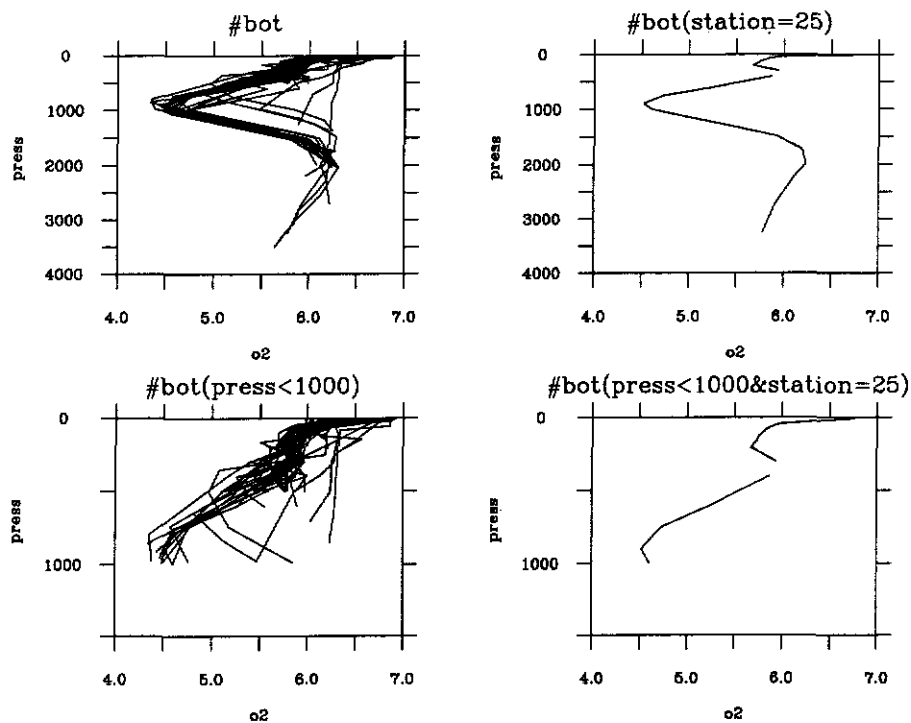
Projection (subsetting by variable name)

Selection (subsetting by variable values)

→ Can minimize transmission of data

→ Individual objects may have other functions

These plots show the results of "selections" on a data object



3) Server works with dictionaries (user's and system's) of object names

→ Names are brief; often dictionary names point to object on another machine

Names of data objects: In order to access a data object, you need to know its *full name*. In its most general form, this includes the location, the method name, and parameters for the method. Often, parts can be abbreviated.

In the dictionary the #bot object is

```
#bot = 18.83.0.11::#bot
```

home machine name on that machine

On 18.83.0.11, the dictionary has

```
#bot=/d5/glenn/bloom/jgbl2("/d5/glenn/bloom/bot")
```

method program required parameters

A request such as #bot(station=25,press,o2) translates effectively into
18.83.0.11::/d5/glenn/bloom/jgbl2(/d5/glenn/bloom/bot,station=25,press,o2)

home method parameters

Database Manipulations: “Constructed Objects”

These are objects which behave much like filters, taking the output from one or more objects, transforming it and passing it to the parent.

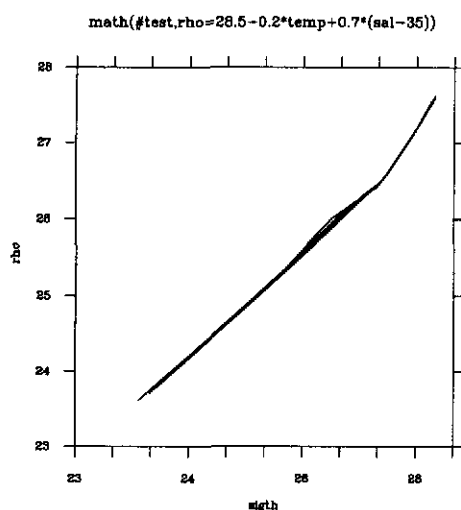
For example, we can add a column to the #test (hydrographic data) which gives a linearized estimate of density:

$$\rho = 28.5 - 0.2T + 0.7(S - 35)$$

by using the “math” constructed object which takes as parameters an input object name and formulae for changing/ adding columns. The new object name is

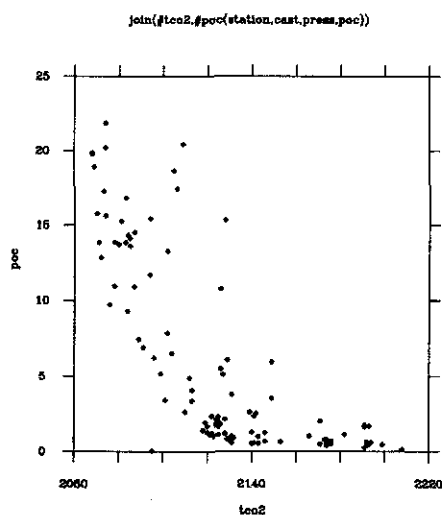
```
math(#test,rho=28.5-0.2*temp+0.7*(sal-35))
```

and this can be used by the lister/plotter/... in exactly the same way as any other object — see figure.



As another example, there is a plot from two data objects joined together by common station, cast, and pressure.

```
join(#tco2,#poc(station,cast,press,poc))
```



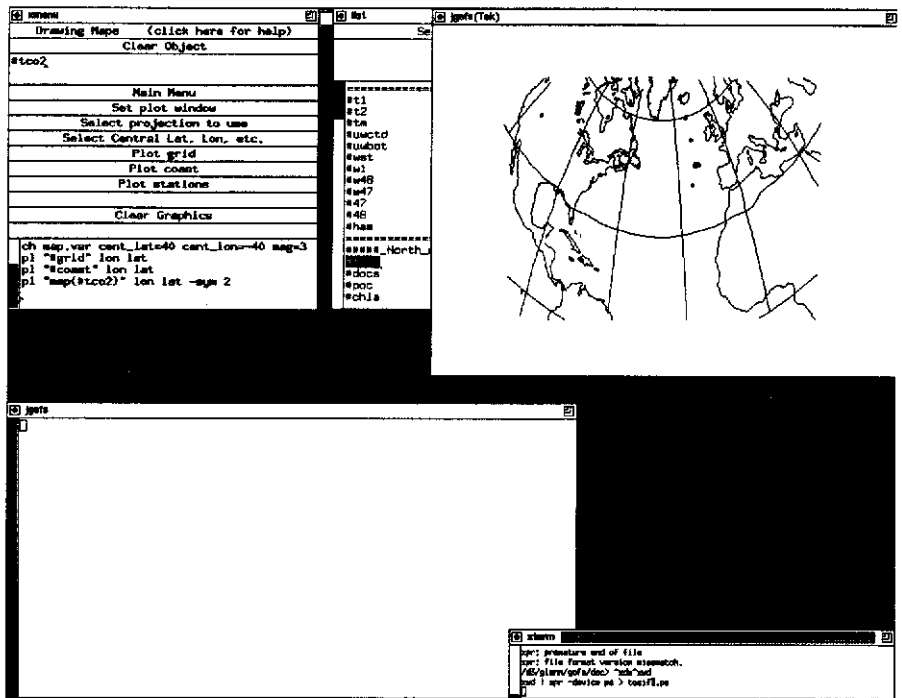
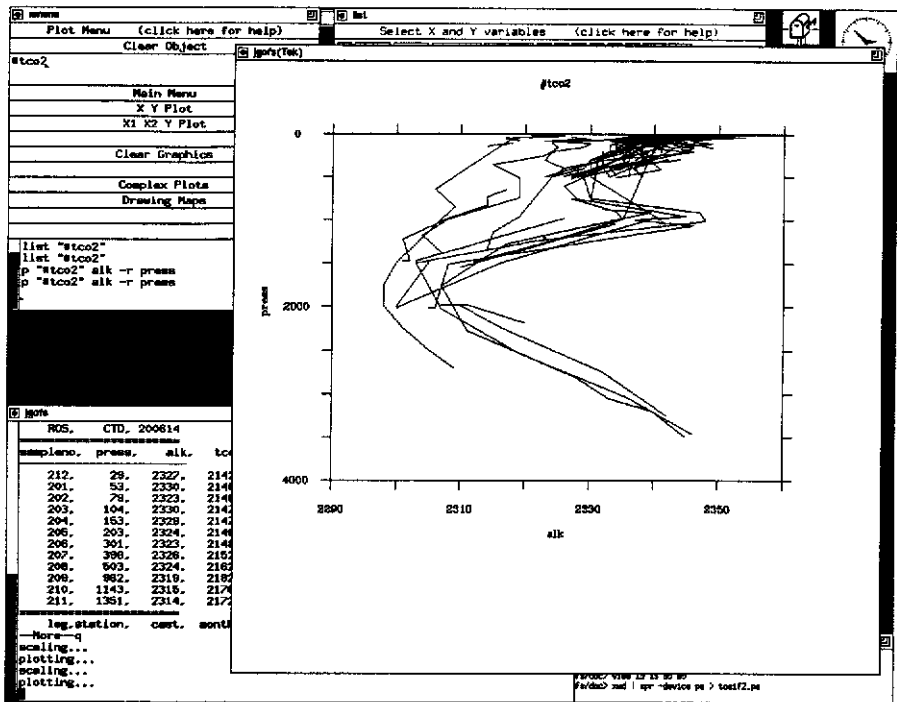
User Interfaces:

The fundamental aspect of the system is the technique for accessing data in many different formats from many different sites. Although we have developed user interfaces for our own work and make those available with the software, the "system" is *not* the interface or plotting package. Below are shown three views of a workstation screen while using our menu interface. On the next figure are two examples using MATLAB with a compiled function, "rjg," which reads data from the JGOFS system into vectors or matrices.

The screenshot displays three windows from a workstation interface:

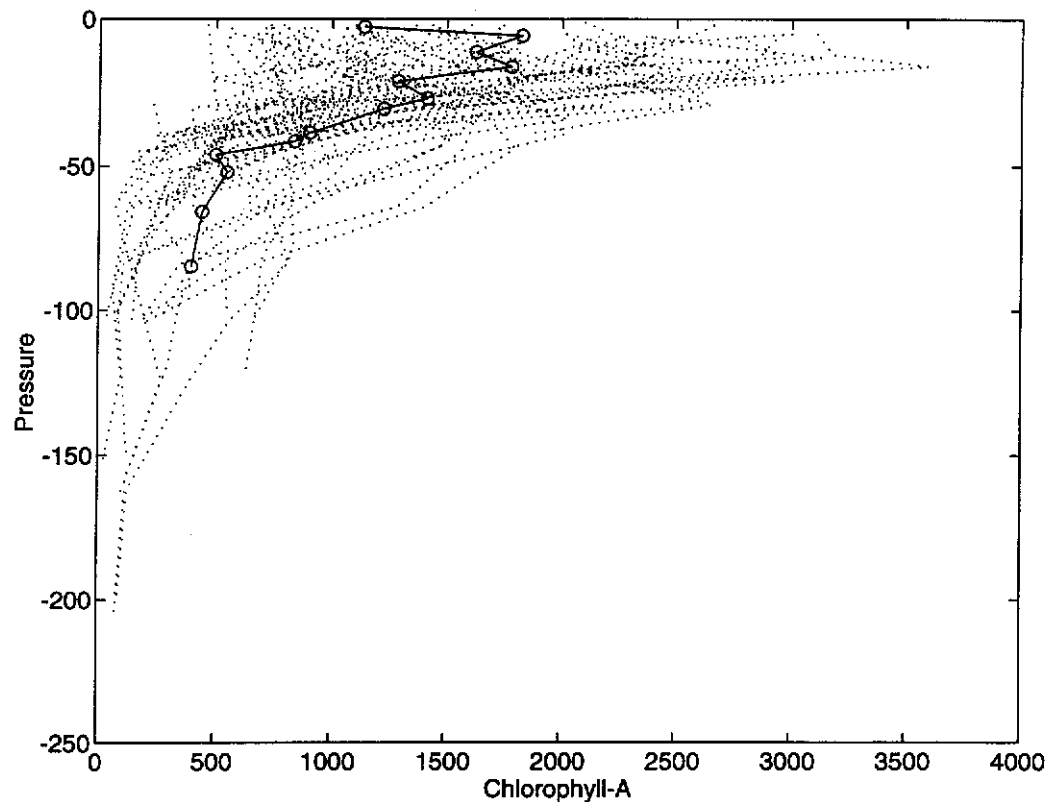
- JGOFS Data System (click here for help):** A menu interface with options: Clear Object, Exit, Select object from dictionary, List Variables, Plotting, Other List Options, and Transformations of Data. Below the menu, it says "Welcome to JGOFS" and "List 'tca2'".
- Select Object (click here for help):** A list of objects:
 - ===== user.dct =====
 - #t1
 - #t2
 - #ta
 - #uactd
 - #uabot
 - #ust
 - #w1
 - #w4b
 - #w47
 - #w7
 - #w8
 - #wan
 - ===== gofs.dct =====
 - #####_NorthAtlantic_Bloom
 - #####
 - #doc3 Carbon_dioxide
 - #doc4 Carbon_Dioxide
 - #doc5 Particulate_C_N
 - #chia Chlorophyll_a
- jgfs:** A data table with columns: leg, station, cast, month, day, year, lat, lon, time, temp, pres, sal, tca2, doc.

leg	station	cast	month	day	year	lat	lon	time	temp	pres	sal	tca2	doc
2	4	2	04	20	1989	59.763	-20.643	0656	2008				
ROS	CID	200814											
212	29	2327	2142	nd									
201	53	2330	2146	nd									
202	79	2323	2146	nd									
203	104	2330	2147	nd									
204	153	2329	2147	nd									
206	203	2324	2146	nd									
206	301	2323	2148	nd									
207	398	2328	2182	nd									
206	503	2324	2182	nd									
209	982	2318	2182	nd									
210	1143	2315	2176	nd									
211	1351	2314	2172	nd									



MATLAB program:

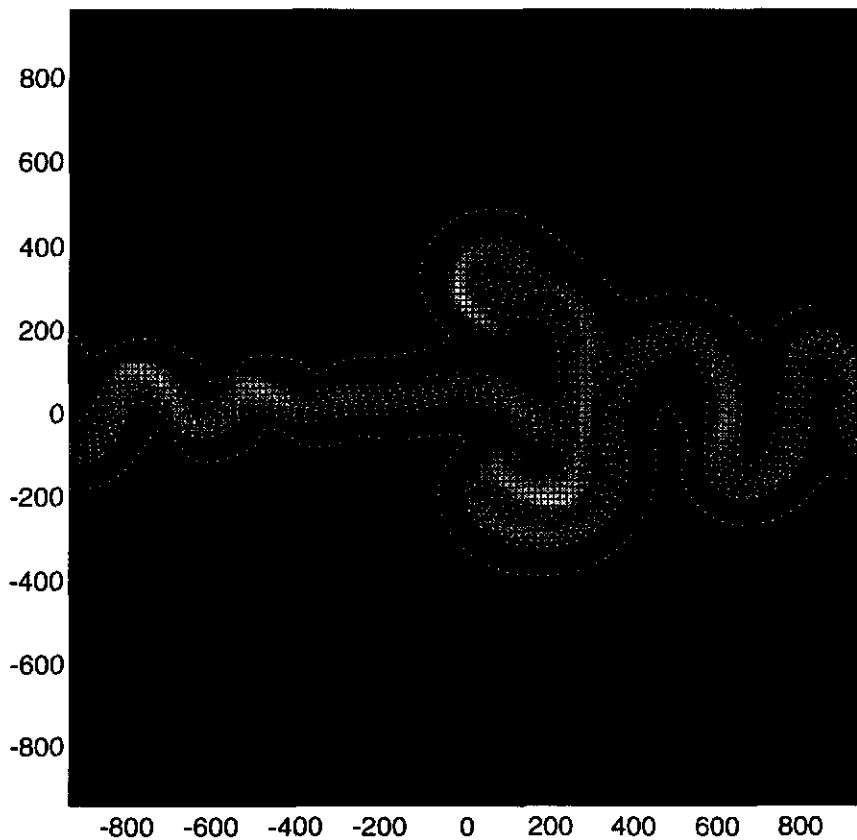
```
%  
% Compute bin-averaged chlorophyll-a profile  
%  
% read in data from JGOFS data system - first, event numbers  
e=rjg('#chla(eventno)');  
% now get eventno,press,chla triplets  
[ee,p,c]=rjg('#chla(eventno,press,chla)');  
% bin-average the press and chlorophyll-A  
pav=binav([0:5:50,60:20:100],p,p);  
cav=binav([0:5:50,60:20:100],p,c);  
% plot the averages  
plot(cav,-pav,cav,-pav,'o');  
axis([0 4000 -250 0]);  
xlabel('Chlorophyll-A');  
ylabel('Pressure');  
hold;  
% plot the individual profiles  
for ex=e'  
    plot(c(ee==ex),-p(ee==ex),'');  
end  
hold;
```



This figure shows the output from a numerical model of the physics and biology of a meandering Gulf Stream. The thermocline depth is contoured. The data resides as floating point binary, direct access records. The MATLAB interface was used to import the data and make the plots.

```
[p1,p2,p]=rjg('#qg(yrday=20,psi1,psi2,phyto)');  
p1=reshape(p1,128,128)';  
p2=reshape(p2,128,128)';  
p=reshape(p,128,128)';  
x=15*[-63:64];y=15*[-63:64]';  
[x0,y0]=meshdom(x,y);  
colormap(gray);  
pcolor(x,y,p);  
hold;  
contour(x,y,p1+4.22917*y0-p2,':k');  
axis('square');  
hold;  
title('Phytoplankton day 20');
```

Phytoplankton day 20



ABSTRACT

xbrowse is one component in a distributed digital data browsing and retrieval system for oceanographers. The system consists of several data servers which provide access to different types of data (e.g., AVHRR imagery and XBT in situ data). Client programs (such as *xbrowse*) have been developed to access these servers. Because the client programs can access several servers at one time and can integrate the data that they receive from those servers, the system facilitates interoperability.

XBrowse, Video Browser, In situ Retrieval systems

Peter Cornillon

University of Rhode Island

XBROWSE

xbrowse is a tool for high speed browsing
of digital imagery

xbrowse is a tool for acquiring digital data

xbrowse is based on the client/server
model

xbrowse uses progressive transmission to
achieve very high compression ratios

xbrowse provides both lossy and lossless
data transmission

The Data:

Primary Data Sets

Satellite

Advanced Very High Resolution Radiometer (AVHRR)

Produces: Sea Surface Temperature (SST)

In Situ

**eXpendable BathyThermograph (XBT), a Free Fall
Temperature Probe**

Produces: Vertical Ocean Temperature Profile

June 12
1988

NOVA
SCOTIA

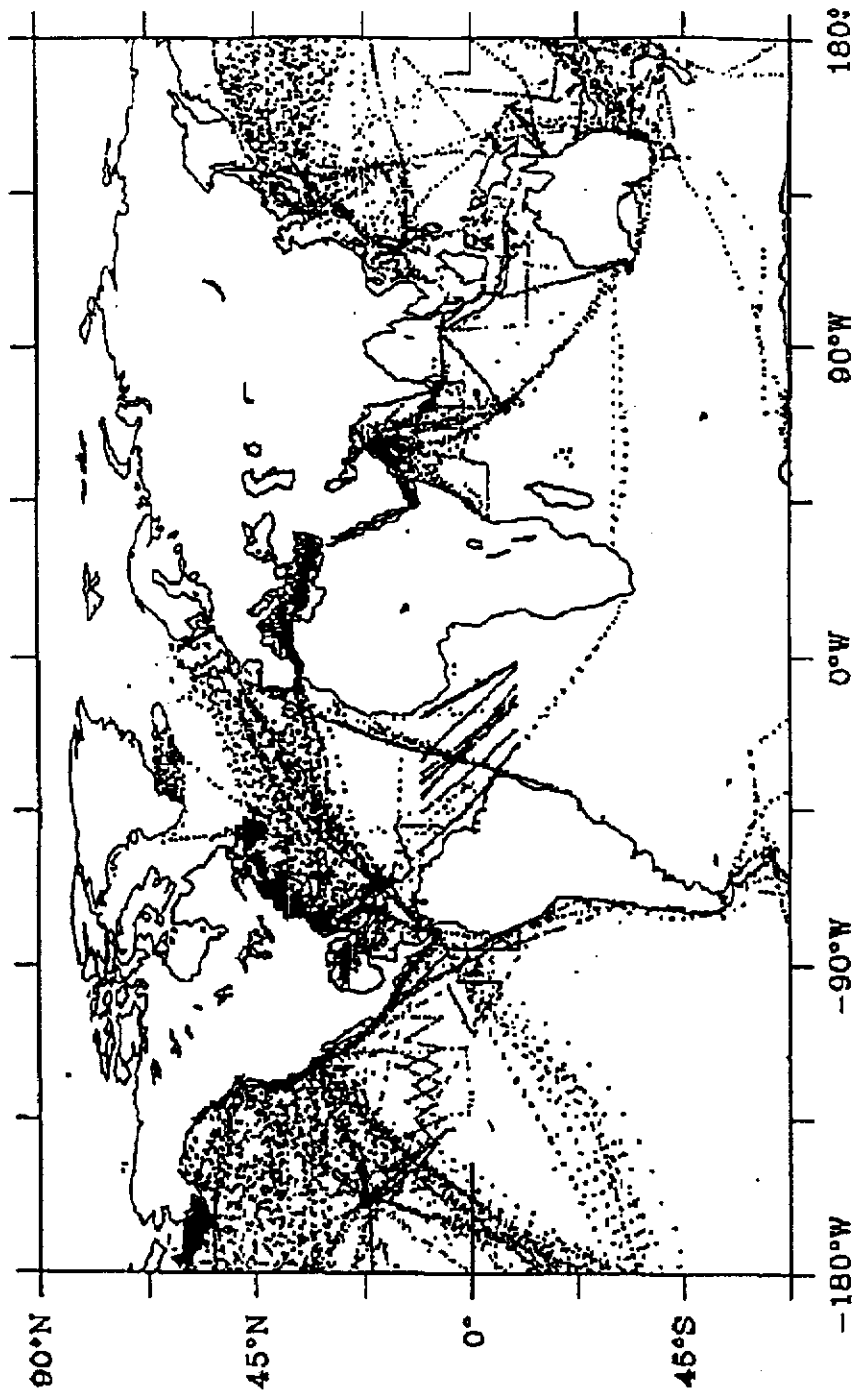
Warm
Core
Ring

Warm
Core
Ring

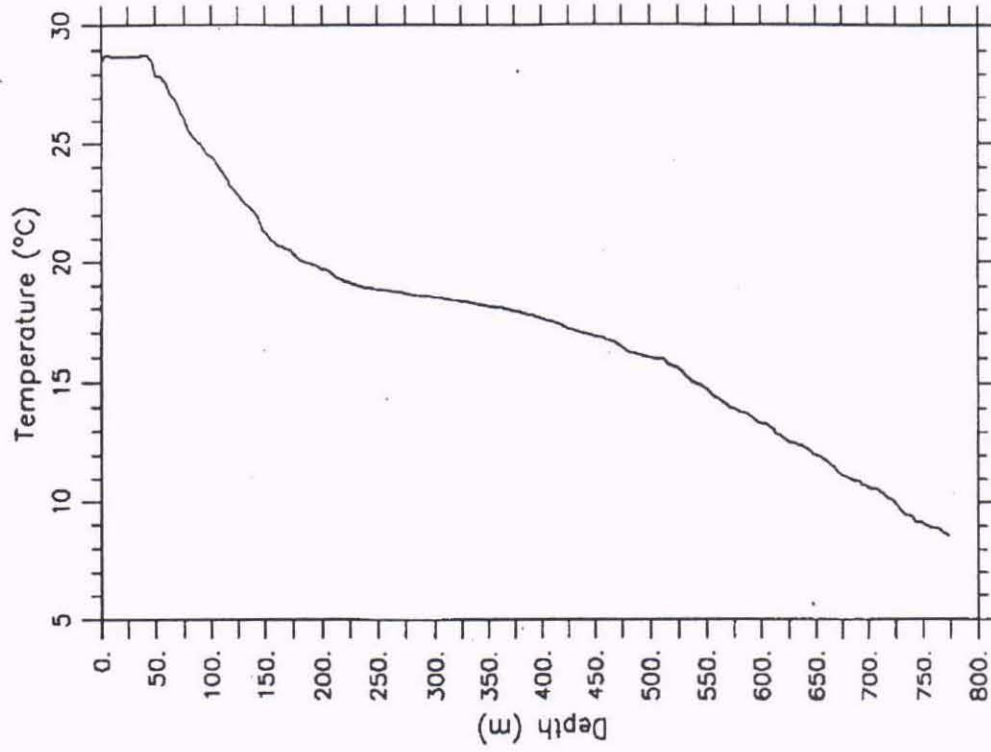
eam

Sargasso
Sea

The Data: XBT Coverage 1982



The Data: XBT Profile



Data Management Issues Related to Scientific Research Requirements

<u>Issue</u>	<u>System Components</u>
Locating Data	AVHRR Inventory
Previewing Data	Video Browse xbrowse
Efficient Access to Imagery from Remote Site	xbrowse
Integration & Correlation of Distributed Data Sets	AVHRR Inventory xbrowse Video browsing

SOLUTION

A Distributed System

Client/Server model allows distributed access

Progressive transmission can reduce bandwidth requirements

CLIENT/SERVER SYSTEMS

Overview

Server provides access to system resources

Client provides interface to remote server

Server and client are separate programs

**Server and client often run on different machines connected
via a network**

**A protocol defines the exchange of information between
server and client**

CLIENT/SERVER SYSTEMS

Interoperability Across Distributed Heterogeneous Systems

The server's protocol describes the resource

A client requests specific resources from server(s) using their protocol(s)

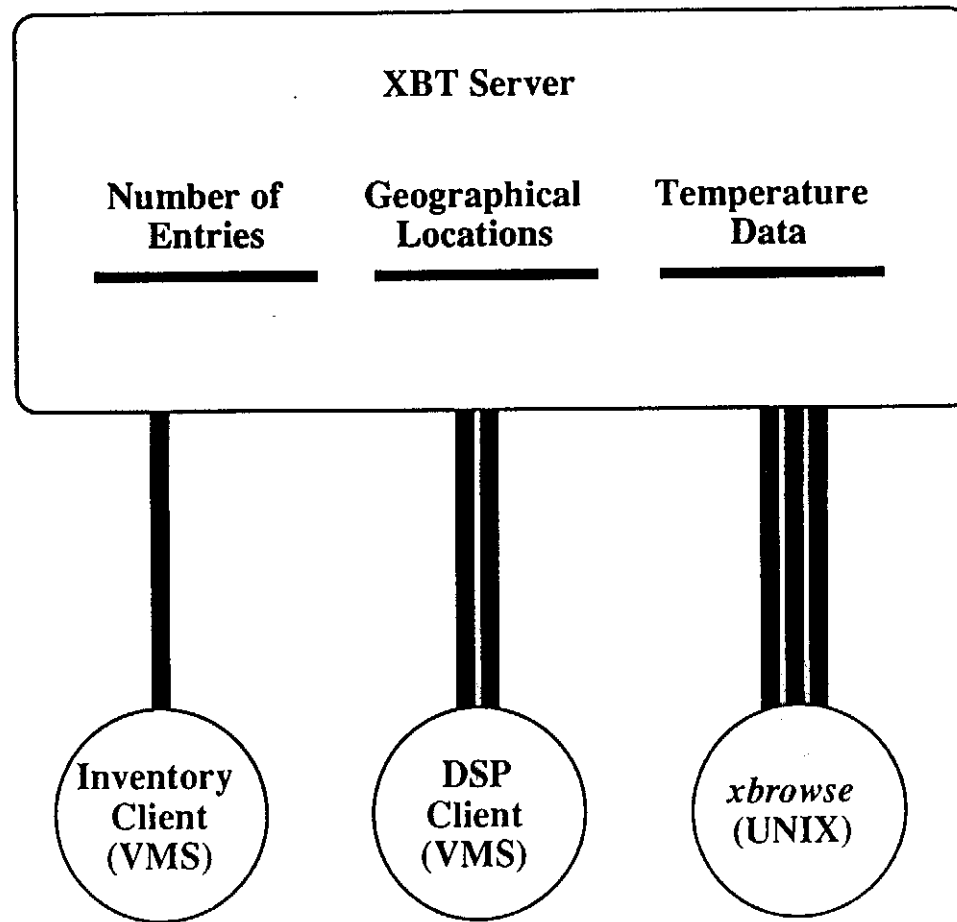
A single client program can access several servers simultaneously

A single server program can provide resources to several clients simultaneously

The client program maps data from servers into a common format (result integration)

MULTIPLE CLIENTS CAN CONNECT TO A SINGLE SERVER

Each Client can Request Different Levels of Information



CLIENT/SERVER SYSTEMS

***xbrowse* Implementation**

Access distributed resources (e.g., AVHRR, XBT)

Two separate servers

Modular

Servers are specialized for their resources

Client is specialized for its application

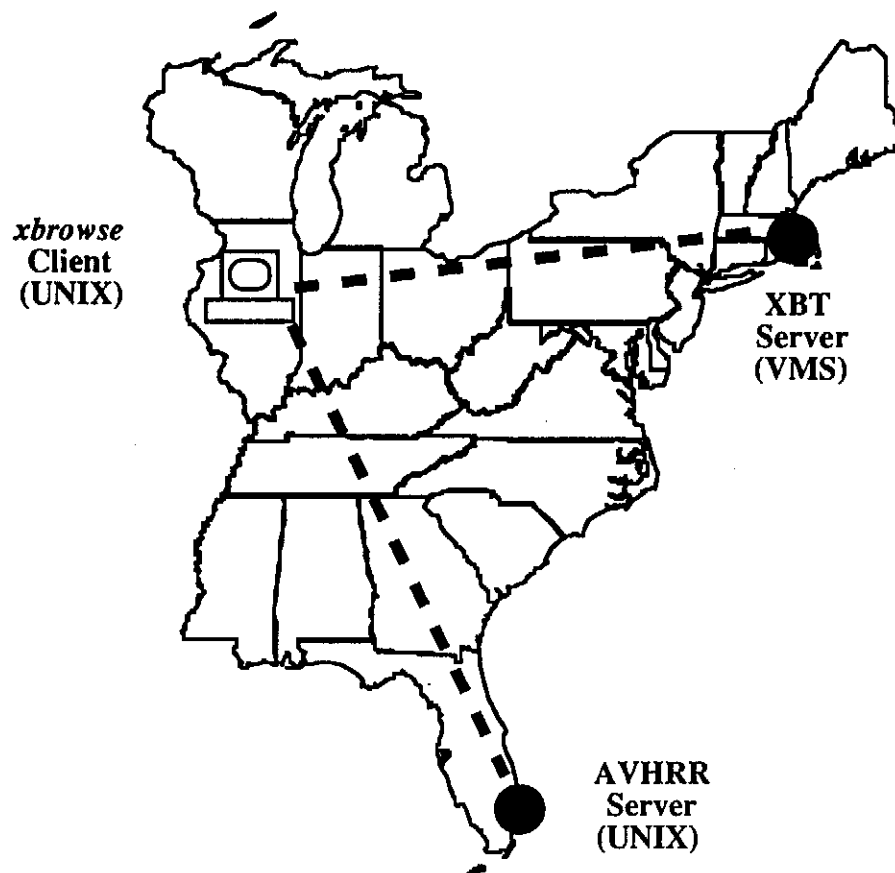
Facilitates interoperability

Multiple data sets are cross correlated

Multiple data sets are integrated in one visual display

SYSTEM OVERVIEW

Client & Server Locations for this Demonstration



PROGRESSIVE TRANSMISSION

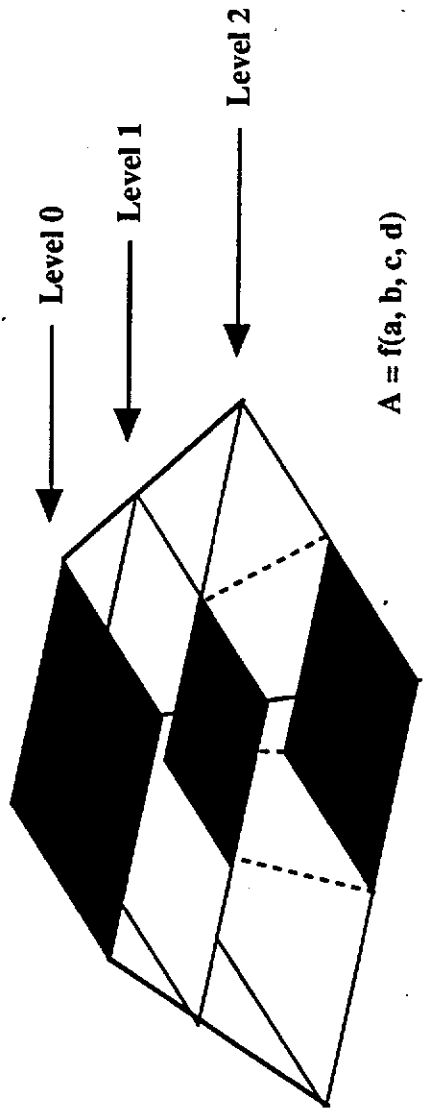
One Way to Browse Imagery

Based on the image pyramid

**Sampling can be used to produce progressive
levels of resolution**

Input can be any raster file

THE IMAGE PYRAMID



$$A = f(a, b, c, d)$$

PROGRESSIVE TRANSMISSION Compression

Level	Bytes	Σ Bytes	Comp.	Time (s)*	Σ Time (s)*
4	256	256	1024	< 1	< 1
5	768	1,024	256	< 1	< 1
6	3,072	4,096	64	< 1	1
7	12,288	16,384	16	3	4
8	49,152	65,536	4	12	16
9	196,608	262,144	1	48	64

* Times based on transmission rate of 4 Kbytes/second

PROGRESSIVE TRANSMISSION

***xbrowse* Implementation**

Gives user control of resolution

**Sampling technique enables each piece of information
to be transmitted only once**

**Image source can be either 'browse products' or real
data**

BROWSING

A Classification Problem

Many images can be classified at low resolution levels

80% of images categorized as cloudy/clear

**Remaining images require near-full to full resolution
inspection for classification**

Much of the image may be superfluous

FEATURES

Display digital AVHRR imagery using progressive transmission

Allow user to select Region of Interest

Integrate display of imagery and *in situ* data

Session logging

Save image information to a file for processing

XBROWSE

The screenshot displays the XBROWSE interface. At the top, there is a large, pixelated image of a person's face. Below the image is a control panel with the following elements:

- Status Indicators:**
 - 85711000-00.png
 - Resolved March 9
 - File Check
 - Aug 26 5:07 (Wed)
 - June - 81.803 (640)
 - 21.4.879 (C)
- Control Buttons:**
 - Open
 - Log
 - Print
 - Define Region
 - File CTS Data
 - Auto Scale
 - Quit
 - Book
 - Home
 - Use Region
 - Plot XST Data
 - Zoom

XGB, a Geographical Database Browser

by
John D. Unger
U. S. Geological Survey
Reston, VA

XGB (aka geobase) is an X-Windows, geographical database browser program that serves as a user-friendly front end to a sophisticated, easy-to-use distributed database system designed by researchers at Lamont-Doherty Earth Observatory. The XGB program provides a geographically-oriented interface to various "viewservers" that provide read-only access through the internet to the different databases that can reside on the computers anywhere on this network (see Appendix I for a complete description of this system). The Lamont distributed database system allows users to store their own data locally using whatever database programs or data structures they require. A viewserver program is then used to make a "view" or table structure of that database available to anyone on the internet who is using XGB or other program that is designed to interface with the viewserver protocol. The viewserver program provides a common "view" or interface for whatever database is used to store the data locally. Currently, viewserver programs have been written for Oracle, Ingress, Unify, dbVista, Sybase, standard SQL, SDM (a public domain Lamont database program), and for simple raster formatted data. This system allows the owner of the data to have complete control of how his data are made available to others using programs like XGB over the network. He can control which fields of a database are part of each "view", and any changes that he makes in the database locally are immediately reflected in the view of the database made available by the viewserver.

XGB is designed to be used with geographical data, and all the data retrieved by the program must have latitude and longitude fields, which are used as indices to the various databases as they are retrieved. Each database is displayed graphically in the geographic region of interest as it is retrieved; the database can also be viewed record-by-record. Each database can be filtered and/or modified by a set of boolean and mathematical operations within XGB. Two databases, a relatively crude coastline and country boundary map of the world and a global raster database of topography and bathymetry at 5'x5' spacing are part of the XGB program.

The license to use XGB, the viewservers, and all associated programs is granted free of charge to nonprofit educational institutions and government agencies. The programs are available via anonymous ftp from lamont.lidgo.columbia.edu (129.236.10.30); appendix 1 contains details of which files are required.

A Voluntary, Public Method for Sharing Earth Science Data over the Internet using the Lamont View-Server System

William Menke*, Paul Friberg, Arthur Lerner-Lam, David Simpson,
Robert Bookbinder and Garry Karner
Lamont-Doherty Geological Observatory of Columbia University
*also Dept. of Geological Science, Columbia University
* Internet E-Mail: menke@lamont.lidgo.columbia.edu

Published in EOS, September 17, 1991

Abstract. We envision a time when an earth scientist from any institution will be able to access high quality data, regardless of where this data may actually be archived. The scientist will not have to know anything about where the data actually resides or what format it is stored in. He or she will only have to specify the type of data and the workstation software will handle the rest of the details of finding them and accessing them. A method - now in use at Lamont and several other institutions - of achieving this vision is described in this paper. Institutions make 'views' of their databases publicly available to users of the Internet, using database serving software that runs on one of their computers. This software completely automates the process of finding out what kind of data are available and of retrieving them. Hence a wide variety of different databases become, from the scientist's perspective, parts of an Internet-wide earth science database. At the same time, the institution that archives the data is not locked into any particular database management system, and is free to provide alternative access methods and to exclude some of its data from the system. The self-teaching nature of this concept is especially useful in large cooperative scientific efforts.

CRITICAL NEED FOR DATA ACCESS AND SOFTWARE TOOLS

In order to make scientific advances in understanding a given region of the earth, an earth scientist may need to study and compare many different data types, having very different structures and data volumes. For instance, in a geological study of a volcanic region, a scientist may need to examine topographic data in order to locate the major volcanic edifices, gravity anomaly data in order to make inferences about crustal thickness, seismicity data in order to locate active geologic faults and regions of magma movement, satellite photos in order to map recent lava flows, and chemical analyses of rocks in order to understand the magma generation process. Furthermore, a large number of subsidiary databases (such as magnetic anomaly, heat flow measurements, well temperatures, meteorological reports, etc.) may play a role in resolving a particular problem.

Research in such integrated studies can be greatly facilitated by a carefully designed database tool. Such a tool must possess certain properties. First, it must be capable of teaching the user about what data are available. Too often in the past a researcher has overlooked a source of data that might further his or her research out of ignorance of its existence (even though a nearby colleague may have it in a desk drawer) or because the structure of the database was perceived as being too complex. As time progresses, more and more data are being collected, in much larger volumes than previously, by a larger and more international group of people, so that this problem

is becoming worse. Second, the data must be readily available (preferably on-line). A delay of a month while a scientist attempts to obtain a tape of a critical data set would greatly interfere with a study, and especially with pursuing promising (but speculative) paths during such a study. The available data should also be complete and up to date, which is to say that they should be obtained as close as possible to their original source, as contrasted to using incomplete copies of unknown pedigree. Third, the method of data access needs to be reasonably uniform, even though the individual data sets may be stored in physically different places and by different methods. A scientist should not have to deal with a dozen or so idiosyncratic data access methods in order to access a dozen kinds of data. Fourth, there is a critical need for user-friendly software applications that can access, display (e.g. in map format) and manipulate data. Finally, the whole system must be flexible enough that it allows for scientific creativity in an exploratory research environment. The various pieces of the system must be coupled sufficiently loosely that they can be replaced with new, experimental, or custom made, pieces. The system must not be self-contained, but instead only one facet of an overall research environment.

DATA ACCESSABILITY CURRENTLY POOR

While the earth science community is gradually moving towards greater use of on-line data, the current situation - with a few exceptions - is poor. Most groups charged with archiving data still do so mainly onto magnetic tape, distributed by mail. Some modernization has taken place, mostly in the form of a move from tape to CD-ROMS, but often without any improvement in the data formats or the level of compatibility between different CD-ROM reading software. Furthermore, continued reliance of distribution by mail leads to access that is much slower and more awkward than can be achieved by electronic means, such as Internet.

The exceptions to this situation are noteworthy. Some institutions make data available through the Internet's File Transfer Protocol (FTP) facility, which can be used to transfer files from an archive to a scientist's computer. A few institutions (e.g. the Center for Seismic Studies and N.A.S.A.) make arrangements for scientists to access their on-site database management system through guest accounts on their computers. A very few have network accessible databases (e.g. the Incorporated Institutions for Seismology's Data Center and Lamont). These exceptions demonstrate that a large segment of the community is both willing and able to use network technology to share data.

There has been considerable discussion among some members of the earth science community on how data access can be improved. Some people favor the establishment of more centralized data centers. Our own feeling is that data archiving must not be far removed from the scientists who extensively use that data, or else data quality suffers. Data archiving centers that are completely divorced from scientific research tend to become merely 'clearing houses' for the distribution of third party data of uncertain quality. Furthermore, without research-driven competitiveness, data archiving centers tend to fall quickly behind technologically. Thus we see distinct advantages to a model in which a combination of smallish, topical data centers and groups of scientists at research institutions make individual data sets available as part of a loose, network-distributed database. The data remains close to the scientists who generate it, but (given software capable of automatically finding the data on the network) is available to everyone else.

One argument that has been voiced against the use of a network-distributed database is the perception that it cuts off members of the earth science community at smaller institutions who lack access to Internet. This argument is specious: an inexpensive telephone modem can be used

to communicate between a personal computer at a small institution and a computer on Internet, provided arrangements for an account have been made. Once connected, the scientists has access not only to that computer's data, but all of the data in the network (albeit at a fairly slow data transmittal rate). Furthermore, the cost of a direct Internet connection is rapidly declining.

LAMONT'S DESIGN OF A DATA ACCESS SYSTEM

No software system which comes close to embodying the features that we have discussed above is currently available (though many data management and display systems contain clever, innovative ideas worth emulating). Most of the commercial Geographical Information Systems (GIS) are either very limited in their capabilities (to, say, satellite photographs) or very rigidly structured, in a way that might be suitable for a production, but not a research, environment. Furthermore, most existing GIS software provide only minimal data management and data access services. We at Lamont therefore developed a prototype system in order to work out some of the theoretical and practical problems involved. Development of these tools has a high potential of leading to scientific advances, since it simultaneously frees the scientist of the drudgery of assembling a data set, and provides a larger set of data from which new insights can be gained.

Our design was primarily motivated by the way in which Lamont (and many other labs) function, with small groups of scientists collecting and archiving a given type of data, but with a much wider community of scientists accessing (reading) that data. The scientists who collect the data want primary control over the data. They want the data to reside on a nearby computer system. They want to choose their own data management system, design their own database structure, and control changes made to the database. Since they are the experts on that data, they need the most sophisticated access to the data. Other scientists have only read access to the data, and may be willing to trade off flexibility for convenience of access. For instance, if the data are stored under a relational database manager, they may be satisfied with a few predefined 'views' (where a view is a single flat table built from the many tables of the database) rather than general access to the complex database itself.

We therefore decided to distinguish between these two types of access to the data. We assume that the data are stored using some database management (DBM) software, which may be of university or commercial origin. This DBM software is used for inputting data into databases, and can be invoked directly by the scientist for 'expert' access to the data. We then designed an intermediate layer of software, a 'view-server', that separates and insulates the ordinary scientist from complexity of the DBM software and of the database itself. The view-server keeps track of which databases are available (there may be many, each controlled by a different small group of scientists, each at a different location on the local - or wide - area network), and their schema, and can get views of data from the database (Figure 1). Each view is a simple, rectangular table of data. The structure of a view is intentionally kept simple so that it is easy to write software that can learn the schema of new views from a view-server and present the data in a view in some desirable form (e.g. as graphs, tables, or maps) to the scientists. The ability of the scientist's display software to manipulate a particular view is not built in, but rather is learned from the view-server. This automatic learning facility improves the software's versatility.

In order to further simplify the design of applications that can communicate with the view-server, we have limited the fields of views to parametric data (integers, floating point numbers, and character strings). We handle times series and images in a rather ad hoc fashion, by assuming that they are stored in a pre-defined format in a file, and including the file pathname as a

field of the view. Standard Internet file transport (e.g. FTP or NFS) is then used to recover the time series or image.

Our main effort in application development has been a geographical database browser (called 'GB'). This application is the main tool for extracting data from databases. It also has most of the features of a traditional GIS application: a scientist can mathematically manipulate parametric data and images and display them on maps. GB can be used in conjunction with other software (including commercial GIS applications), as part of a larger research environment. A key part of the overall philosophy of the Lamont system is to keep the different parts of the system as loosely coupled as possible, with just enough standards to maintain a 'look and feel' similarity between applications and to facilitate interchange of data between them. Thus the system can easily be built upon by specialized users. For example, Lamont's Seismology division has developed several substantial seismogram manipulation and earthquake location applications which are on the one hand completely distinct from the data access system, and on the other hand function smoothly together with it.

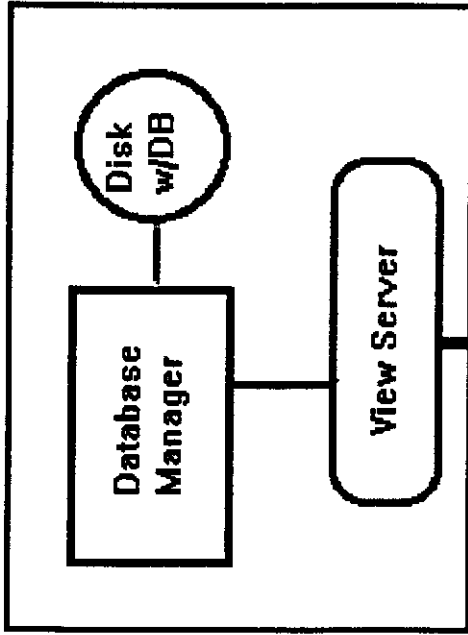
VIEW-SERVERS FOR PUBLIC ACCESS OF PARAMETRIC DATA

One of the pervasive problems with dealing with data provided by different institutions, or even different data sets provided by the same institution, is that they are usually in incompatible formats. Furthermore, software provided by the institution to read one data set often cannot read another. Even making a simple plot or map containing two types of data can be very difficult and time consuming. This problem would be solved if institutions could all agree to use a common data format and common conventions for data units. However, given the wide variety of data formats in use (ranging from simple text files of FORTRAN 'card-images' to formats imposed by commercial database management products), agreement on formats seems far away. Furthermore, database technology has not yet provided a single solution to all data archiving problems. Formats that might be appropriate for one kind of database might be completely unsuitable for another. The view-server concept is an attempt to define a compromise - a standard in the absence of standards. No requirements are placed on the underlying format in which the data are stored at the institution. Instead, the data are reformatted as they are transmitted over the network to the scientist. The data format issue is effectively sidestepped, providing the scientist with a uniform, network-based access method for publicly accessible geographical databases. Furthermore, the view-servers provide indexing and documentation services not usually provided by formats alone. Unfortunately, one of the negative aspects of the compromise is that uniformity in the data archiving process is left unimproved, and some of the complex (but useful) relationships that might exist between data in its native format might be lost when that data is reformatted for distribution over the network.

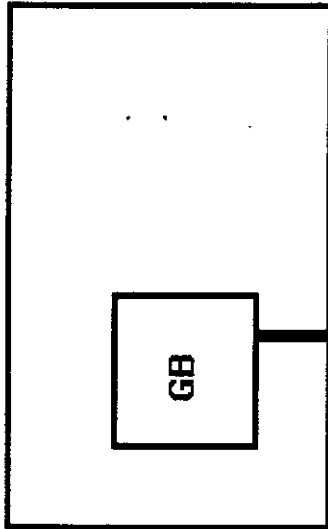
The view-server is implemented as a standard network service, build upon a client-server computer architecture. The scientist runs a software application (the 'client') to retrieve and display data, which communicates over the network to another software module (the 'server') running on a different computer that supplies the data (Figure 1). Communication is by messages in a standard 'protocol' that we have developed.

Before designing the view-server and its underlying network protocol, we considered whether any existing standards could be adapted and applied. One strong candidate we evaluated is Structured Query Language (SQL), a database access programming language used in many commercial database management systems. We decided it could not be used directly, for three reasons: such a choice would seem to preclude the use of fast, simple non-SQL databases which are already

Site 2 – Database Server

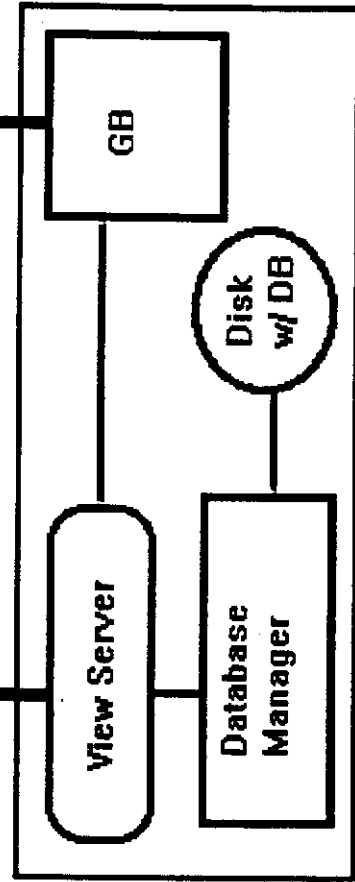


Site 1 – User



INTERNET

Site 3 – Database Server and User



in use in many university data archives; in actual practice, relational database managers from different vendors use substantially incompatible versions of SQL and completely different methods of communicating SQL queries to database managers; and finally, SQL lacks standard features for documenting the contents of databases. In our opinion, the ability to recover information *about* the data in a standard way is every bit as important as the ability to recover the data itself.

The protocol we have developed implements only three basic functions: Provide the client with a detailed description of the database views, including their names and a detailed description of their contents; Given a view name, suggest other, related, view names; Satisfy a request for data for a specific view. The first two functions are essentially 'contents' and 'index' functions, respectively. They automatically documents the contents of its databases and provide information that helps scientists assess whether a particular type of data is relevant to their work. They are used to answer questions such as: What kinds of data does this institution have? Does the institution have any of a particular type of data in the geographical region I am interested in? Where did the data come from? What is the meaning of a particular field in the database? Equally important, the view-server provides this information in a machine-readable format that can be used by software applications that present the data to the scientist. This feature makes possible software that can simultaneously display a very wide variety of data.

We have designed the view-server protocol to be simple enough that a view-server can be implemented in a few days of modest programming effort. It uses widely available UNIX sub-routines, such as Berkeley sockets and External Data Representation. Special purpose view-servers can be written to provide a network interface even to one-of-the kind databases. For instance, Lamont has a view-server for its 'GMT' marine geophysical database. We are currently providing view-servers software for four different database management systems: and ascii-view-server for databases based on simple text files; the 'dm-view-server' and 'sdm-view-server' for two very fast binary database managers written at Lamont; and the 'sql-view-server' for relational DBM's based on the popular SQL database language (e.g. UNIFY and SYBASE).

SIMPLE COMMAND-LINE ACCESS TO VIEW-SERVERS

A 'user-interface' is a program that provides a scientist with access to data made available by the view-servers. The simplest user interface that we provide is a set of command-line oriented programs that can be run on a simple, non-graphical terminal connected to a UNIX computer. Since the connection might be made through a telephone modem, this simple user interface makes the entire distributed database accessible to a scientist with access to an account to any computer on Internet. Information about the views made available by the view-server running on a given computer can be obtained with the command:

```
rshowviews hostname
```

where 'hostname' is the Internet name of the computer running the view-server. Once the scientists is aware of an interesting view, data from that view for a particular part of the world can be obtained with the command:

```
rviewsdm host view_name query left right bottom top
```

Here 'query' is a logical expression (in a C-like syntax) operating on the fields of a single record

of the view, and 'left', 'right', 'bottom' and 'top' are the longitude and latitude bounds of a rectangular region on the earth. For instance, if the view 'ISC_hypocenters' contained a field 'magnitude', the scientist could recover all hypocenters from a given region with the query 'magnitude>5.5'. The data retrieved by this command are in a simple, ASCII 'spreadsheet' format.

The scientist then can inquire whether there are any other views related to a particular view with the command:

```
rassocviews host view_name
```

Perhaps, following the example above, the 'ISC_hypocenters' view is associated with a 'ISC_arrival_time' view. Then the scientist could go on to recover the arrival times associated with a particular earthquake with the command:

```
rviewsdm host view_name query target_id
```

Here the 'target_id' would be the ID of the hypocenter of interest.

HIGH PERFORMANCE MAP-BASED USER INTERFACE

The Lamont Geographical Database Browser (GB), is an X-Windows (and SunView) based GIS application which can display data in the form of symbols on maps (Figure 2). This application learns from network view-servers about what types of data are currently available anywhere on the Internet and determines their schema. The user can make queries to retrieve these data and can manipulate them graphically and symbolically. This software employs a powerful mathematical expression interpreter that allows the user to make queries from the database, perform sorting and arithmetic functions, and control symbol attributes, all through symbolic expressions operating on the fields of the views. The user can call up any number of views, and plot them on a base map (either as discrete symbols for parameters, or by contouring images). GB can also be used to transfer data to other programs, including spreadsheets, time-series analysis programs and 2- and 3-dimensional plotting programs, all of which use common data formats.

CURRENTLY AVAILABLE DATA

Lamont currently makes the following kinds of parametric data available through its view-servers:

Geodetic: Vector coastlines at a variety of resolutions; Rivers, lakes, political boundaries, etc. for selected parts of the world.

Seismological data: Earthquake hypocentral catalogs, including the ISC and PDE global catalogues and local catalogues such as Lamont's New York State and Aleutian catalogues, C.D.-M.G.'s California Catalogue, D.N.A.G.'s U.S. Catalogue, E.P.R.I.'s Eastern U.S. Catalogue.; Centroid moment tensors, mapped faults for selected regions of the world; Strong motion accelerogram from the United States; locations of seismological observatories;

Geophysical Data: Along track measurements of gravity anomalies, magnetic anomalies and bathymetry for UNOLS research ships; global land gravity stations; satellite geoid, seafloor magnetic anomalies and isochrons; marine heat flow measurements;

Geological: Deep Sea Drilling Program and Ocean Drilling Program drill sites; the Lamont and Oregon State University piston core catalogues.

Our expectation is that the list of available data will grow as data are contributed by other institutions. We provide a directory service for view-servers at other institutions through the view named 'view_servers' on chaos.ldgo.columbia.edu. E-Mail should be addressed to menke@lamont.ldgo.columbia.edu to announce the existence of view-servers.

AVAILABILITY OF SOFTWARE AND ACCESS TO VIEW-SERVERS

We encourage interested scientists and data management staff to consider running view-servers on their institution's computers. Such an effort will be particularly easy if you have access to Sun workstations, since recompilation of the software will not be necessary. First transfer the manual, gb.inst.ps, and the rshowviews, rviewsdm, rassocviews programs to your computer (see Table 1). You will then be able to access view information and data from view-servers (see Table 2). Then transfer and install GB, which retrieves data and displays them on maps. Finally, pick one of the simpler view-servers (e.g. ASCII or DM), install it on your computer, and build some of your own databases. Unfortunately, due to our limited resources, we cannot offer you any technical support in this endeavour.

License to use the software described in this article is granted free of charge to non-profit educational institutions and U.S. Government Agencies, by simply completing and returning the License Agreement in the document gb.lic.ps (see Table 1). All material can be obtained by anonymous FTP to lamont.ldgo.columbia.edu (129.236.10.30) and consists of compressed UNIX 'tar' files in the directory ~ftp/pub (see Table 2). In return for the use of this software, we request that your host institutions be diligent in maintaining databases and making them publicly available. Some interesting time series and raster data are available on the public NFS-mountable file system duke.ldgo.columbia.edu:/duke/nceer_data.

Acknowledgments. We thank all the people who over the years contributed to Lamont's software development efforts, including members of our programming staff: Richard Boaz, Roger Davis, Ken Howard, Yalin Liu, Brenda Murphy, Suzanne O'Hara, Doug Shearer; engineers: Dale Chayes, David Lentricchia, Larry Shengold and scientists: Dallas Abbott, Tom Boyd, Alberto Malinverno, Walter Smith, Jeffrey Weissel, Paul Wessel, and Dean Witte. We also thank people from other institutions for their comments and assistance, including Bill Leith and Janet Walz of the U.S. Geological Survey. Parts of this research was funded by the National Science Foundation under grants EAR87-96177 and IRI93-13126. Lamont-Doherty Contribution Number 4832.

Table 1. Compressed UNIX TAR files and their contents.

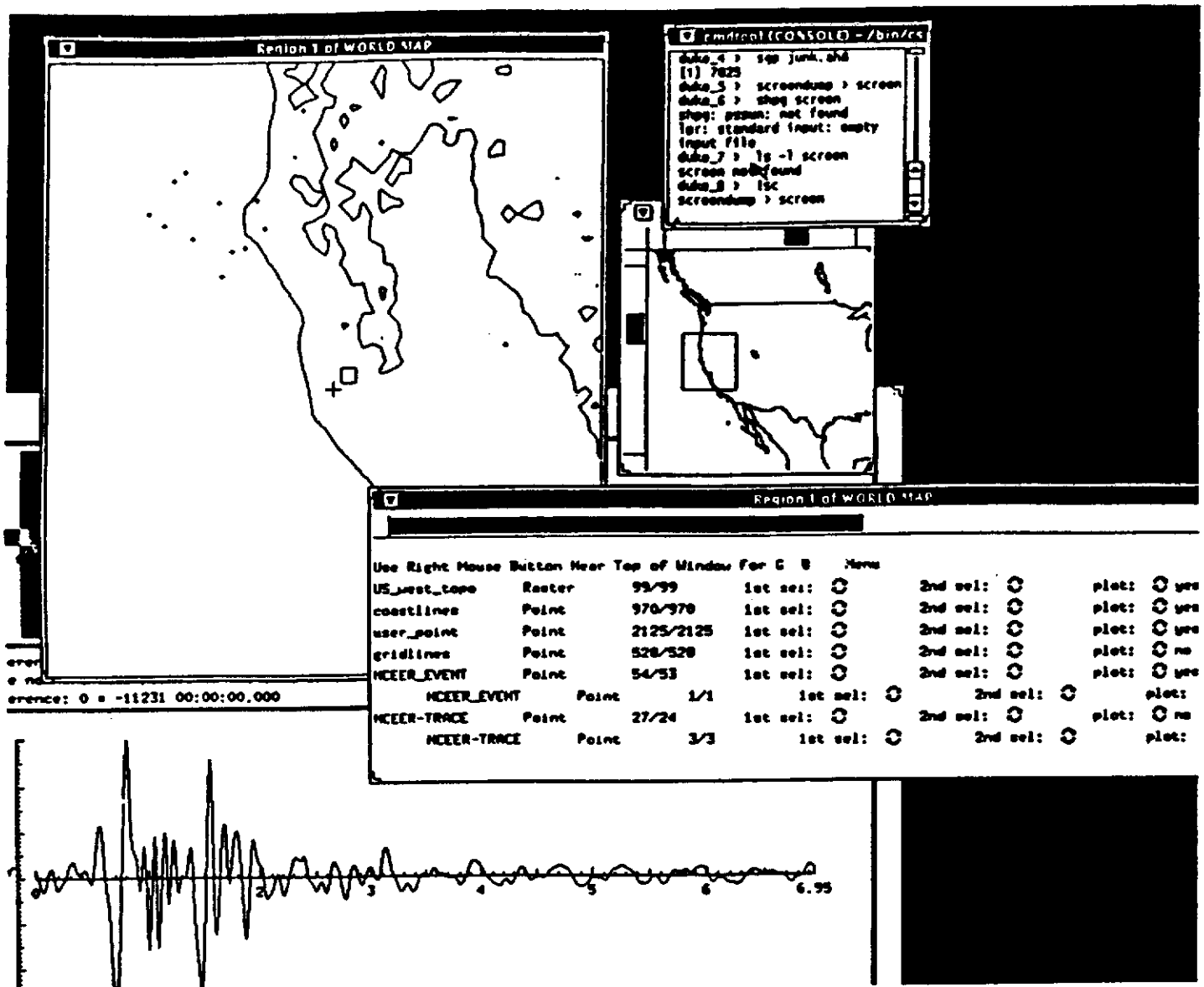
<u>name</u>	<u>contents</u>
gb.lic.ps	License Agreement (PostScript Format)
gb.tar.Z	GB geographical browser, and manual for entire system, gb/doc/gb.inst.ps (PostScript format)
dm.tar.Z	DM database manager and view-server
ascii.tar.Z	ASCII database manager and view-server
sdm.tar.Z	SDM database manager and view-server, and rshowview, rviewsdm, rassocviews commands
sql.tar.Z	SYBASE and UNIFY sql-view-servers
parse.tar.Z	mathematical expression parser subroutines
lxt.tar.Z	Lamont X-Toolkit subroutine library

Table 2. Public view-servers at the Lamont-Doherty Geological Observatory.

<u>Name</u>	<u>Number</u>	<u>Type</u>	<u>Contents</u>
miles.ldgo.columbia.edu	129.236.10.70	DM	misc. geophysical
duke.ldgo.columbia.edu	129.236.10.50	SYBASE	NCEER strong motion
trane.ldgo.columbia.edu	129.236.10.90	SDM	misc. seismic
chaos.ldgo.columbia.edu	129.236.10.20	ASCII	misc. geophysical
clipper.ldgo.columbia.edu	129.236.20.90	GMT	marine geophysical ship data
ocean.ldgo.columbia.edu	129.236.20.71	GMT	marine geophysical ship data

Fig. 1. Data accessing system. The scientist's computer at Site 1, which is running a map based geographical database browsing program (e.g. GB) can retrieve data from data serving computers at Site 2 and Site 3, located elsewhere on Internet.

Fig. 2. Screen of scientist's workstation during typical use of the Lamont Geographical Database Browser (GB). The scientist is displaying a map of western North America. Coast lines and 1000 meter topographic profiles are shown for orientation. Earthquake hypocenters from the NCEER_-EVENT view (crosses) have been retrieved from the view-server on the computer duke.ldgo.columbia.edu, which is maintained by the National Center for Earthquake Engineering Research. The scientist has selected one of these events (large cross) and retrieved records from the NCEER_-TRACE view for that earthquake. The scientists has selected one record from this view, corresponding to a seismic station located near the hypocenter (square). One of the fields of this view contains the pathname of the strong motion accelerogram recorded at that station. This time series is displayed using a separate program.



Regional Systems

11 10 9 8 7 6 5 4 3 2 1

PROTOTYPE
ENVIRONMENTAL DATA AND INFORMATION MANAGEMENT SYSTEM
FOR THE GULF OF MAINE

by

Wendell S. Brown and Karen M. Garrison
Ocean Process Analysis Laboratory
Institute for the Study of Earth, Oceans, and Space
University of New Hampshire
Durham, NH 03824

25 October 1993

ABSTRACT

An Environmental Data and Information Management System (EDIMS) is being developed under the auspices of the Council on the Marine Environment for the Gulf of Maine region. When fully operational, EDIMS will incorporate a diverse blend of archived and real-time data from operational and research sources in a decentralized database. Data users/suppliers from the states, provinces and federal agencies will be able to exchange data and information via Internet. The prototype EDIMS consists of a few representative databases, which are being accessed by a small data user/supplier group. We are developing a set of protocols that will enable the data users/suppliers to: query a directory of the regional Gulf of Maine data sets; electronically transfer a selected subset of these data to their own computing environment; and communicate generally with the prototype EDIMS user community.

A subset of the prototype EDIMS databases—including a Gulf of Maine database directory, Who's Who, Gulf Watch mussel data, and the NOAA real-time data—reside on the EDIMS host client/server computer at the University of New Hampshire. When fully implemented, most of the EDIMS databases will reside at their respective remote storage sites. During the prototype EDIMS development, most of the databases will be static, except for the regularly updated Gulf Watch, NOAA, and river discharge databases.

EDIMS has a SQL/ORACLE based data directory which enables EDIMS users to browse the EDIMS database directory for information. The user will be able to initiate the instant retrieval of any of the electronic EDIMS databases; consisting of a data description header, and a flat ASCII data file. Internet is the means for distributing the prototype EDIMS data and information. We monitor user access to the different EDIMS databases and will distribute logs to data suppliers on a regular basis.

CONTEXT

The Gulf of Maine Council on the Marine Environment (GOM/CME) has developed a ten-year Gulf Action Plan to address issues concerning the environmental health of the Gulf and management of the marine resource. Toward that end, the GOM/CME established a marine environmental quality monitoring program focusing on mussels (Gulf Watch). The GOM/CME Working Group also established a Data and Information Management Committee (DIMC) charged with the task of developing what we call an Environmental Data and Information Management System (EDIMS) for the Gulf of Maine region. The EDIMS will enable a diverse set of users in the region to exchange pertinent information and data (incl.

Gulf Watch) among themselves. The DIMC has subdivided the implementation of an EDIMS into a set of tasks including the development of a regional database directory and a prototype EDIMS. The latter will accommodate the Gulf Watch mussel data.

THE EDIMS CONCEPT

The EDIMS is designed to make a data directory and a relevant set of data bases accessible to EDIMS users. The prototype EDIMS is designed to be simple and flexible enough to accommodate anticipated changes. We expect that the eventual user group will include marine environment and resource managers; state and provincial planners; and ocean scientists and engineers. As users become more familiar with the prototype EDIMS they will articulate their needs more clearly. The DIMC will analyze that feedback and direct future development accordingly.

The fully operational EDIMS will need to deliver to its users a large and diverse blend of archived and real-time data from operational and research sources. Thus EDIMS is structured around a decentralized database. Figure 1 illustrates schematically how data users/suppliers from the states, provinces and federal agencies will be able to exchange data and information via the Internet network. This approach relies on the effort and resources of the data users/suppliers and thus can be expanded relatively easily.

While the EDIMS data directory, Gulf Watch data, and a few other selected data sets--will reside at a host computer site, most of the EDIMS databases will reside at remote locations. EDIMS data users/suppliers will link to the network and thus have access to all of the information and data at the EDIMS host site (GOM EDIMS), as well as the remote EDIMS sites. As envisioned, the GOM/CME will ask EDIMS data users/ suppliers to assume a major share of the responsibility for database quality and maintenance. They will be assisted by an EDIMS manager at the host site who will oversee the operations of the EDIMS and implement DIMC initiatives. We expect that different data users/suppliers in the region will commit to and support such an EDIMS because of their need to access the comprehensive EDIMS database. This information will enable them to conduct research, protect public health, and/or manage the Gulf of Maine marine resource better than ever before.

PROTOTYPE EDIMS DESCRIPTION

A UNH development team has constructed and is testing the prototype EDIMS. The prototype EDIMS consists of a few representative databases (see below), which are being accessed by a small DIMC data user/supplier group. We are developing a set of protocols that will enable the group of data users/suppliers to (a) query a directory of the regional Gulf of Maine data sets; (b) electronically transfer a selected subset of these data to their own computing environment; and (c) communicate generally with the prototype EDIMS user community.

We have selected a diverse set of databases to be part of the prototype EDIMS. The prototype EDIMS includes:

- A Gulf of Maine database directory
- Documentation
- Gulf of Maine maps and bathymetry
- Massachusetts Bays Program physical oceanographic data archive
- Real-time satellite imagery and meteorology from the NOAA/NOS Ocean Products Division
- The New Brunswick Department of Environment real-time river discharges
- The USGS sediment texture data archive
- Dartmouth model "data" for the Gulf of Maine
- The Bedford Institute of Oceanography Atlantic Fisheries Adjustment Program (AFAP) hydrographic data archive

Our goal in the next six months is to add:

- A "Who's Who" in the Gulf of Maine
- The Gulf Watch mussel data
- The EOEA Massachusetts shellfish data archive

In the prototype Gulf of Maine database directory, documentation and all but the Dartmouth databases reside on the EDIMS host client/server computer at the University of New Hampshire. (When fully implemented, most of the EDIMS databases will reside at their remote storage sites.) During the prototype EDIMS development, most of the databases will be static, except for the regularly updated NOAA and river discharge databases.

Internet is the conduit for the prototype EDIMS data and information. It is an established and well-documented international network, with data and mail transfer protocols that can be implemented on a variety of platforms. The Internet File Transfer Protocol (FTP) feature is used to retrieve selected data from the host and/or remote storage sites. In an effort to keep track of EDIMS use, we will monitor access to the different EDIMS databases. Logs detailing access to EDIMS will be maintained and distributed to data suppliers on a regular basis.

EDIMS includes a SQL/ORACLE-based data directory. A simple query system enables EDIMS users to browse the EDIMS database directory for information in user-specified time and space domains. This documentation describes how to browse the EDIMS directory and how to "capture" any of the electronic EDIMS databases, which consist of a data description header, and a flat ASCII data file.

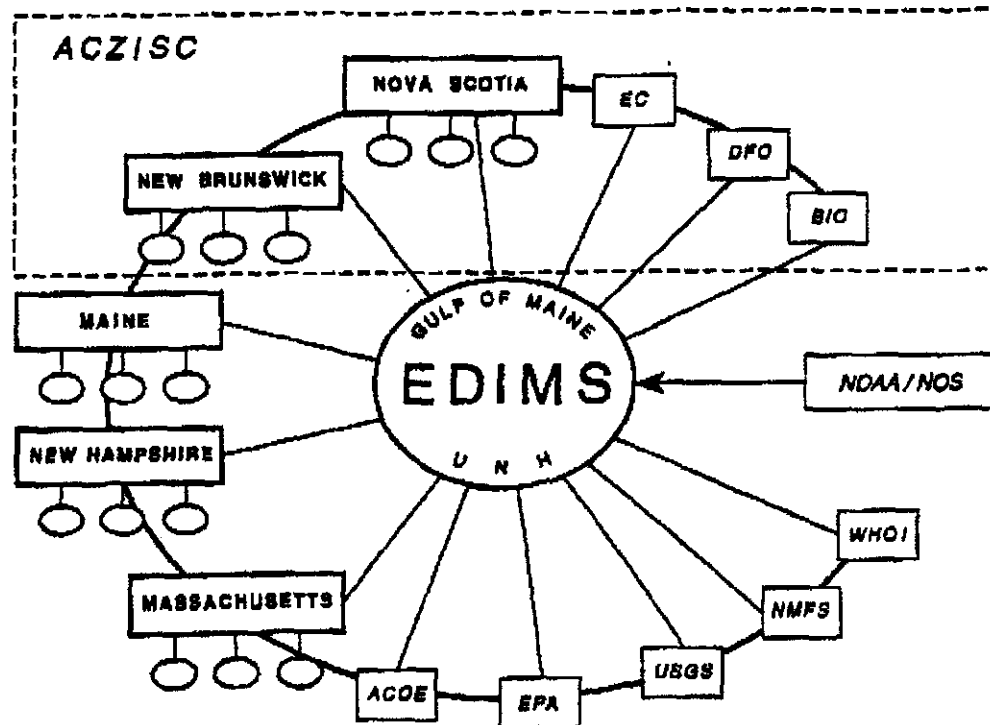


Figure 1. Conceptual design of the GOM/CME Environmental Data and Information Management System EDIMS. The host site (GOM EDIMS) and remote databases are linked together by Internet. NOAA/NOS provides realtime data that is captured at the host site.

USER DOCUMENTATION

This draft EDIMS User's Documentation seeks to help users in (a) accessing EDIMS, (b) navigating the EDIMS data directory (Figure 2), (c) using the Internet e-mail utility and (d) acquiring EDIMS data. The best guidance for the future development of EDIMS will come from users. Thus we encourage users to critique the existing prototype EDIMS.

●ACCESSING EDIMS

- 1) Log into your own INTERNET node.
- 2) Type what is in the brackets { ... }.
- 3) OK, (telnet metro.unh.edu)

[Telnet Rev. 2.4 Copyright (c) 1991, Prime Computer, Inc.]

Please wait...

Connecting to: metro.unh.edu (132.177.128.67)

This is
UNH-Specific

Entering INPUT mode. Your escape character is: ^]

Entering INPUT mode. Your escape character is: ^]
RISC/os (metro)

- 4) login: {edims}
- 5) password: {ledims!}

RISC/os (UMIPS) 4.52.p3.2 metro
(c) Copyright 1986-1990, MIPS Computer Systems
(c) Copyright 1990, PRIME Computer Systems
All Rights Reserved

The above puts the user into the EDIMS MAIN MENU (see page 7). The three options (see Fig. 2) are discussed starting on the next page.

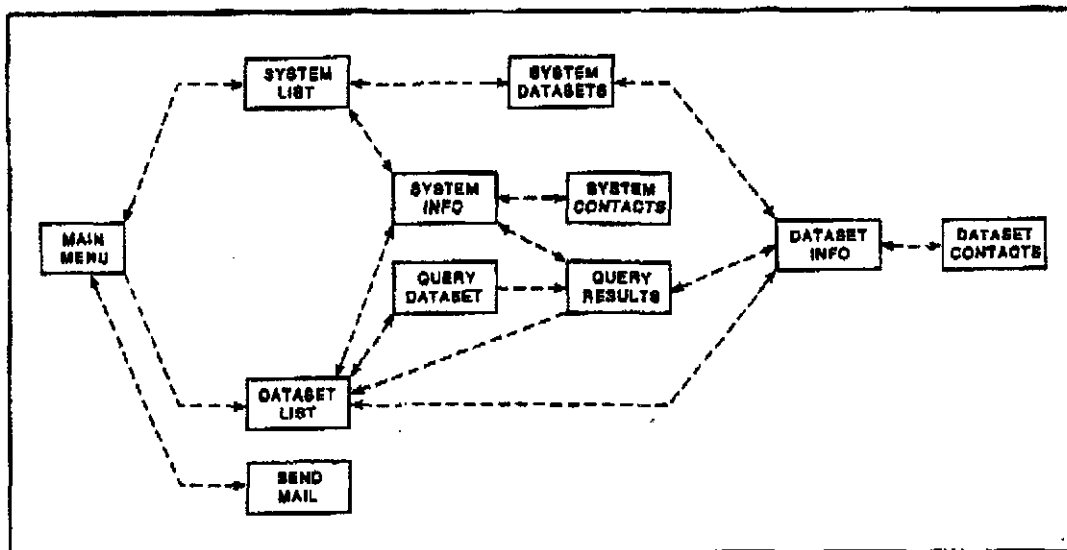


Figure 2. Block diagram of the EDIMS Data Directory.

*Keyboard entries on the EDIMS computer are case-sensitive and you must emulate a vt100 terminal.

DATA ACQUISITION DEMONSTRATION

To transfer EDIMS NOAA/NOCN meteorological data electronically to your own computer via Internet:

PREVIEW DATA HEADER

OK, ftp metro.unh.edu

[FTP Rev. 2.4 Copyright (c) 1991, Prime Computer, Inc.]
FTP is connecting to remote host: metro.unh.edu
220 metro FTP server (from 4.3BSD Tahoe) ready.
Remote host connected.

Username (metro.unh.edu: ...): anonymous
331 Guest login OK, send id as password
Password (metro.unh.edu: ...): user@internet.address
230 Guest login OK, access restrictions apply

ftp > cd /pub/nocn/metro/1992/01/1992

250 CWD command successful.

ftp > get b44007.dat
ftp > bye

Note: At this point, the user can read the ASCII header file using an "editor" with which the user is familiar or see next page.

RETRIEVE DATA

Note: ftp metro.unh.edu as described above.

ftp > get b44007.dat

ftp > bye
221 Goodbye.
Remote host disconnected.
Exiting due to quit request.

Exiting FTP.
OK,

Note: Contents of b44007.dat appear on page after next.

- Logon to anonymous ftp area on the EDIMS host computer (UNH-specific).

- Move to proper directory.
- Transfer header file to user computer and leave ftp.

- Transfer data file

These data cannot be used for commercial purposes without express permission of the data provider.

Header File (buoy.asc)

NOAA Realtime Meteorological Data

The National Data Buoy Center (NDBC) maintains a network of ocean buoys and C-Man island stations that automatically log and transmit near-realtime meteorological and oceanographic data to the National Ocean Service in Maryland. The data from the stations listed below are being archived in EDIMS central on a daily basis.

<u>Station ID</u>	<u>Location</u>	<u>Type</u>	<u>Latitude</u>	<u>Longitude</u>	<u>Filename</u>
44005	Gulf of Maine	moored	42.6N	68.6W	b44005.dat
44007	Portland	moored	43.5N	70.1W	b44007.dat
44008	Nantucket	moored	40.5N	69.4W	b44008.dat
44011	Georges Bank	moored	41.1N	66.6W	b44011.dat
44013	Boston	moored	42.4N	70.8W	b44013.dat
IOSN3	Isles of Shoals	C-Man	43.0N	70.6W	iiosn3.dat
MDRM1	Mt. Desert Rock	C-Man	44.0N	68.1W	imdrml.dat
MISM1	Matinicus Rock	C-Man	43.8N	68.9W	imisml.dat

A set of 3-hourly data records from each site are captured from the NOAA Ocean Communications Network (NOCN) computer on a daily basis. The data are linear interpolated to standard three hour intervals and written to separate files (see table above for filenames). Each ASCII data record is configured as:

Column 1	Year
Column 2	Month
Column 3	Day
Column 4	Hour (UTC)
Column 5	Air Temperature (°C)
Column 6	Sea Surface Temperature (°C)
Column 7	Atmospheric Pressure (mb)
Column 8	Wind Speed (m/sec)
Column 9	Wind Direction (deg)
Column 10	Wave Height (m)

Note: Missing data are replaced with -99's.

Data Example

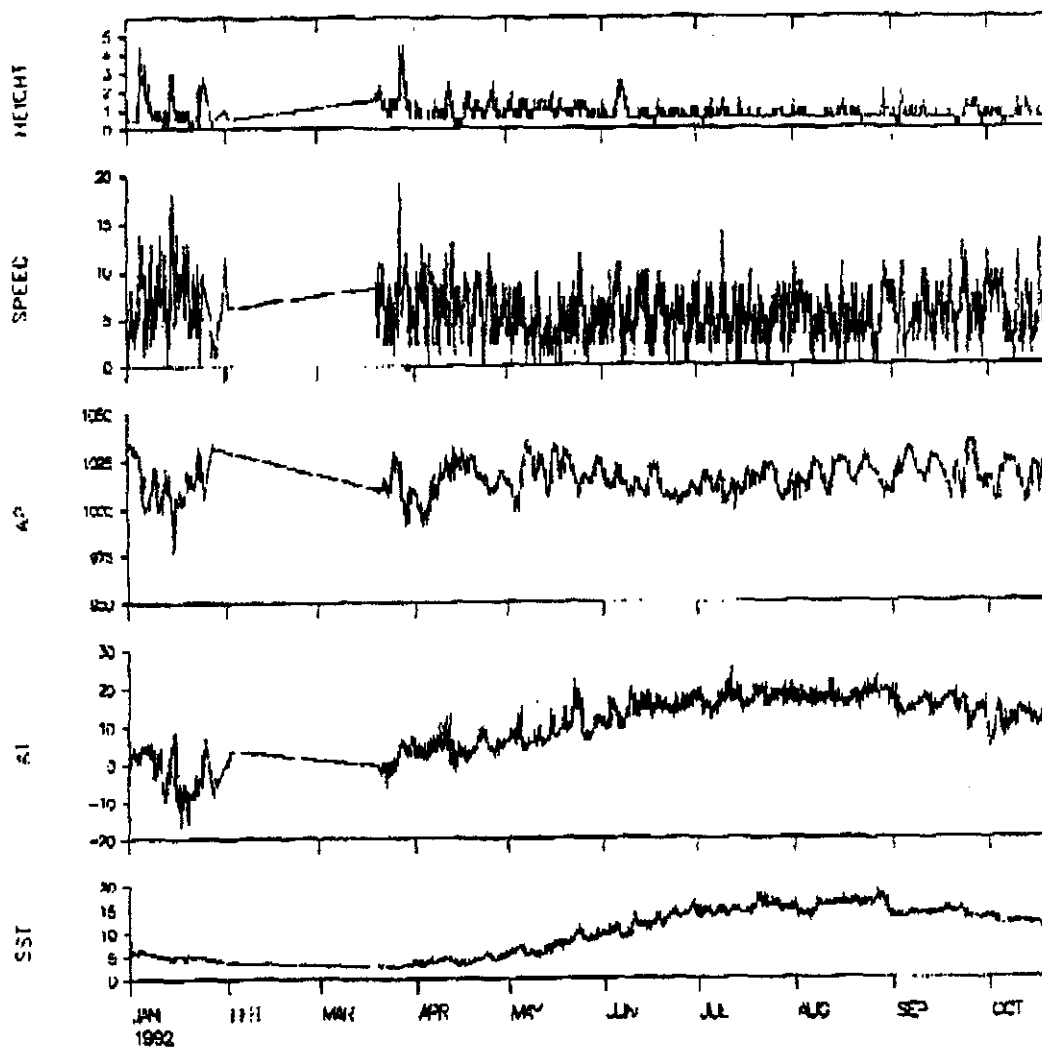
C> ~~TYPE 14007.DAT~~

93	1	1	0	3.4	5.7	1004.8	9.8	220.0	1.0
93	1	1	300	4.0	5.8	1006.5	4.1	290.0	1.0
93	1	1	600	4.0	5.8	1007.2	1.0	270.0	1.0
93	1	1	900	3.8	5.9	1007.8	2.1	20.0	1.0
93	1	1	1500	1.4	5.8	1011.3	11.8	300.0	1.0
93	1	1	1800	-1.7	5.9	1014.1	10.8	320.0	1.0
93	1	1	2100	-3.6	5.8	1016.8	8.7	320.0	0.5
93	1	2	0	-5.7	5.6	1020.9	8.7	310.0	0.5
93	1	2	300	-7.4	5.8	1023.1	5.1	350.0	0.5
93	1	2	600	-8.5	5.7	1024.0	6.2	300.0	0.5
93	1	2	800	-7.7	5.7	1024.2	6.2	260.0	0.5
93	1	2	1000	-7.2	5.7	1023.8	8.7	270.0	0.5
93	1	2	1200	-6.6	5.6	1025.2	8.2	270.0	0.5
93	1	2	1500	-6.1	5.6	1028.8	9.8	330.0	0.5
93	1	2	1800	-6.7	5.6	1030.2	9.8	340.0	0.5
93	1	2	2100	-6.3	5.6	1032.1	8.2	340.0	0.5
93	1	3	0	-6.5	5.6	1034.7	7.2	340.0	0.5
93	1	3	300	-8.3	5.6	1037.7	9.8	360.0	0.5
93	1	3	600	-9.2	5.6	1039.2	6.2	10.0	0.5
93	1	3	900	-9.1	5.5	1039.3	4.1	360.0	0.0
93	1	3	1200	-7.8	5.5	1040.1	3.1	70.0	0.0
93	1	3	1500	-5.9	5.5	1041.0	2.1	100.0	0.5
93	1	3	1800	-2.1	5.1	1038.2	5.1	230.0	0.5
93	1	3	2100	-0.6	5.5	1035.0	9.8	190.0	0.5

In "OPAL-land" we use our own data processing package (OASP) to

- READIN data ... to put in Ocean Format
- EDIT data ... to eliminate spurious data
- TIMPLTGKS ... to generate plot file
- SCREEN ... to view plot on screen (see below)

PORTLAND MAINE BUOY



COMMENTS ON CHESAPEAKE BAY PROGRAM INFORMATION AND DATA MANAGEMENT TOOLS

Joseph Macknis
Monitoring Coordinator
Chesapeake Bay Program

1. CHESSE

- hierarchical, menu driven system to access text and data files installed on Chesapeake Bay VMS computer system
- outdated, created in 1984; Folio and Infobase software currently available provide much greater capability to access and link data bases and to display data
- lessons learned
 - * costly to maintain, required contractor support
 - * schedule regular updates
 - * assign responsibility for update close to data generation

2. Review of Chesapeake Bay Data Management System

- converting from VMS to ORACLE
 - * EPA agency standard
- identifying linkages through subcommittees by proposing management questions
 - * effective process to link data bases in order to meet management questions

3. Chesapeake Bay Information Tools

- developed to support monitoring program, the scientific and technical cornerstone of the Chesapeake Bay Program
- routine, coordinated monitoring provided basis for model calibration and verification, assessment of effectiveness of management efforts, development of water quality targets for living resources and hypothesis generation and verification to fuel research
- CAMS (Chesapeake Automated Data Transfer System)
 - * customized PC software to facilitate data submission and data verification
 - * output is standardized data set
 - * resolves QA problems close to data source

- **DAITS (Data Analysis Issue Tracking System)**
 - * provides process to resolve and document technical issues from all program areas
 - * ties up loose cannons, applicable to other areas
- **Guide to Using Chesapeake Bay Water Quality Data**
 - * describes water quality data base
 - * informs user/data analyst of changes in detection limits, methods, sampling design that are critical to interpreting the data
- **Data Management Plan**
 - * establishes common data attributes across program areas in similar or translatable formats
 - * describes forms and procedures for the submission, retrieval, and storage of Chesapeake Bay Program data
- **Data Directory**
 - * set of definitions for all data items in the CBP data base

4. Information and Data

- data are numeric values that require interpretation to become information
- information is product to managers and citizens that builds support to continue/maintain program
- examples of information products
 - * Bay Journal
 - * State of the Bay Report

Agency Systems

**Data and Information Management
at the Branch of Atlantic Marine Geology
and the
Contaminated-Sediment Database**

Marilyn R. Buchholtz ten Brink , Frank T. Manheim,
Nancy K. Soderberg, and John C. Hathaway
U.S. Geological Survey, Woods Hole, MA 02543

Introduction

The U.S. Geological Survey Branch of Atlantic Marine Geology (BAMG) at Woods Hole MA is part of the Geological Division of the USGS. The other two divisions are the Mapping Division (which produces the well-known topographic maps), and the Water Resources Division (WRD). Our Branch is responsible for a full range of geological and environmentally related basic and applied research in marine and coastal geology along the Atlantic Continental Margin, the Gulf of Mexico, and the Caribbean Sea. In addition to the largest staff at Woods Hole, other groups are located at National Headquarters in Reston VA, in St. Petersburg Florida, and a small office in San Juan Puerto Rico. A corresponding Branch that deals with the Pacific and Arctic areas is located in Menlo Park California.

The objective of the current workshop is to review information management systems to answer the following questions:

- What data is available and pertinent to the needs of the Gulf of Maine research effort?
- How have participating organizations solved or been frustrated by data and information management problems?
- How can our mutual experience work to the benefit of all and aid efficiency, solve problems, promote cooperation, and get the job done more efficiently?

Data and Information Management at the USGS

The U.S. Geological Survey

The USGS mission is "Earth Science in the Public Service". To this end USGS BAMG provides basic and applied research in the geological and related environmental sciences, mapping and interpretation, data and information distribution and public education. As a research organization without regulatory obligations we do not make policy, but attempt to provide the most accurate and objective background information possible to serve the needs of scientists, managers, and the general public. Wide cooperation with Federal, and state agencies, as well as academic and other research organizations, private firms, and citizen and environmental groups is a key element of our operations.

The types of data and products generated by the USGS-BAMG span a very large spectrum, including:

- Sediment sampling and analysis (chemical, geological-physical, radiochemical (principally for dating) and paleontological (for stratigraphic and age-dating purposes), ranging from surficial to coring and drilling.
- Geophysical measurements, including seismic profiling and interpretation, magnetic and gravity measurements, sidescan sonar records and mosaics to map the nature of the sea floor and its deposits and hazards.

- Current meter and other physical oceanographic measurements that help model water movements, sediment and contaminant transport processes in coastal and estuarine areas, as well as sediment trap and nephelometer records that measure and recover transported sediments for further analysis.
- Video and photographic records of the sea floor.
- Use and improvement of base maps and data such as bathymetry and coastlines
- Interpretive maps, involving geology, environment, habitat, contaminants, erosion and sedimentation, and hazards.
- In addition, we access monitoring data provided by other parts of the USGS, such as riverine flow and load measurements by the Water Resources Division.

Most work passes through the stages of raw data, documented and edited data, synthesis and interpretation and publication. In recent years more and more attention has been focused on computer visualization of data, and output of large compilations and syntheses on CD-ROMS produced in house. Examples of these products will be available at the workshop.

TYPE	AN EXAMPLE
1) Digital Data	CD-ROM demonstrations
2) Topographic Maps	
3) Lists of Publications	Description of Atlantic Marine Programs
4) Technical Publications	Journal articles and books
5) Public Education	Map-making working manual
6) Data Management Systems	Federal Geographic Data Committee
7) Contaminated-Sediment Database	MacIntosh demonstration

At the National level, the USGS responds directly to data needs of the public and other agencies. It is currently chairing the Federal Geographic Data Committee (FGDC) and participates in its subcommittees, such as the Bathymetry Subcommittee. The FGDC is responsible for building a National Spatial Data Infrastructure (Fig. 1) with initial goals of building good communication links, recommending common procedures and standards, and establishing partnerships for transferring data or common databases. Among FGDC's publications, the Manual of Geographic Data Products (Fig. 2) lists data products available from all Federal agencies by topic.

USGS-BAMG data management operations

As with many other members of the GOM community, we also are struggling with how to provide efficient data access, information exchange, and publication internally and externally. The massive influx of data from many sources, as well as the dramatic improvements and price reductions in computer technology have caused us to place special emphasis on two areas: the establishment of sound database structures and databases, and efficient computer technology to hold, manage and interface data types.

AMG data management groups have been organized and have struggled to find options that meet our needs (Fig. 3). We are moving from the traditional (existing) model of limited access in a system comprised of one-PI who has one-project with one database, or limited databases for one parameter and all PIs, towards a moving target of a data and information system in which the following occurs :

- a) CD-ROM publication of all digital data and information types for each project (Fig. 4 for example)
- b) distributed databases through GOPHER and other systems
- c) toolboxes available for data access and manipulation
- d) the system is flexible, alluring and easy-to-use, and
- e) multiple media are maintained for information distribution.

We operate both desktop, workstation, and central computers (using a variety of operating systems) and geographic, geologic, mapping and visualization software, such as UNIX, MIPS, MAPGEN, GMT, ISM, AVS, MATLAB, SPYGLASS and ArcInfo³, on various platforms. Data is linked together internally by local networks while INTERNET provides access off site through rapidly developing data distribution systems like GOPHER, email, and FTP. This diversity of computer equipment and approaches helps assure our ability to cooperate with other organizations that use diverse computer technology. We also strive to make our data easily accessible and have found that digital publication, especially via CD-ROMs, allows full dissemination of all types of data and information for each project (e.g. Fig. 4). A CD ROM can include raw and qualifying information, maps and interpretive products, photographs plus "tool boxes" that permit easy data access and manipulation for inexperienced users.

As will be noted below, our largest specific sediment databases, which includes the East Coast Sediment Database (Fig. 5) are now managed by high-end desktop database management software like PARADOX³, which is compatible with most IBM databases and spreadsheets, and readily transferable to Macintosh systems, as well as uploadable or even directly linked (e.g. SQL Link) to syntactic language mainframe databases like SyBase, Oracle and InterBase³.

Contaminated-sediment database

The remainder of this discussion will focus on construction, use and distribution of contaminated-sediment databases in impacted coastal waters, which is one aspect of our data management and utilization efforts. We began with Boston Harbor and the Massachusetts Bays as a pilot project and are now expanding this (with our collaborators¹) into a Contaminated-Sediment Database for the Gulf of Maine. We also have contaminated-sediment database construction occurring in the Chesapeake Bay and Gulf of Mexico regions. (A demonstration was presented at the workshop of the Boston Harbor Contaminated-Sediment Database on Macintosh and DOS³ platforms).

Current situation

Contaminants are ubiquitous in our nation's ecosystems; coastal zones are often severely affected. Contaminants in bottom sediments may have long-term effects on components of the ecosystem even after the source is eliminated. National programs recognize the need for assessment of contaminant distribution. Contaminant inputs are changing; we need to know the distribution in the marine environment and the sources, sinks, transport paths, and residence times. Many current projects are targeted at monitoring selected stations, providing syntheses for management purposes, or are restricted to small geographical areas. A vast and poorly documented volume of published and unpublished data exists which has not been compiled and integrated into forms usable for the many potential users of the data. Many scientific questions and policy issues related to contaminants in sediments require data of historical, regional, and interdisciplinary scope. Acquisition of new data is expensive and may duplicate previous efforts if a full interpretation of existent data has not occurred. In addition, historic data represent samples in time that cannot be repeated. Consequently, it is logical and cost-effective to utilize existing data to its maximum so that it can better serve as a foundation, baseline, and starting point for further work.

It is important to obtain good geographical coverage in glacial regimes; this requires compilation from numerous studies. We believe that "representative data" from a few stations are not adequate for local sediment quality characterization and management. Local and regional

bathymetry and geologic substrate information profoundly affect contaminant distribution. "Hot spot" and "representative sampling" areas should be delineated more quantitatively and behavior of contaminants in these areas linked with geophysical and geological data.

What environmental managers, policy-makers, and scientific researchers need is an easily accessible, well-qualified and documented, and simple-to-use compilation of existing edited basic data on contaminated sediments and related sediment properties. The compilation should comprehensively contain original data from many heterogeneous sources, be able to grow and evolve, be an active scientific tool, and remain applicable to questions and models developed in the future.

The database

The Branch of Atlantic Marine Geology of the U.S. Geological Survey Branch of Atlantic Marine Geology (and collaborators²) has compiled a comprehensive chemical database on sediments in Boston Harbor-Massachusetts Bay, which contains some of the most contaminated sediments in the nation. Our database work in Boston Harbor and the Atlantic margin has shown that such a database can be successfully made and is widely requested. The principles and methods of the database are given in Table 1. Outgrowths from it are baselines for contaminants in sediments, locations for temporal changes in contaminant sources, sinks, and transport paths, and a geochemical context in which to interpret biological and ecological information. This effort has also served to test new approaches for retrieving and utilizing heterogeneous sources of historical sediment chemical and geological data (utilizing large but poorly documented volumes of published and unpublished data has been regarded as difficult).

Table 1. Database principles and methods

DATABASE PRINCIPLES AND METHODS

- 1. Cooperative effort:** Cooperation among agencies and organizations utilizes local expertise, shared effort and resources. The Boston Harbor Contaminated-Sediment Database cooperation includes the US Geological Survey, Mass. Water Resources Authority, U.S. Environmental Protection Agency (Region I), and The U.S. Army Corps of Engineers, New England District (Waltham, MA) and interdisciplinary specialists in our and other scientific organizations.
- 2. Decentralized data entry and preliminary processing:** Use widely accessible desktop computer hardware and software, especially spreadsheets and database management software.
- 3. Data dictionary:** Mutually acceptable definitions and a comprehensive data dictionary are required to meet the needs of multiple cooperators for data fields, documentation, and quality assurance. Formats and scope must be flexible to meet new needs.
- 4. Validation and quality control:** These are accomplished by batch screening, including mapping techniques. The database retains comprehensive original and derived data from many heterogeneous sources and lets users determine appropriateness of specific analyses for their purposes.
- 5. Scientific, rather than archival objectives:** Display and interpretive mapping techniques are in active flux. Maps generated from GIS systems (Arc Info), AVS, USGS MIPS systems, and MacIntosh plotting programs are displayed here.
- 6. Data dissemination:** Accessibility and active use by scientists and managers are primary goals. The database, ancillary data, maps, and viewing and mapping tools in a variety of interactive, user-friendly, and readily exchangeable forms such as CD-ROM, diskettes, FTP, and paper copy.

Database construction

Task sharing among participants is an integral aspect of our efforts. These tasks include collaboration and identification of data sources, construction and quality assurance of the database, database documentation, acquisition of bathymetry and coastline data, mapping and data display, analysis of existing samples, connections to related databases, data analysis and synthesis, and utilization of the database and scientific findings generated by it. Construction of the database is the most labor intensive stage of this project. Existing compilations and data must be acquired, data entered from the many heterogeneous sources, it must be scientifically edited to assure quality of the documentation and data, and the data must be integrated into a cohesive format. A database is widely useful though, only if it is accessible and comprehensive.

Data dictionary

The database has a data dictionary (Figure 6) of parameters, definitions and formats for the contaminants and associated information. The data dictionary explains the structure and content of the database (Figure 7). The simplest display of the database is a flat matrix where each sample is a row and the columns contain location information, analyzed parameters and observations. It includes station and sample information, metal and other inorganic data, general and specific organic contaminants, sediment grain size and other physical properties, references, and a working glossary. There is generous use of qualifiers and comments so that the compilation is comprehensive and no information is lost. Additional fields can be added as necessary and few formatting limits are required. The paper-trail information and format accommodates most chemical and physical data and the completeness improves the value of the database to future users. The structure allows for flexibility, growth and ease of data exchange.

We work with other agencies towards having nationally acceptable "universal" fields and formats that are inclusive yet flexible. We actively pursue collaboration with other agencies and scientists in promoting regional and national efforts in assessing the distribution, transport paths, and impact of contaminants in our coastal environments. We have developed a training program for the data screening and entry needed to construct the databases and present our methods and scientific interpretations at conferences and in publications.

Data products

The products generated are data access and synthesis at both local and regional levels. Foremost is a comprehensive, validated database of contaminated sediments and sediment properties for the region of interest, e.g. Boston Harbor or the Gulf of Maine. Associated with the database is the documentation provided in a data dictionary, bibliography, user guide, and statistical analysis of the data contained in the database. Then there are interpretive text and maps, such as two and three dimensional maps showing the distribution and concentrations of contaminants, bathymetry, shorelines and sedimentary environments. We also are developing and will distribute software tools for data analysis. The ultimate goal of the database effort is help provide another tool, the database, that will help the community address specific scientific questions.

We rely on multiple distribution routes to insure access of many to the database. Various routes that are evolving are floppy disk, CD-ROM, hard copy, FTP, and inclusion in existing national data archives. We also are working out the bugs that frequently are associated with cross-platform transfers or translations so that the data can be obtained in a variety of widely-used formats.

Database distribution and access

We have found that translation difficulties often prevent data exchange. One example of this is the difficulties that arise moving information about contaminants in sediments between different databases. To remedy this situation, we have created a concordance of fields in the data dictionary for our Contaminated-Sediment database (Fig. 8 and Table 2). This concordance shows the names used for equivalent fields in more than eleven major sediment databases used around the country.

Table 2. Data parameter concordance

USGS Contaminated -Sediment Database Dictionary includes concordance of fields	
RASS-PLUTO	ODES
EMAP	WATSTOR
STORET	NWIS
WASH. STATE DEPT. ECOLOGY	BEDFORD INSTITUTE
USACE-NED	NGDC
MWRA	

Translation difficulties also frequently occur across platforms. Again the Contaminated-Sediment database is one example of our efforts to work around these barriers:

- We use a **flat spreadsheet file as the least common denominator** and can convert it to most spreadsheet types, or import to database managers when complex manipulation is needed.
- **Field names are provided a choice of formats and with database concordances.** A choice of 10, 23, or unlimited character format meets most application needs.
- We use **inexpensive commercial or public software** that has shorter learning curves and is more widely accessible than many of the mainframe database managers or dedicated database programs.
- We use **desktop computers (DOS, WINDOWS, and MACINTOSH³)** so that products will be available to a greater number of users throughout the community. A network or mainframe is not required but can be used.

Other sticky issues arise when dealing with database construction, particularly when the data is from heterogeneous sources. For the Contaminated-Sediment Database, we deal with some of these issues in the following manner:

- **Data quality** is assured by 1) maintaining liberal qualifier and comment fields to preserve the full information content of the original data set and let user decide appropriateness for purpose intended and 2) subjecting the compiled database to VALIDS scientific editing, which flags questionable, redundant, or missing data.
- **Acknowledgment** of original references and sufficient information to locate sources is accomplished by maintaining a paper trail to publications, sponsors, researchers and analysts within the database.
- Compilation from **heterogeneous data** and comparison between them is accomplished because 1) the data dictionary provides precise definition of fields and field name concordance and 2) ample comment and qualifier fields allow text elaboration about non-standard items.
- The compilation focuses on contaminants in sediments, however some users may want additional information on biota or water. The **limited scope of the compilation** is expanded by field which note the availability of associated data or information.
- In the Contaminated-Sediment databases for Boston Harbor, Gulf of Maine and other locations, rescue of historical data generated during the last 25 years or so is a one-time effort. **Continuity** into the future will be maintained by continually feeding newly released data into the database and issuing periodic updates.

Validation by Interactive Data Screening (VALIDS)

Quality assurance and validation needs scientific expertise to recognize problems in data quality and address "quality gaps" in the database documentation. Data is available from sources

as divergent as permit applications, compliance records, surveys, and academic research. The degree of documentation, levels of detection, and overall quality of the analysis also vary widely. Despite the high degree of heterogeneity, we have found that scientific editing and thorough documentation allow construction of a very usable database. Currently, there are more than 2,300 sediment samples from Boston Harbor and nearby Massachusetts Bay in our database. By far the largest number of problems in data quality arise from gaps in documentation, especially location, and in clerical errors. Such errors can escape some standard quality control procedures and create major bias in data synthesis; however, they are relatively easy to detect using the Validation by Interactive Data Screening (VALIDS) technique, which uses sorting and graphics techniques to identify outliers, inconsistencies and other data that needs further inspection. These entries and their references can then be checked in detail and editorial comments attached. An important feature of the database is that it provides sufficient information with the data for the user to make her own decisions about the appropriateness of specific analysis for a given purpose and maintains a "paper-trail" so that the user can return to the original reference or analytical lab for clarification when needed. Figure 9 illustrates the editing procedure in a spreadsheet format as a sort on Cd or Hg by descending values. Values are tagged that have questionable data, questionably high concentration values from a single data source or lack locations. Figure 10 illustrates graphical editing procedures where questionable information is identified by plotting of unreasonable locations or concentrations.

Table 3. Batch validation procedure using VALIDS

BRIEF OVERVIEW OF BATCH VALIDATION PROCEDURES

Location reasonable: a. Plot points - do they fall within designated area?

b. Sort on latitudes and longitudes or state plane coordinates to check for sample location and designation discrepancies

Paper trail: Sponsoring, operating agencies, laboratories, dates, source references; if missing note authors or identified individuals, agencies for further search check.

Depth in sediment, sediment description and variability; note possible availability of additional data (e.g. texture) not given in report.

Questionable sampling, parameters, procedural details, e.g. ignition loss as a measure of organic content (clays generally contribute greater loss). Flag originating organization for examination of other data.

Concentration values: reasonable range, units? Anomalous relationships, e.g. some contaminant constituents in unpolluted range, others in toxic range; toxic-range constituents in sediments from cored strata supposed to be unpolluted - sign of sampling, lab contamination or systematic error?

Cross plotting & comparison: e.g. values from same region but different data sources; comparison with well-qualified data set; anomalous values vs. data sources, color-coded mapping of values - check anomalies.

Discussion: Validation is an iterative process and continues from prescreening through the display and synthesis stages. When many data are available one may more readily exclude questionable items from specific applications, using specific criteria to do so.

Inclusions vs. exclusion: it is important to retain and properly label questionable data, rather than eliminating them from the database. Why? 1) Criteria for acceptability may vary with user; 2) if suspect data are excluded future database users may add them to an expanded data set or use them independently, not realizing they have already been considered and questioned.

Application

The Contaminated-sediment database is used to study the distribution, transport, and fate of contaminants in the marine system. Histograms and maps illustrate how use of the database provides information about distribution and processes of contaminants that would be difficult to identify with the smaller datasets available from any individual study. The compilation allows identification of components or regions where samples are sparse (Figure 11). A histogram of all the samples for any given parameter (e.g., Figure 12) shows the scale of variability, which further inspection of the data can identify as due to differences in proximity to a source, geochemical processes such as variations in affinity for the contaminant to the sediment, variability in sediment composition, deposition or transport, or analytical uncertainty. Concentration values have a log-normal distribution. This means that a few erroneous values will not effect interpretations based on the average behavior as long as the measured average is representative of the true behavior. It also suggests that care should be taken to obtain representative results when selecting monitoring sites, monitoring contaminant levels and setting regulatory standards for near-shore waters and fisheries. For Boston Harbor, this compilation of data from over 500 references shows the relative proportion of samples having possible acute toxic side effects. Histograms are color-coded to show "effects-based toxicity levels"⁴ published by Long and Morgan (1990). More than 50% of the analyzed samples for at least six metals fall above the lowest acute toxicity threshold (Manheim and Hathaway, 1991).

Assessment of the fate and persistence of contaminants in sediments must consider the composition, texture and sedimentary environment of the host sediment. Maps of the distribution of samples and concentration values (Fig. 13-15) help elucidate the transport paths of contaminants in the environment and identify heavily impacted areas. The concentration levels are patchy, reflecting variability in sedimentary environments (Knebel et al., 1991) and water depths. Contaminant accumulation can be focused in depositional areas that are a few meters away from rocky or sandy regions of little accumulation. The highest concentrations of metals in surface sediments occur where high sediment deposition rate coincides with a nearby source such as a combined storm-sewer outfall. The region having highest concentrations of contaminants is the Boston Inner Harbor, with values decreasing in the Outer Harbor, and towards Massachusetts Bay.

Integration

The contaminant data must be put in the context of the geological and physical features which effect the distribution and transport of the contaminants. Superposition of the contaminant map on maps of sedimentary environments (Fig. 16) or bathymetry (Fig. 17) is a helpful technique for analysis and display. Detailed bathymetry is necessary to interpret the sediment data and predict the transport of contaminants. The USGS is working with NOAA and state agencies to obtain good digital bathymetry for Gulf of Maine coastal regions. New visualization and data manipulation programs, both in-house ones, and commercial ones (e.g., GIS and data visualization applications), are also making it easier for scientists and resource managers to link parameters.

Summary

The USGS Branch of Atlantic Marine Geology

- generates diverse types of marine data and information
- is committed to providing earth science information and data to the community
- interacts with and exchanges data with other agencies at the national and local scale
- is currently exploring and testing new modes of data access and distribution
- our functional model for databases is:
distributed databases

centralized (local) archives
QA/QC by complete documentation
multiple distribution platforms
collaboration at all levels

The Contaminated-Sediment Database

- Is an example of one of the databases resident at the USGS
- Is a user-friendly archive of chemical data
- Successfully integrates heterogeneous data and provides QA
- Relies on and promotes collaboration and cooperation
- Assures access by many through distribution via multiple platforms (with tools)
- Provides a valuable scientific tool to the GOM community to address questions and concerns about contaminated sediments in estuarine and coastal waters

Footnotes:

¹The U.S.G.S. work complements and is closely coordinated with other research and monitoring studies in the area, including those supported by the Massachusetts Environmental Trust, the Massachusetts Bays Program, the Massachusetts Water Resources Authority, the U.S. Environmental Protection Agency, Region 1, and the New England Army Corps of Engineers.

²Primary collaborators in construction of the Gulf of Maine Contaminated sediment database are Steve Jones and Larry Ward (Univ. New Hampshire), Peter Larson (Bigelow Marine Lab), Gordon Wallace (U Mass Boston), John Farrington and Bruce Tripp (Woods Hole Oceanographic Institution) and Bill Hubbard (Army Corps of Engineers).

³ Use of any trade names in this document does not constitute endorsement of any particular product by the USGS or US government.

⁴ER-M is the metal concentration level at which 50 percent of the test sediments are found to have consistent adverse effects on biota. ER-L is the metal concentration level at which 10 percent of the test sediments are found to have consistent adverse effects on biota. These values are considered screening levels and are subject to revision.

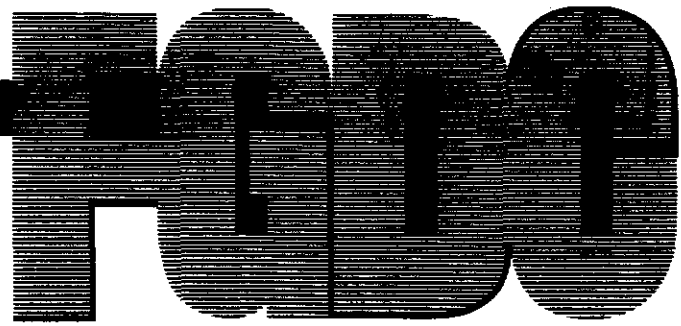
References:

J.C. Hathaway, 1993. Atlantic Margin Sediment Texture Database., U.S. Geological Survey Open File Report 93-xxx, Woods Hole, MA.

H.J. Knebel, R.R. Rendigs, and M.H. Bothner. 1991. Modern sedimentary environments in Boston Harbor, Massachusetts. *J. Sedimentary Petrology*, 61, 791-804.

E.R. Long and L.G. Morgan. 1990. The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program, NOAA Technical Memorandum NOS OMA 52.

F.T. Manheim and J.C. Hathaway, 1991. Polluted sediments in Boston Harbor-Massachusetts Bay: Progress report on the Boston Harbor Data Management File. U.S. Geological Survey Open File Report 91-331, Woods Hole, MA.



A Publication of the Federal Geographic Data Committee

Building the National Spatial Data Infrastructure

NSDI

Ask 20 people to define the national spatial data infrastructure (NSDI), and you are likely to get 20 definitions that vary in scope, application, responsibility, and format. The Federal Geographic Data Committee (FGDC) is hoping to build consensus on what the NSDI is and why it is needed.

The FGDC envisions that the NSDI consists of organizations and individuals who generate or use spatial data, of the technologies that facilitate use and transfer of spatial data, and of the actual data. The reasons we need the NSDI are to find, retrieve, create partnerships to collect, and make effective use of spatial data.

If you are reading this newsletter, you likely are a node in the NSDI. On a basic level the infrastructure consists of individuals who are interested in geographic or spatial data: data users and producers. A critical first step in developing the infrastructure is establishing the ability to communicate and building links among individuals and organizations within the spatial data

in using the data. A draft metadata standard for consistency in documentation is being reviewed (see accompanying article). If members of the spatial data network can agree to use standards such as this or the Spatial Data Transfer Standard (SDTS) for data transfer, building another tier in the NSDI becomes possible. We can move one step closer to finding and retrieving data.

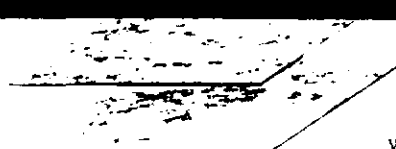
A third step in infrastructure development is establishing partnerships for transferring data or for developing common data bases. This step builds on previous actions of communication and agreement to use standards. For many individuals or agencies, agreeing to create and use common spatial data sets is a major achievement. Partnerships to exchange (sell, trade, or share) data are common; partnerships to generate data are rare. Will the NSDI ever be one, agreed-upon set of spatial data representations of the Nation (for example, one road network) to which we can all attach our own attributes for analysis?



FOR MANY INDIVIDUALS OR AGENCIES, AGREEING TO CREATE AND USE COMMON SPATIAL DATA SETS IS A MAJOR ACHIEVEMENT.

community. The most technically sophisticated can communicate electronically using vehicles such as Internet or proprietary networks within agencies. Others can communicate by telephone or newsletters or conferences. Regardless of the means, the key is that knowledge about data availability is proliferating because communications are improving.

A second step in developing the infrastructure recognizes that data sharing can be facilitated through the use of common procedures and standards. Metadata, or data about data, are critical first to determine that specific data exist, and second to understand the nature and quality of the data to determine whether they are useful in a specific application. Consistency in how, when, and where to document data will assist others



Could this be a data base for which we know the inherent accuracy of each coordinate and can select resolutions as we see fit? Or is the NSDI really a large collection of multiple representations of spatial reality that we hope to understand sufficiently to choose the representations that meet our needs? Probably the NSDI partly encompasses both these concepts, and this point will be debated in many future forums. But the fact is that we and the data we collect are all components of the NSDI. Many of us have come to realize the benefits of having more cohesiveness within our spatial data community, of building partnerships, making agreements, and communicating what we know. We can all share the success of developing the NSDI.

Figure 1. Newsletter from the Federal Geographic Data Committee (FGDC) describing the National Spatial Data Infrastructure. The first steps are to 1) establish communication and build links; 2) facilitate data sharing through common procedures and standards, and 3) establish partnerships for transferring data or developing common databases.

United States Geological Survey

Data Category Keywords <i>From FGDC manual of Federal Geographic Data Products Data Product Page 2 of 3</i>	Page Number	Administrative boundaries	Aerial photography	Airports	Atmospheric	Bathymetry	Boundaries	Census geography	Climats	Coastal	Contours	County boundaries	County land ownership	Demographic	Ecological	Economic	Elevation	Federal land ownership	Flood	Floodplain	Geocetic	Geographic names	Geology	Geophysics	
	Aerial Photographs & Satellite Photographs	63		•																				•	•
Bulletins	66									•														•	•
Coal Investigations Maps (C Series)	67																							•	•
Digital Airborne & Satellite Imagery	68																							•	•
Digital Data Series (DDS)	68																							•	•
Digital Elevation Models (DEMs)	70																•							•	•
Digital Line Graphs (DLGs)	71	•	•							•	•	•	•				•	•						•	•
Dig. Sidescan Sonar Images of the Sea Floor (GLORIA Atlases)	72				•																			•	•
Flood-Prone Area Maps	73																		•	•				•	•
Geographic Names Information System (GNIS)	74																							•	•
Geologic Quadrangle Maps (GQ Series)	75																							•	•
Geophysical Investigations Maps (GP Series)	77																							•	•
Hydrologic Investigations Atlases (HA Series)	79																							•	•
Hydrologic Unit Maps	80																							•	•
Land Use & Land Cover Digital Data	82	•	•							•	•	•	•					•						•	•
Mineral Investigations Resources Maps (MR Series)	83																							•	•
Misc. Field Studies Maps (MF Series)	84																							•	•
Misc. Investigations Series Maps (I Series)	86																							•	•
National Earthquake Information Center Products	89																							•	•
National Uranium Resource Evaluation (NURE) Data	91																							•	•
National Water Conditions (Report Series)	93																							•	•
National Water Summary (Report Series)	95																							•	•
National Water-Use Information (Report Series)	97																							•	•
Oil & Gas Investigations Charts (OC Series)	99																							•	•
Oil & Gas Investigations Maps (OM Series)	100																							•	•
Open-File Reports	101																							•	•
Outcrop photographs	102		•																					•	•
Other Forest Maps	103	•	•		•	•	•				•	•	•	•			•	•						•	•
Other Map Products	104	•	•	•	•	•	•				•	•	•	•			•	•						•	•
Photographic Products from Airborne & Satellite Imagery	105		•																					•	•
Professional Papers	106																							•	•
Quadrangle Forme Maps	107	•	•																					•	•
Quaternary Geologic Atlas of the United States	109																							•	•
Regional Aquifer-System Analysis (Report Series)	111																							•	•
Satellite Image Maps	112																							•	•
State Geologic Maps	113																							•	•
Surface Water Map	114																							•	•
Thematic Maps	115	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Water Data	116																							•	•
Water Supply Papers & Other Book Series	117																							•	•

Figure 2. Page from the FGDC produced Manual of Geographic Data Products. The manual is a useful reference material that cross-indexes subject matter and types of product available from all federal agencies.

Information in the IMS

Information contained: The IMS contains the data, information, and knowledge needed by the Branch. It contains information with a wide range of users, reliabilities, and formats, including:

- a. Original (raw) digital and analog data, and samples, with corresponding ancillary information. This is intended for use within the Branch.
- b. Documented, self-describing, intelligent data sets or information, intended for internal use and release to scientists at cooperating institutions.
- c. Published, verified and refereed information, including highly reliable standard data sets of extensive general use. This is intended for distribution to the scientific community.
- d. Synthesized information produced by the combination of information from different sources and by the combination of information with model outputs, theories, and other forms of knowledge. This information represents our best understanding of the natural world and is used to address a wide variety of critical issues and for distribution to the public.

Information quality: The IMS assures the quality of its information by preserving the full information content of each data set, documenting the reliability and history of all information, and providing security so that information cannot be altered improperly.

Figure 3. Schematic of the data and information and management system (IMS) envisioned by the Branch Information and management Team (BIMT).

USGS-BAMG SEDIMENT DATABASE

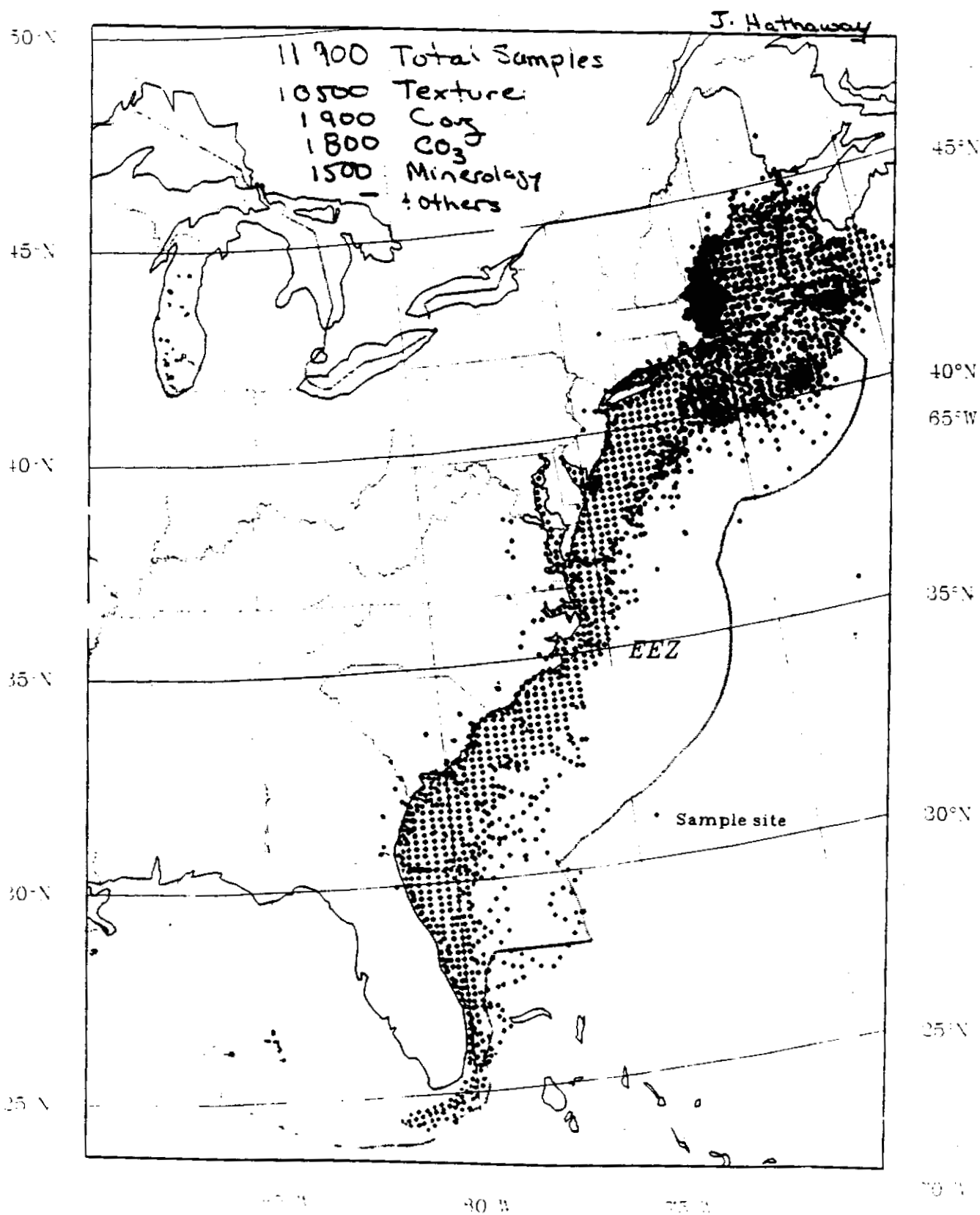


Figure 5. The sediment texture database for the Atlantic margin (Hathaway, 1993) contains data for 11900 samples at the locations indicated and is an example of one of AMG's distributed databases. The data is currently available through Internet via GOPHER or FTP and will soon be available, accompanied by interpretive maps and display tools, in a CD-ROM.

USGS Contaminated-Sediment Data Dictionary

USGS Row #	Short Field Name 10 characters long	CAS NUMBER	Medium length Field Name 25 characters long	Full Length Field Name	Description of Parameter and its Fields
	XXXXXXXXXX		XXXXXXXXXXXXXXXXXXXXXXXXXXXX		
	TABLE OF STATION DATA				
1	LOCAL ID		Local row or ID No	Local Row or local ID Number	ID for use by user in maintaining sample order. This number can be changed by users.
2	UNIQUE_ID		Unique Sample ID	Unique Sample #	Database ID unique to this specific sample; assigned by USGS, this is the common parameter for joining all tables in this database
3	ORIGNL_ID		Preceding database ID	Preceding database sample ID	Database ID unique to this specific sample in the preceding database.
4	REPLCT_NO		Repl. #	Repl. #	A number for a replicate where, for any parameter(s), more than one analysis have been made of the same sample. For the purpose of unambiguous joining of multiple tables, this number is incorporated in the Unique sample identification number
5	ST_PLANE_E		State Plane E	State Plane E	Easting coordinate in the State Plane System for the datum shown in Datum, Position
6	ST_PLANE_N		State Plane N	State Plane N	Northing coordinate in the State Plane System for the datum shown in Datum, Position
7	LATITUDE		Latitude (Decimal)	Latitude (decimal)	Latitude in decimal-degrees (south latitudes are negative)
8	LAT_DEG_N		Lat(Deg)N	Lat(deg)n	degrees of latitude in whole degrees
9	LAT_MIN_N		Lat(Min)N	Lat(min)n	minutes of latitude in whole or decimal units
10	LAT_SEC_N		Lat(Sec)N	Lat(sec)n	seconds of latitude in whole or decimal units
11	LAT_ORIG		Lat Orig. Position Format	Lat Orig. position format and precision	format and precision of originally recorded latitude (indicate format and significant digits, e.g., ddmms.s; dd.dddd; ffff)
12	LONGITUDE		Longitude (Dec.)(- = W)	Longitude (decimal) (- = West long.)	Longitude in decimal-degrees (West longitude is depicted by negative values.)
13	LON_DEG_W		Lon(Deg)W	Lon(deg)w	degrees of longitude (West) in whole degrees
14	LON_MIN_W		Lon(Min)W	Lon(min)w	minutes of longitude (West) in whole or decimal minutes
15	LON_SEC_W		Lon(Sec)W	Lon(sec)w	seconds of longitude (west) in whole or decimal units
16	LON_ORIG		Lon Orig. Position Format	Lon Orig. position format and precision	format and precision of originally recorded longitude (west), eg. ddmms.s; dd.dddd; fffff
17	ORIG_LOC		Orig. Locn. if Unusual	Original location in orig. format, if unusual	Original location if given in units other than latitude and longitude (e.g., State Plane or distance from a point)
18	DATUM_POS		Datum, Pos.	Datum, pos.	North American Datum (NAD) for the state plane position coordinates given
19	DATUM_ELEV		Datum, Elev.	Datum, elev.	Datum standard for determining lat/lon from location map
20	NAV_MODE		Navigational Mode	Navigational mode	Navigational system used, e.g. LORAN C, GPS, triangulation, read from a sketch map
21	TIME_DELAY1		Time Delay 1	Time Delay 1	Navigational time delay (first value) used to calculate position from signals
22	TIME_DELAY2		Time Delay 2	Time Delay 2	Navigational time delay (second value) used to calculate position from signals
23	SOUNDING_M		Sounding (M)	Sounding (m)	Measured depth of water overlying sediment at sample time, in meters
24	SNDNG_ORIG		Sounding (In Orig Units)	Sounding (in original units, if not m)	Measured depth of water overlying sediment at sample time, in original units
25	SNDG_UNITS		Sounding Original Units	Sounding original units	Depth units (meters, feet, fathoms, etc.)
26	AGNC1_SPON		Agency1 (Sponsoring)	Agency1 (Sponsoring)	Agency or researcher sponsoring or publishing the work, see listing for abbreviations
27	AGNC2_SPON		Agency2 (Sponsoring)	Agency2 (Sponsoring)	Agency or researcher doing the sampling or research

Figure 6. A portion of the Contaminated-Sediment data dictionary illustrates choices in parameter names, comprehensive definitions, and flexibility.

Figure 6. A portion of the Contaminated-Sediment data dictionary illustrates choices in parameter names, comprehensive definitions, and flexibility.

Station and miscellaneous information

<i>entry identification</i>	<i>sampling</i>
Local Row or local ID Number	Samp. Day1
Unique Sample #	Mo1
Preceding database sample ID	Year1
Repl. #	to Samp. Day2
<i>location information</i>	Mo2
State Plane E	Year2
State Plane N	Date q
Latitude (decimal)	Hour1
Lat(deg)n	Min1
Lat(min)n	Samp. Date 1 formatted
Lat(sec)n	to Samp. Date 2 formatted
Lat Orig. position format and precision	Tide
Longitude (decimal) West	Sample ID or original Sample Field Number
Lon(deg)w	Cruise ID
Lon(min)w	Orig. sta. #
Lon(sec)w	Core Or Grab #
Lon Orig. position format and precision	Sampling Device
Original location in orig. format Datum, elev.*	Sample type
Navigational mode	Depth In Core or sediment, when interval not given
Time Delay 1	Depth interval, TOP of Core or sample (cm)*
Time Delay 2	Depth interval, BOTTOM of Core or sample (cm)
Sounding (m)	Original Depth in sediment in original units if not cm
Sounding (in original units, if not m)	Original depth units
Sounding original units	Sediment depth comments
<i>paper-trail</i>	Sediment depth code
Agency1 (Sponsoring)	Compositing Scheme
Agency2 (Contracted)	General Comments Pertaining To Sample
Agency3 (Subcontracted)	Description/Color
Agency4(Other)	Estimated Volume of material to be disposed
ID No. for source of Information or Reference	Disposal Area Code
Source of Information or Reference	Proposed disposal area
Original database ID	<i>content</i>
Regulatory File Number	Metals and other inorganics analyzed?
Project Name (I.E. Lex Atlantic)	Organic contaminants analyzed ?
State name	Grain sizes analyzed?
Quad name	Bioassay data available?
General location name	Comments-Bioassay
Specific name of location of water body	Bio reference
Area Code	Other types of anal. in ref. but not in this database?
Location Comments	Comments- other analysis
	data entry day
	data entry mo
	data entry year
	entry date, formatted
	initials of data enterer

Figure 7. List of some of the parameters or fields included in the USGS Contaminated-Sediment Database. The specific fields and parameters have been chosen, based on the frequency and format of information typically found in the literature, to avoid ambiguity and to make the process of locating fields, entering data, and validating it as simple yet comprehensive as possible. There are more than 800 fields ; however, fields can be added at any time to encompass additional information or only a subset can be used for data entry from a reference with limited information.

Metal and other inorganics data*entry identification*

Local Row or ID Number
 Cover-Id Unique Sample Identifier (US#)
 Source of Information, Reference, or Ref. No.

Ba (barium) $\mu\text{g/g}$
 Ba q
 Ba det. lim.
 Ba original units

analytical

Metals Testing Lab
 Laboratory's sample ID number
 Laboratory's JOB number (metals)
 Analytical technique (metals)

Ba value in original units
 Be (beryllium) $\mu\text{g/g}$
 Be q
 Be det. lim.
 Ca (calcium) $\mu\text{g/g}$

Analytical comments1 (metals)
 Analytical comments2 (metals)
 Replicate no ___ of n (metals)
 Total replicates n (metals)
 Testing Date (metals)
 Test month (metals)
 Test day (metals)
 Test year (metals)

Ca q
 Ca det. lim.
 Ca original units
 Ca value in original units, if not $\mu\text{g/g}$
 Cd (cadmium) $\mu\text{g/g}$
 Cd q
 Cd det. lim.
 Cl (chloride) $\mu\text{g/g}$

Other inorganic testing lab
 Laboratory's internal ID number (Other inorgs)
 Laboratory's Job number (Other inorganics)
 Analytical technique (Other inorganics)
 Analytical comments (Other inorganics)
 Replicate no ___ of n (Other inorganics)
 Total replicates n (Other inorganics)
 Testing Date (Other inorganics)

Cl q
 Cl det. lim.
 Cl original units
 Cl value in original units, if not $\mu\text{g/g}$
 Co (cobalt) $\mu\text{g/g}$
 Co q
 Co det. lim.
 Cr (chromium) $\mu\text{g/g}$

Test day (Other inorganics)
 Test month (Other inorganics)
 Test year (Other inorganics)

Cr q
 Cr det. lim.
 Cu (copper) $\mu\text{g/g}$

majors and trace metals

Ag (silver) $\mu\text{g/g}$
 Ag q
 Ag det. lim.
 Al (aluminum) $\mu\text{g/g}$
 Al q
 Al det. lim.
 Al original units
 Al value in original units, if not $\mu\text{g/g}$
 As (arsenic) $\mu\text{g/g}$
 As q
 As det. lim.
 Au (gold) $\mu\text{g/g}$
 Au q

Cu q
 Cu det. lim.
 Fe (iron) $\mu\text{g/g}$
 Fe q
 Fe det. lim.
 Fe original units
 Fe value in original units, if not $\mu\text{g/g}$
 Hg (mercury) $\mu\text{g/g}$
 Hg q
 Hg det. lim.
 K (potassium) $\mu\text{g/g}$
 K q
 K det. lim.
 K original units

USGS contaminated sediment database parameter	BEDFORD INSTITUTE	STATE OF WASHINGTON	EMAP	NWISI	WATSTOR	STORET
Longitude (decimal) (-) = west long.		N				
Lon(deg)w	LONG_DEGREE	LON_DEG	EVNTLN			
Lon(min)w	LONG_MINUTES	LON_MIN	EVNTLN			
Lon(sec)w		LON_SEC	EVNTLN			
Lon Orig. position format and precision						
Original location in orig. format, if unusual						
Datum, pos.						
Datum, elev.						
Navigational mode						
Agency1 (Sponsoring)		AGENCY				
Agency2 (Contracted)					COLLECTING AGENCY	COLLECTING AGENCY
Agency3 (Subcontracted)					ANALYZING AGENCY	ANALYZING AGENCY
Agency4(Other)						
Source of Information, Reference, or Reference ID No.						
Original database ID						
Regulatory File Number						
Project Name	PROJECT		PROJECT			
State name		STATE				
Quad name						
General location name	GEOGRAPHIC AREA					
Specific name of location of water body	GEOGRAPHIC AREA					
Area Code 2						
Location Comments						
Analytical technique (Other inorganics)			ANALMETH			
Analytical comments (Other inorganics)						
Replicate no ___ of n (Other inorganics)						
Total replicates n (Other inorganics)						
Ag (silver) ug/g				SILVER	SILVER	SILVER BOT
Ag q						
Ag det. lim.						
Al (aluminum) ug/g				ALUMINUM	ALUMINUM BOT MAT	ALUMINUM BOT MAT
Al q						
Al det. lim.						
Al original units						

Figure 8. Examples from the concordance (DRAFT) of fields for some national databases.

Boston Harbor metals sorted by descending mercury values

Unio Sample Id	Latitude	Longitude	Cd µg/g	Cr µg/g	Cu µg/g	Hg µg/g	Pb µg/g	Zn µg/g
US00004	42.38600	-71.04611	0.5	0.8	80.0	5760.0	1200.0	465.0
US00005	42.34222	-70.97417	4.0	257.0	64.0	4730.0	4.3	117.0
US00002	42.38800	-71.04611	22.5	1.9	260.0	3160.0	400.0	29.5
US00009	42.38500	-71.04611	4.0	1.2	200.0	2350.0	505.0	7.5
US00690	42.38600	-71.04611	485.0	2.4	250.0	1980.0	1100.0	86250.0
US00695	42.38500	-71.04611	75.0	1.4	47.0	1140.0	450.0	175.0
US00339	42.38417	-71.04722	1.1	57.8	68.8	68.8	151.2	264.0
US00192	42.36750	-71.04361	7.4	250.0	460.0	26.0	400.0	800.0
US00208	42.38667	-71.04028	3.0	97.0	1650.0	23.0	2000.0	7200.0
US00183	42.35833	-71.05000	6.2	120.0	480.0	20.0	230.0	1100.0
US01739	41.66667	-70.18333	2.4	85.9	391.0	18.9	107.5	189.4
US00308	42.34472	-70.96528	2.5	136.0	75.0	12.0	77.0	210.0
US01662	42.20000	-70.71667	1.7	164.8	83.3	10.8	141.9	159.3
US01740	41.66667	-70.18333	1.9	73.0	188.3	10.2	126.0	143.6
US01741	41.66667	-70.18333	3.0	67.8	349.0	9.9	87.3	151.2
US00406	42.30833	-70.97639	3.0	302.0	192.0	9.4	234.0	242.0
US00317	42.35278	-71.05139	6.8	195.0	218.2	9.4	500.0	595.0
US00209	42.31806	-71.02167	3.2	240.0	110.0	8.0	142.5	190.0
US00416	42.34306	-71.01389	2.8	133.0	81.0	6.8	84.0	160.0
US01332			11.0	451.0	338.0	6.8	222.0	410.0
US00419	42.35833	-70.98472	2.8	134.0	292.0	6.8	84.0	160.0
US00064	42.30722	-71.03694	6.3	128.0	108.0	6.7	117.0	183.0
US01331			14.0	653.0	649.0	6.6	337.0	583.0
US00182	42.35750	-71.05000	7.2	250.0	390.0	6.5	310.0	950.0
US00420	42.36628	-71.04861	8.8	418.0	339.0	6.3	402.0	506.0
US00414	42.33472	-70.96972	4.4	255.0	239.0	6.1	215.0	317.0

Figure 9 Spreadsheet illustrating VALIDs editing using sorting. Shaded cells indicate missing location data, unreasonably high concentration values for Hg and Zn, some questionably low values for Cr and Pb. Further inspection of information in the database identified a data entry error and an analytical lab that had recurring high values as contributors to some of these points; appropriate corrections and comments were added to the database.

Mercury in Boston Harbor Surface Sediments

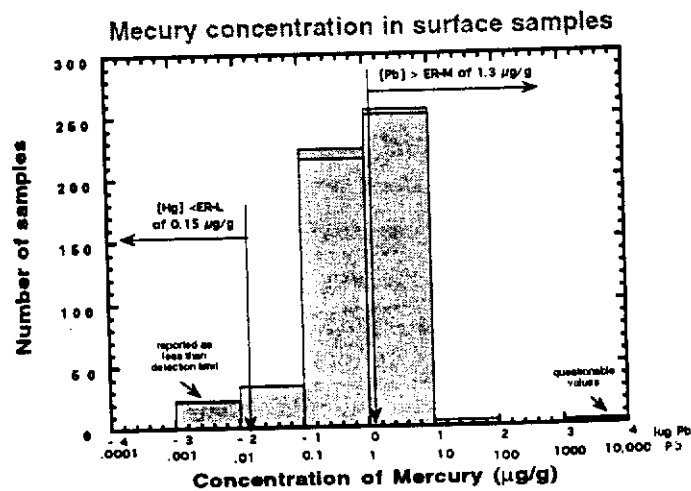
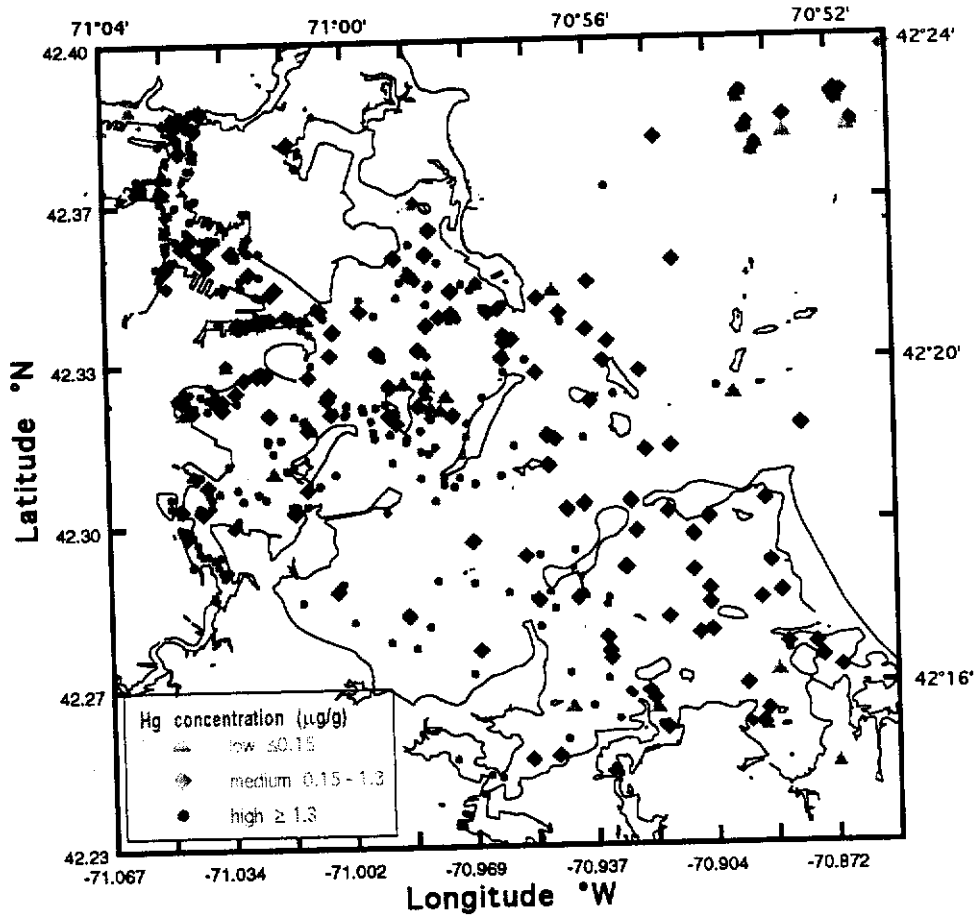


Figure 10a. A map illustrating VALIDS editing using graphical tools points out values that differ greatly from those which are located nearby and samples that are "located" on land. Histograms identify samples having outlying concentration values.

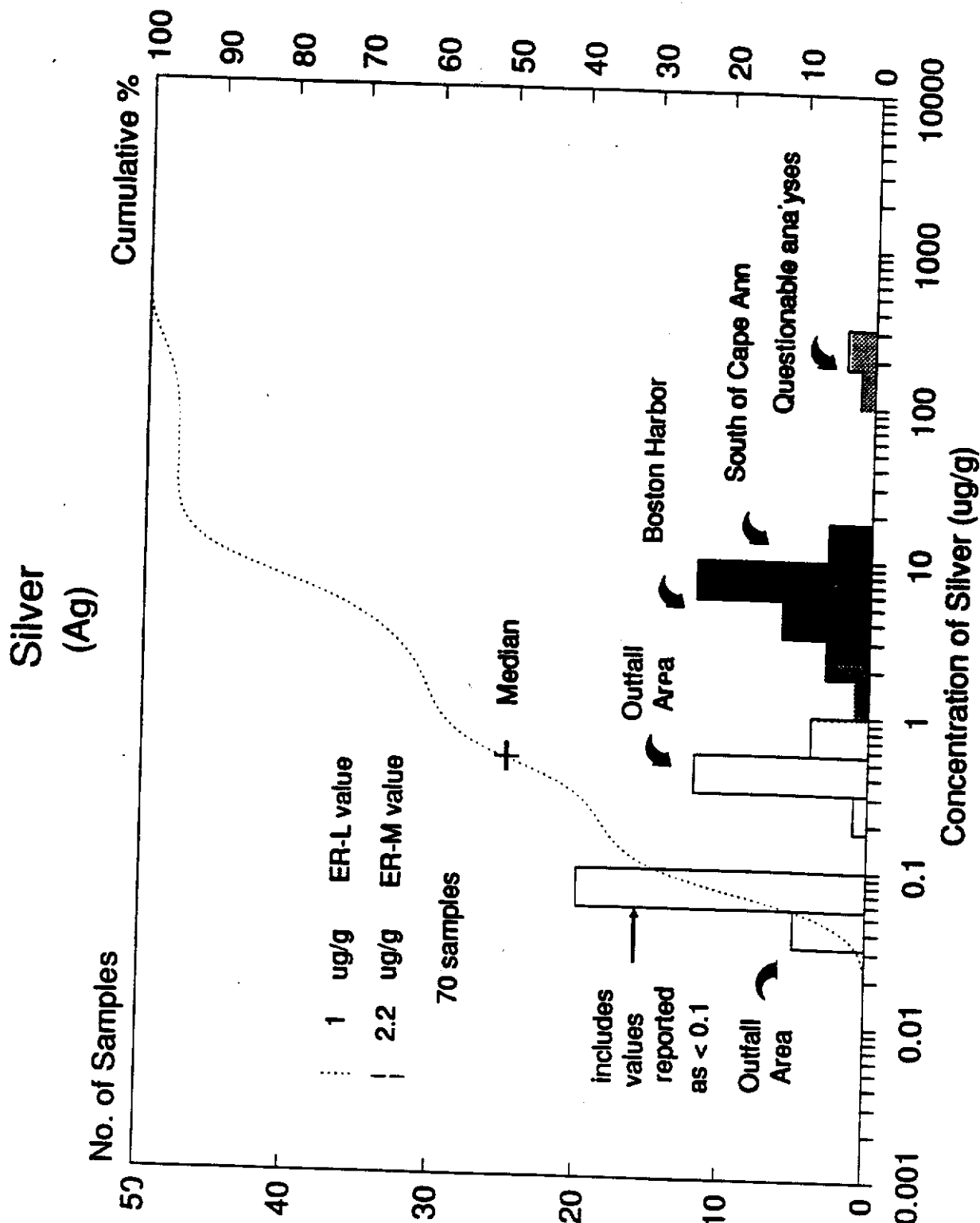


Figure 10b. Histogram illustrating VALIDS editing using graphical tools. Values falling below naturally occurring concentrations, detection limit values improperly reported as sample concentrations, and other outliers are easily identified for further inspection.

Frequency of Metal Values in Boston Harbor-Mass Bay sediment samples

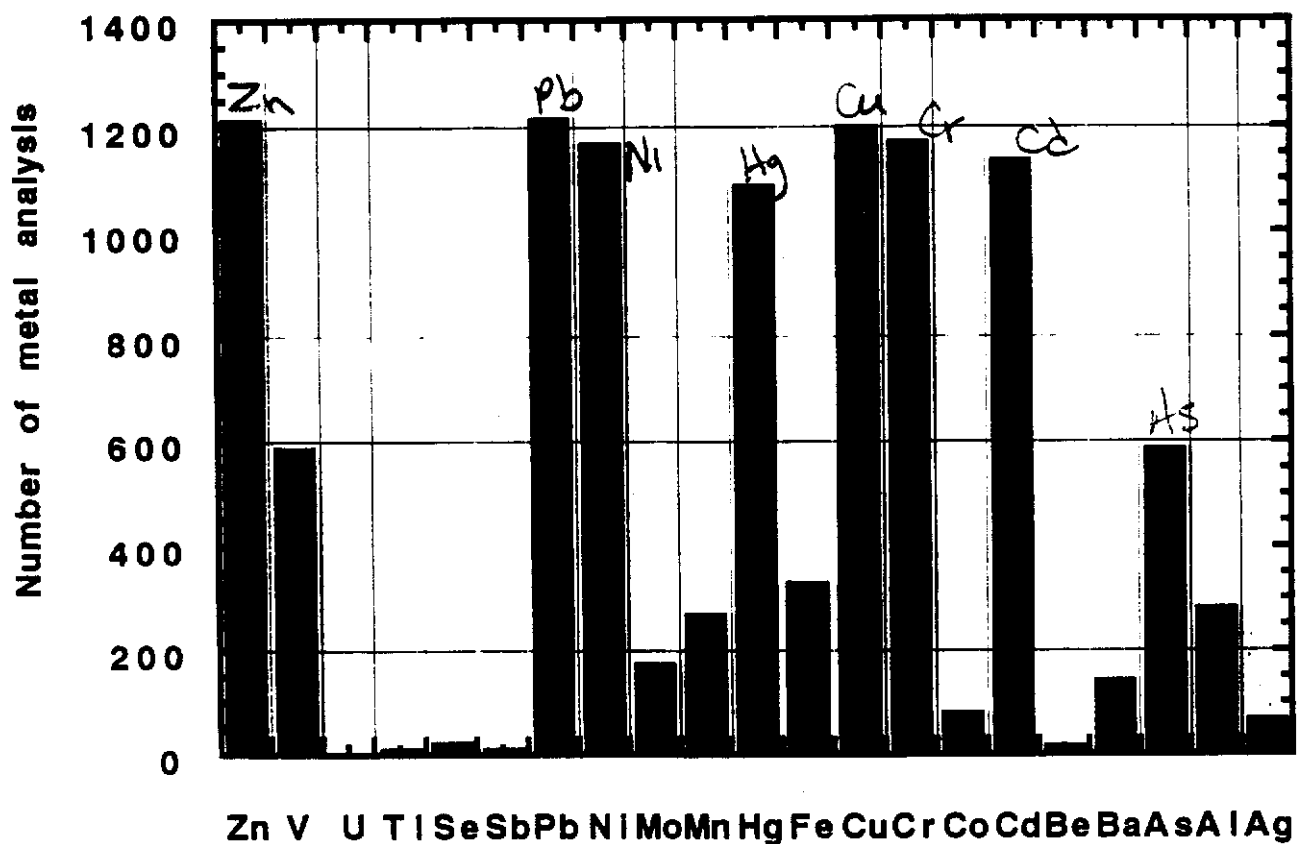


Figure 11. Histogram identifies components or, when done after sorting by geographical area, regions where samples are sparse. The sample density varies greatly for various parameters in the dataset. Coverage is good for V, Cr, Ni, Cu, Zn, As, Cd, Hg, and Pb. Coverage for Fe, a useful geochemical parameter, is poor.

Boston Harbor Sediments ¹

(USGS & EPA Region I Sources)

No. of Samples

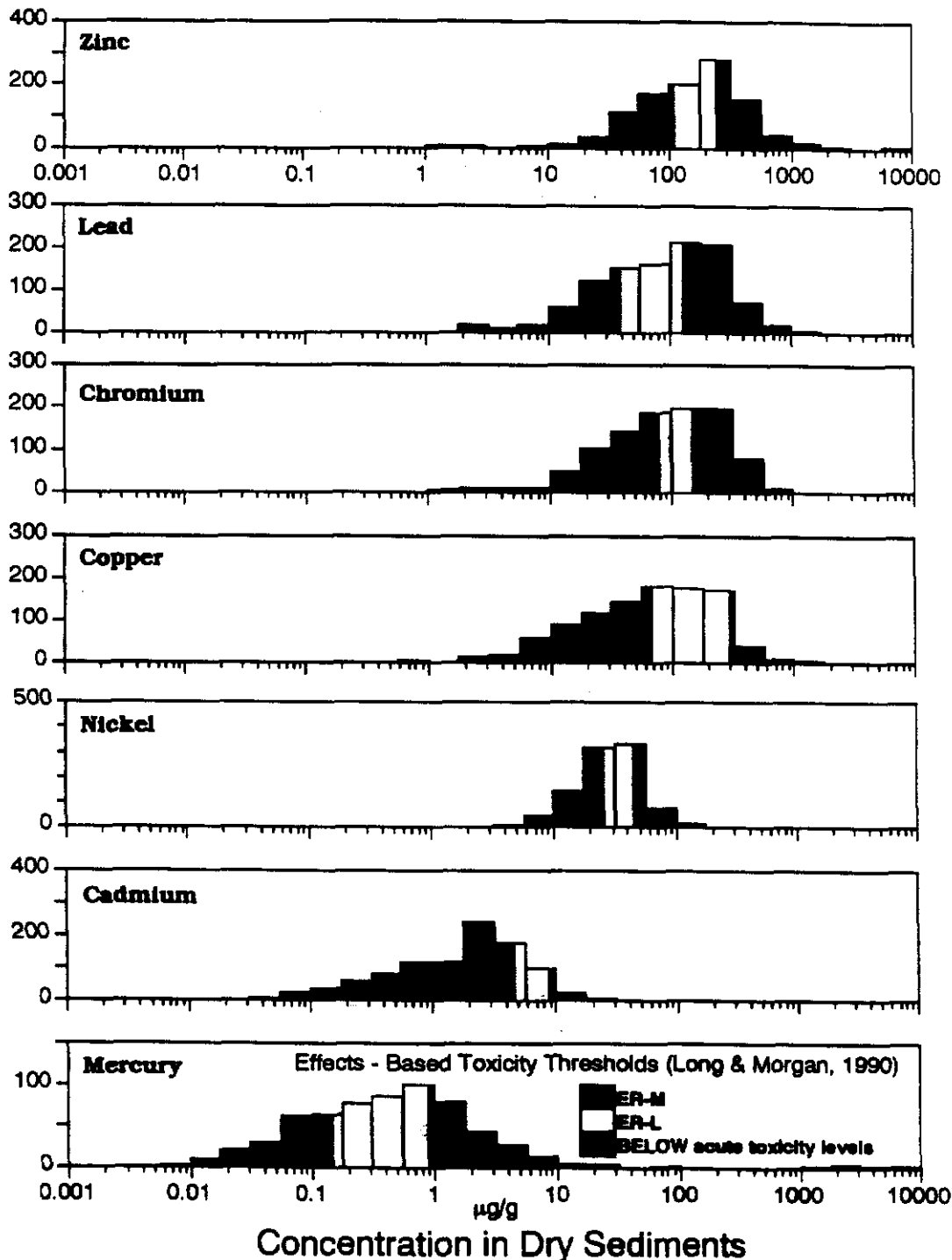


Figure 12. Histogram of all the samples for any given parameter showing spread of data. Concentration values have a log-normal distribution. This means that a few erroneous values will not effect interpretations based on the average behavior. It also suggests that care should be taken to obtain representative results when selecting monitoring sites. More than 50% of the analyzed samples for at least six metals fall above the lowest acute toxicity threshold (Manheim and Hathaway, 1991).

Lead in Boston Harbor Surface Sediments

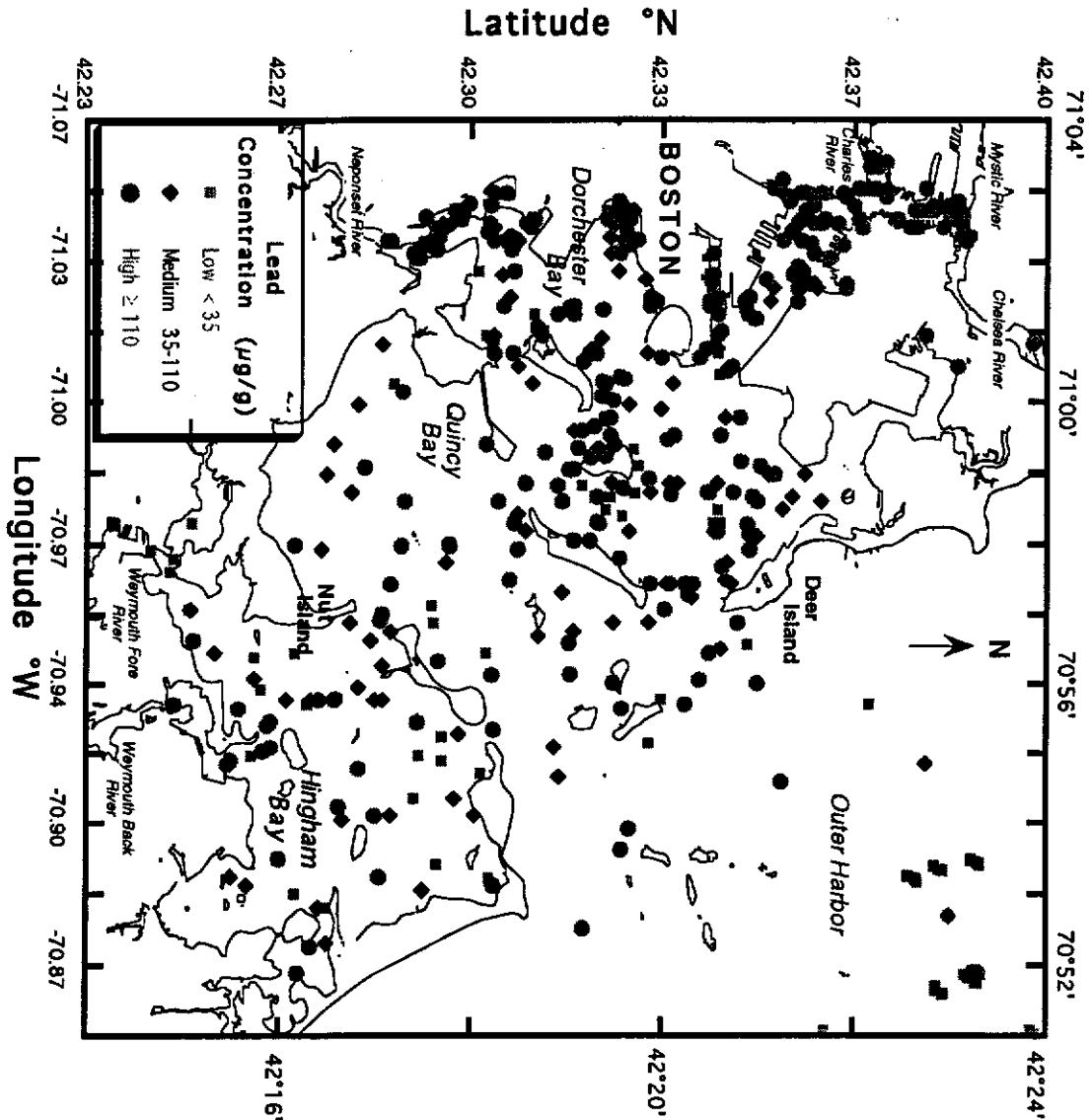


Figure 13. Distribution of samples and concentration values for lead in Boston Harbor. The region having highest concentrations of lead and mercury in surface sediments is the Boston Inner Harbor, with values decreasing in the Outer Harbor, and towards Massachusetts Bay.

Mercury in Boston Harbor Surface Sediments

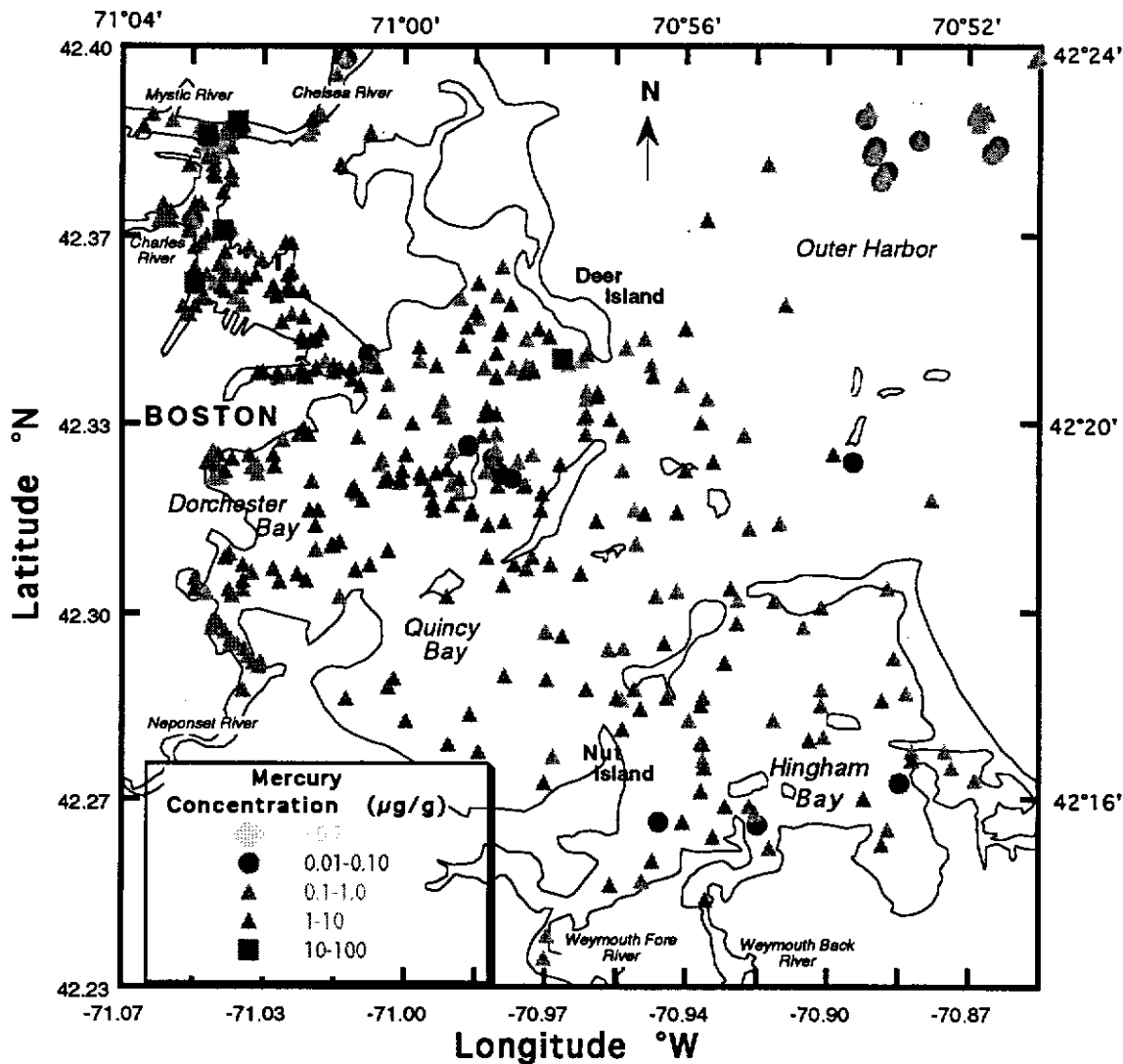


Figure 14. Distribution of samples and concentrations for mercury in Boston Harbor.

Mercury in Mass Bays Surface Sediments

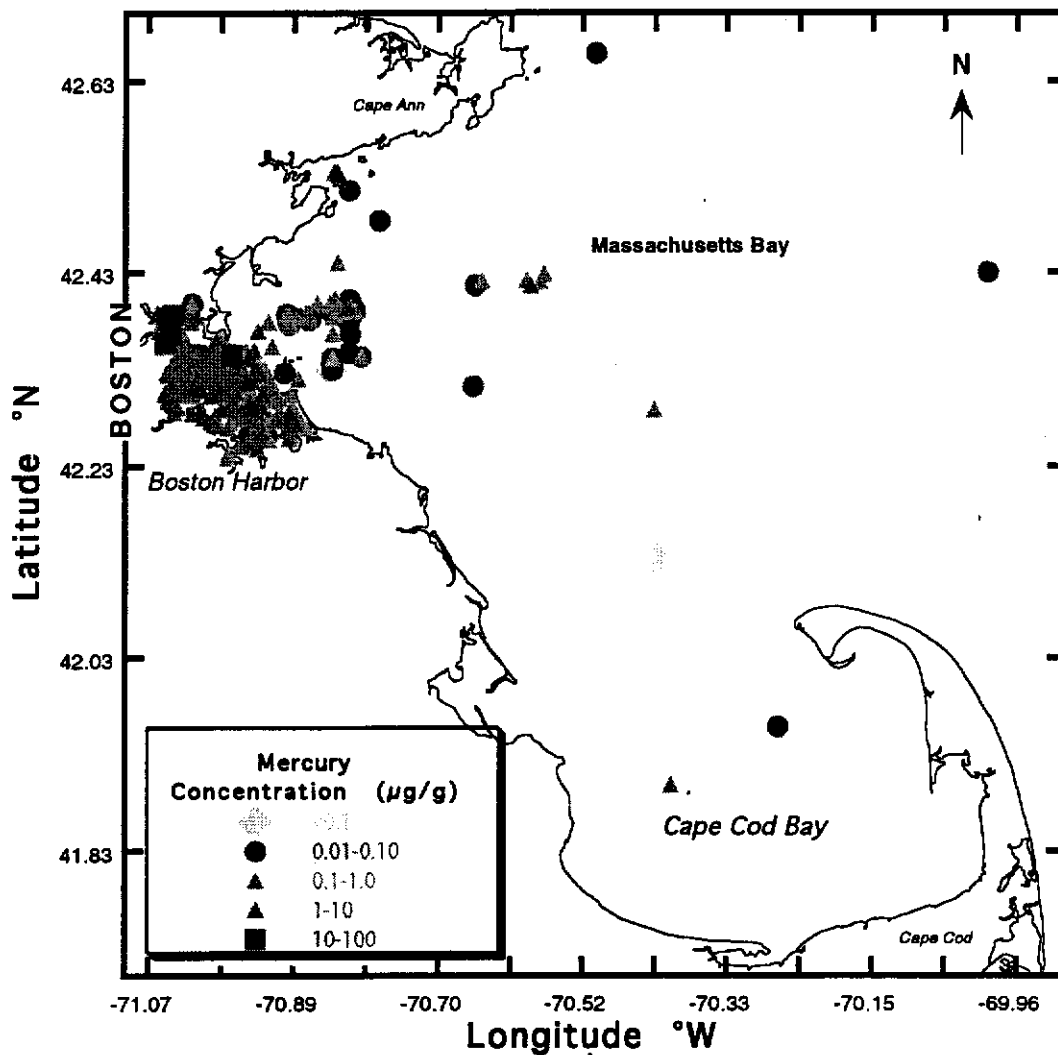


Figure 15. Distribution of samples and concentrations for mercury in the Mass Bays.

COPPER DISTRIBUTION AND SEDIMENTARY ENVIRONMENT IN BOSTON HARBOR

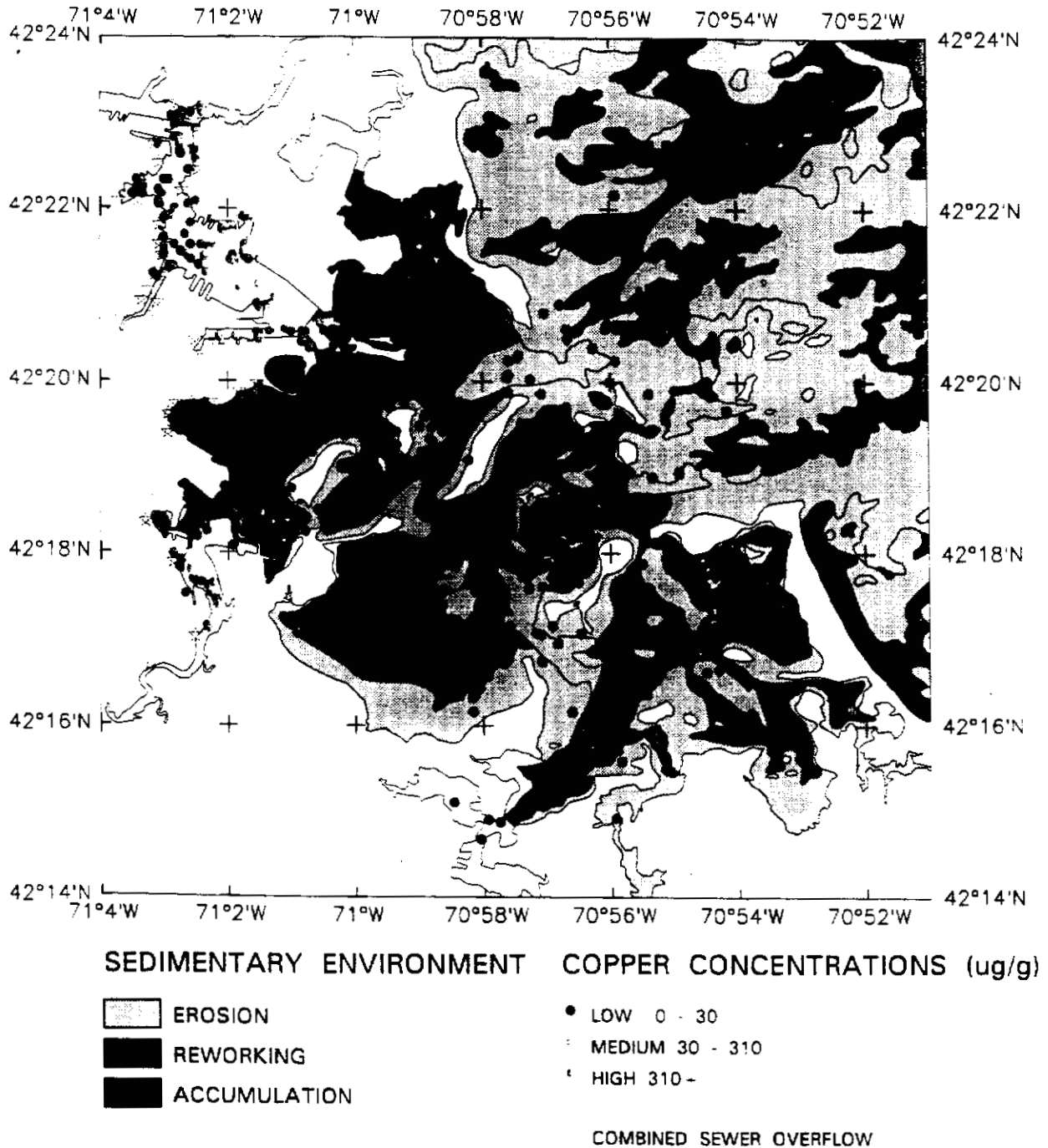


Figure 16. Map of contaminant concentrations with sedimentary environments. The spread and patchiness in the concentration values is partly due to the patchiness in the sedimentary environment in Boston Harbor. Contaminant accumulation can be focused in depositional areas that are a few meters away from rocky or sandy regions of little accumulation.

Boston Harbor Bathymetry

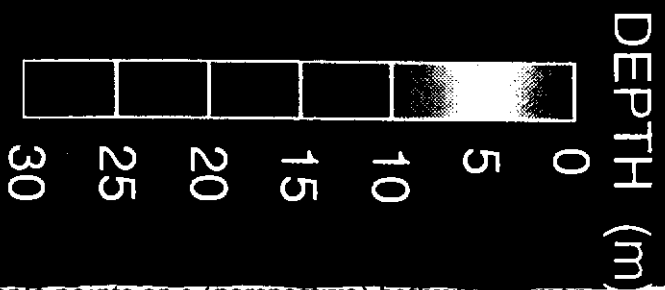
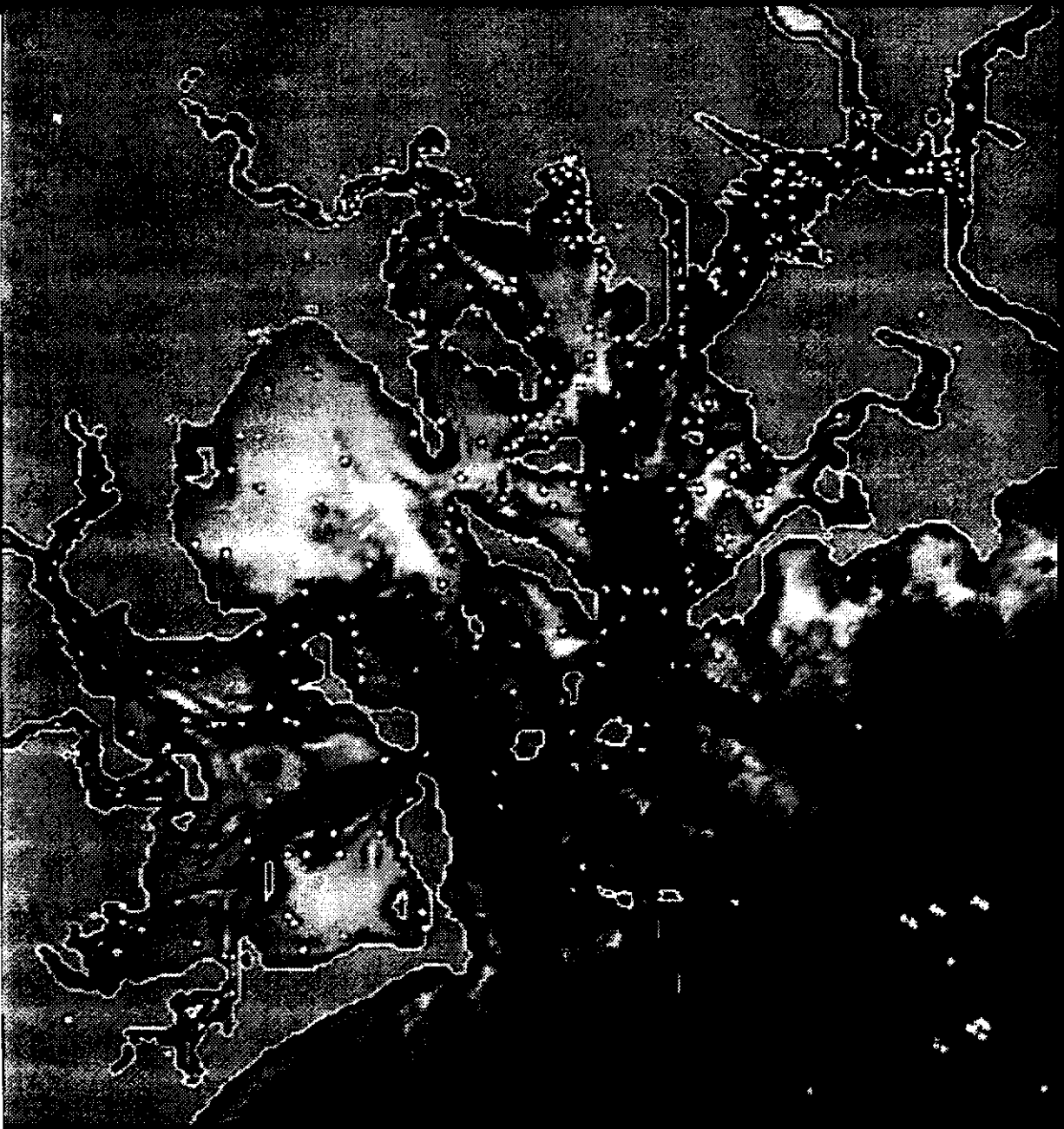


Figure 17. Distribution of sample points on a (perspective) bathymetry map.

EMAP

**An Approach to
Ecological Research, Monitoring
and Assessment**

John Paul
Environmental Protection Agency

EPA EPM-1

What Is EMAP?

- A **Research** Program That Establishes the Scientific Foundation
- A **Monitoring** Program That Provides Data On All Ecological Resources Nationwide
- An **Assessment** Program That Translates Scientific Results Into Information Useful for Decision Making



EMAP Resource Groups

- **Estuaries**
- **Wetlands**
- **Surface Waters**
- **Great Lakes**
- **Agricultural Lands**
- **Arid Lands**
- **Forests**
- **Landscape Ecology**

The EMAP Approach

- **Regional Scope**
- **Ecological/Biological indicators**
- **Probability Based Sampling**
- **Long Term**
- **Interagency Partnerships**

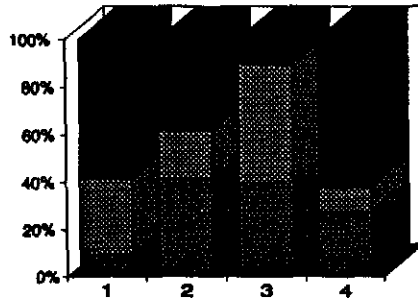
The Twenty Year Test

" Will someone 20 years from now, not familiar with the data or how they were obtained , be able to find datasets of interest and then fully understand and use the data solely with the aid of the documentation archived with the dataset?"

**Solving the Global Change Puzzle,
National Academy Press, 1991.**

NRC Committee on Geophysical Data.

EMAP Products



Annual Statistical Summaries



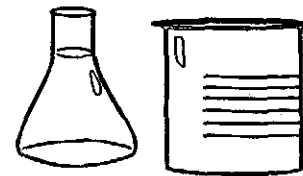
**Ecological Assessments
(Single and Multi-resource)**



**On-line Computer Access
to EMAP Information
Management System**



Ecological Indicators



**Standardized Methods
and Quality Assurance**



Indicators

Response Indicators

- Benthic species composition, biomass and abundance
- Fish species composition, size and abundance
- Fish pathology/histopathology

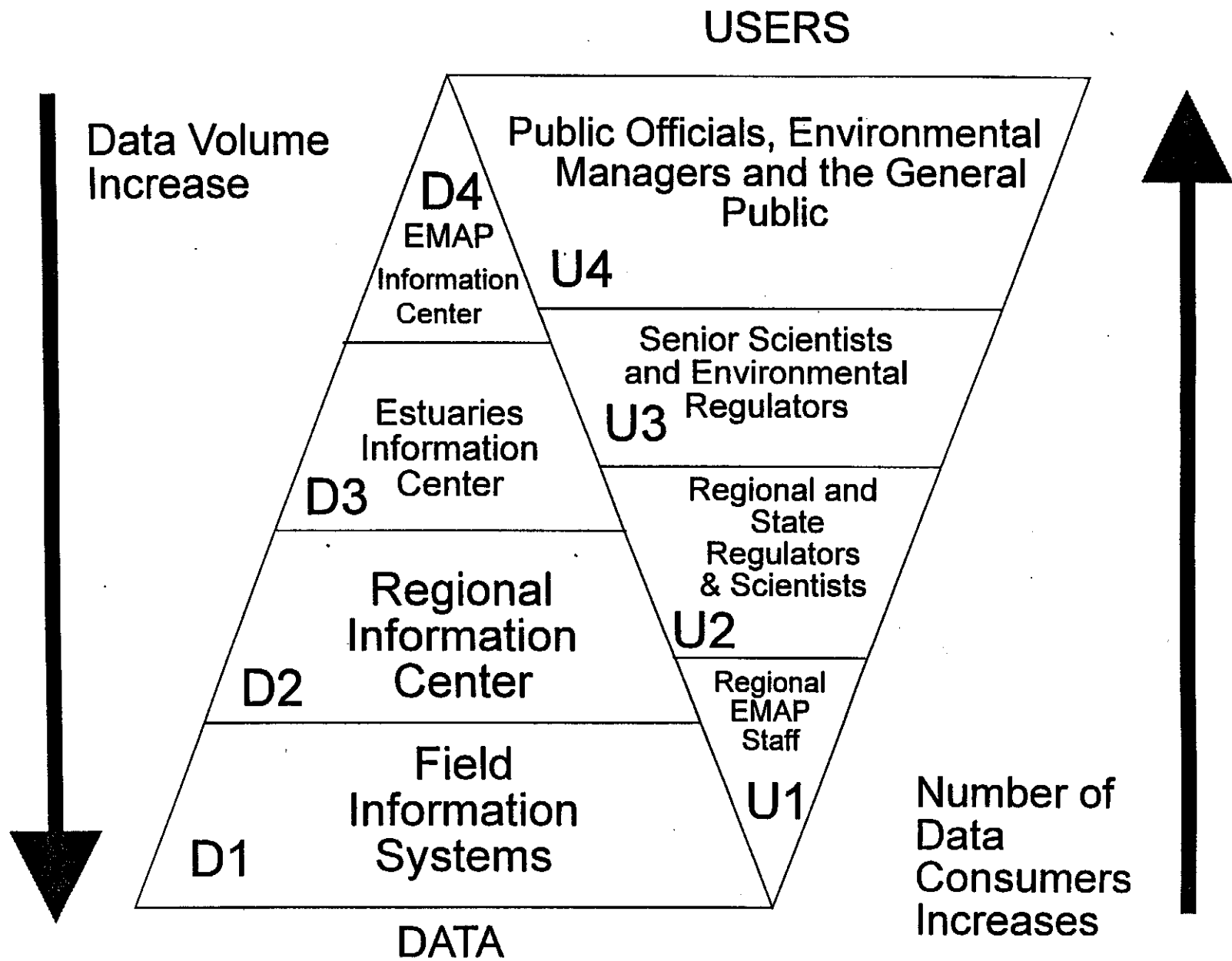
Exposure Indicators

- Sediment contaminants
- Sediment toxicity
- Fish tissue contaminants
- Dissolved oxygen concentration

Habitat Indicators

- Water Temperature
- Salinity
- Water Depth
- Sediment Characteristics
- Water Clarity

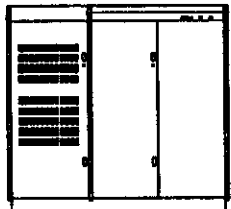
Conceptual Data Levels and User Groups



Data Distribution Prototype

Experiment Under Way : Working with Chesapeake Bay Program, CES, EMAP Reporting and assessment, and Las Vegas to make the data available over the EPA network.

Access all 1990 data aggregates on-line using SAS interface.



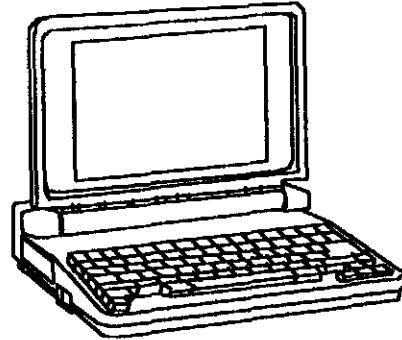
**Data Reside on
Disks in Narragansett**

Request for Data

**Temporary data
file transferred**

**Users run SAS JOB on
Remote VAX . Data retrieved
automatically from
Narragansett**

PC Prototype



Interacts with VAX interface

**Allows user to print on a local printer
connected to PC**

Allows download of data file

**Creates files which can be easily imported
into Lotus 1-2-3, Excel, DBase, etc.**

Also has some graphics built in

Information Management Framework

System Name: EPA EMAP

Users Type	Administrators	Researchers	Conservationists	Public and	ADP Staff
Data Type			Environmentalists	others	Professionals
Historical					
Characterization	R	GRM	R	R	GRM
Monitoring	R	GRMD	R	R	GRMD
Remote Sensing	R	RM	R	R	RM
Research	R	GRMD	R	R	GRMD

Explanation of GRMD - The contents of each cell indicate the interaction between the groups in the columns and the types of data in the rows.

- G = Generate data
- R = Read data
- M = Maintains data
- D = Delete data.

Presenter: JOHN F. PAUL

DRAFT

EPA's Environmental Monitoring and Assessment Program: Availability of Broad-Scale Environmental Data and Opportunities For Use In Environmental Modeling Applications

Richard W. Latimer¹, John F. Paul¹, and Jeffrey S. Rosen²

Abstract

The Environmental Monitoring and Assessment Program (EMAP) has collected a suite of environmental data over a four year period from estuarine systems in the mid-Atlantic and Gulf of Mexico. Data are currently available for secondary users including environmental modelers. The data were collected using a probability-based sampling design which allows unbiased estimates of the status and trends of ecological resources. The parameters collected were chosen as integrative indicators and include both biotic and abiotic condition measures. An information management system has been developed to facilitate the collection, processing, quality assurance, analysis, and distribution of data to a variety of users. The potential usefulness and applicability of EMAP data for modeling studies is presented. EMAP-Estuaries data provide a comparative scale that can be used for placing environmental studies in a regional context. Added benefits include data that are collected and processed in a consistent manner, well documented, can be tracked temporally, and are accessible through a variety of pathways including EPA's Wide Area Network.

Introduction

The Environmental Monitoring and Assessment Program (EMAP) is an innovative nationwide program developed by the Environmental Protection Agency's (EPA) Office of Research and Development (ORD). EMAP was

¹U.S. Environmental Protection Agency, Narragansett, RI 02882

²American Management Systems, Inc., Kingston, RI 02881

DRAFT

designed in response to EPA's Science Advisory Board recommendation to implement a program to monitor ecological status and trends that would identify emerging environmental problems before they reach crisis proportions (SAB, 1988). EMAP is an integrated federal program, ORD is coordinating the planning and implementation of EMAP with other federal agencies including the Agricultural Research Service, the Bureau of Land Management, the U.S. Fish and Wildlife Service, the Forest Service, the U.S. Geological Survey, and the National Oceanic and Atmospheric Administration. EMAP objectives are to: 1) estimate the current status, trends, and changes in selected indicators of the condition of the Nation's ecological resources on a regional basis with known statistical confidence; 2) estimate the geographic coverage and extent of the Nation's ecological resources with known statistical confidence; 3) seek associations between selected indicators of natural and anthropogenic stresses and indicators of condition of ecological resources; and 4) provide annual statistical summaries and periodic assessments of the Nation's ecological resources (EPA, 1993).

The remainder of this paper will provide additional background on EMAP-Estuaries including characteristics of its sampling design and ecological indicators; a description of the important features of Estuaries Information Management System; a discussion of the usefulness of EMAP data for modeling applications; and an explanation of what data are available and how those data may be accessed. Details on EMAP-Estuaries can be found in Holland (1990), Paul *et al.* (1992), Rosen *et al.* (1992) and Weisberg *et al.* (1993).

Background

EMAP is unique in many aspects, while drawing on the successes of other monitoring programs: 1) its scale is national and regional across all ecological resources; 2) it is a long-term program; 3) EMAP uses a probability-based sampling scheme that permits the estimation of resource condition with known confidence; and 4) EMAP emphasizes biological indicators. EMAP-Estuaries represents that part of the overall program dealing with estuarine resources. In 1990 EMAP-Estuaries initiated a demonstration project in the estuaries, bays and sounds between Cape Cod, MA and Cape Henry, VA (Weisberg *et al.*, 1993). Monitoring has continued through 1993. In 1991 monitoring was initiated in the estuaries of the Gulf of Mexico and continues to this date. Pilot studies were initiated in the estuaries along the southeastern coast in 1993 and similar studies are planned in the Gulf of Maine and Southern Florida in 1994. Full implementation will include all coastal areas in the United States.

A unique aspect of EMAP is its reliance on a probability-based sampling design (Overton *et al.*, 1991). There are four essential features of the design for EMAP-Estuaries (Holland, 1990; Paul *et al.*, 1992): regionalization,

DRAFT

classification, index period, and statistical sampling. The regionalization scheme partitions the estuarine and coastal resources of the United States into geographical provinces with similar ecological properties which provide reasonable reporting units. The classification scheme defines certain subpopulations of interest within provinces that are functionally similar and can be sampled using a common approach (e.g., large estuaries, small estuarine systems, and large tidal rivers). The statistical sampling allows for the determination of unbiased estimates of the status and trends of the estuarine ecological resources in a cost-effective manner. Figure 1 illustrates the systematic nature of the sampling design in the Virginian Province. The index period or sampling period for estuaries is in the summer (July-August) when the dissolved oxygen concentrations are at a minimum and biological activity at a maximum.

EMAP monitors ecological indicators to assess status and trends in the condition of the nation's ecological resources. These indicators are measurable environmental attributes that quantify ecological condition or environmental stress (Hunsaker and Carpenter, 1990). Biotic condition indicators measured by EMAP-Estuaries include: benthic species, abundance, community composition, and biomass; fish species, abundance, size, community composition, and incidence of gross external pathologies (e.g., growths, lumps, ulcers, fin erosion, etc.). Abiotic condition indicators historically have been the mainstay of environmental monitoring programs, because these indicators quantify the levels of stress to which organisms are exposed. They also provide many of the parameters needed to develop hydrodynamic and water quality models. Abiotic condition indicators measured by EMAP-Estuaries include: dissolved oxygen concentrations, water temperature, salinity, pH, depth, and water clarity; sediment characteristics; and sediment contaminants and toxicity.

The success of any monitoring program depends upon the degree the data collected are used to answer the questions for which the program was designed (Wolfe, *et al.*, 1987; NRC, 1990). EMAP is designed to assess the condition of the nations ecological resources. However, the Estuaries sampling design also allows questions to be answered for major classes of estuaries (e.g., large estuaries) and certain large estuarine systems like Chesapeake Bay. The large volume of data collected can be used for a wide variety of analysis and interpretations. In order to effectively manage this volume of data, EMAP-Estuaries has established a decentralized Information Management System. All data collection, raw data processing, quality assurance and analysis is conducted on regional systems. The EMAP-Estuaries Information Management Center, located in Narragansett, RI, is responsible for standardizing information systems, cataloging the data available from regional systems, and for the storage and dissemination of aggregated data.

DRAFT

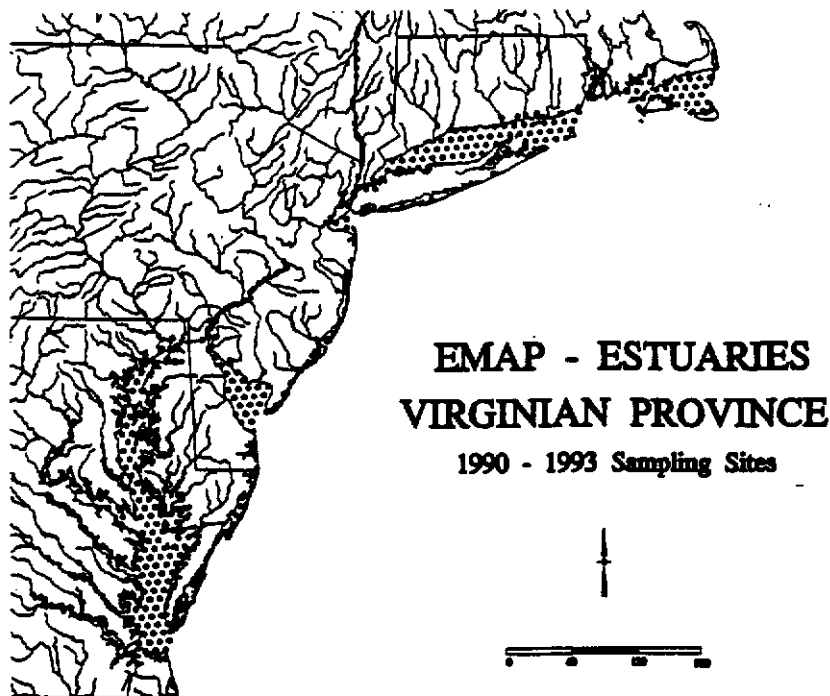


Figure 1. EMAP-Estuaries Probability-Based Sampling Stations for the Virginian Biogeographic Province for the Period 1990-93.

Information Management Systems

The data being collected by EMAP-Estuaries is intended for use by a wide variety of data analysts and environmental managers. The complexity of the use of the data and the variety of users requires that the data be processed and stored at a variety of levels of aggregation. The objective of the EMAP-Estuaries Information Management System is to develop a distributed information system which facilitates the collection, processing, quality assurance, analysis, aggregation and distribution of data to a wide variety of users and at the appropriate technical level. Figure 2 below illustrates the different data levels and the different user groups. The data aggregates (D1 through D4) are discreet data sets designed to meet the needs of primary users.

Raw data (D1) are generated through a sophisticated field computer system which facilitates data collection and transfer to regional information centers. This component is the most highly distributed component of the system. The raw data coming into the regional data center is processed and checked for gross errors. The processed data are then loaded into a data base system for further checking, analysis and aggregation.

The regional information systems (D2) are currently implemented in SAS (SAS, 1990). The system is a completely normalized relation system.

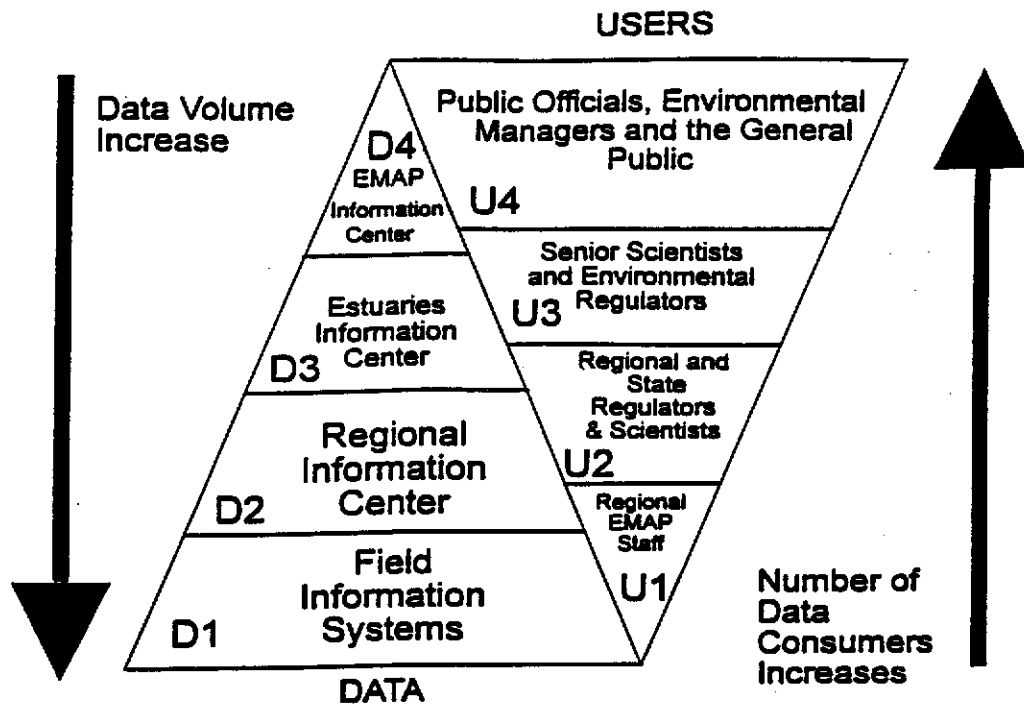


Figure 2. EMAP-Estuaries information management system conceptual diagram for different data levels and user groups.

Specifics of the technical design can be found in Rosen *et al.* (1992). The systems are currently being converted from SAS data bases to ORACLE data base to improve performance and simplify maintenance. The D2 data bases are stored on Digital Equipment Corp. (DEC) VAX computers and interact on a peer to peer basis over a DEC Wide Area Network (WAN) supported by the EPA.

Data aggregates (D3) developed by the individual regional data centers are sent to the EMAP-Estuaries Information Center in Narragansett, RI. These data aggregates include summary results for each parameter measured at a particular station on a specific date. These aggregate data can be cross referenced on-line back to the raw data from which they were developed. This cross referencing is done over the distributed system using straight relation joins. Once the aggregates are joined to the raw data both levels of resolution are available from the system in the same data search.

A central EMAP information system is under development. Results of data analysis and interpretation of analysts will be created and stored as objects on the central information system which will be connected via a WAN to D3 and D2. Summary data tables are planned with summary graphics and tools for using the EMAP data to determine regional distributions and for use in management decisions.

DRAFT

Modeling Applications

The potential usefulness and applicability of EMAP data for modeling studies will be presented by discussing three characteristics of the program and illustrating with actual data.

A primary EMAP objective is to report information on areal extent of conditions for large geographic regions. For EMAP-Estuarines, these regions are the biogeographic provinces. EMAP was not designed to fully describe what is happening on small, local scales; however, information that EMAP collects is actual data from individual sites at specified times, which can be utilized as are other individual datum. One does need to recognize that EMAP data were not designed to provide detailed description at local scales. As a result, multiple samples or replicates at sampling stations are typically not collected as part of EMAP field activities. As an example of EMAP site data, Figure 3 shows the salinity, temperature, and dissolved oxygen vertical profiles collected at EMAP-Estuarines station VA90-026 (41° 0' 7.8" N, 73° 27' 40.8" W) in Long Island Sound on 22 July 1990.

In summary, EMAP site data can be viewed in terms of usefulness as any other individual site data. The added benefits are that EMAP data are collected and processed in a consistent manner, and that quality assurance (QA) for all the data is well documented and stored as part of the information management system.

EMAP uses a probability-based sampling design to select sites for sampling. Thus, sites are selected based upon their probability of occurrence in a specified statistical population. The population that EMAP-Estuarines describes is the total estuarine area in each of the biogeographic provinces. The provinces are further divided into classes based upon physical characteristics of the estuarine systems. Samples in large estuaries are distributed according to the areal extent of each of the large systems. For example, Chesapeake Bay represents approximately 49% of the estuarine area in the Virginian Province and, therefore, approximately 49% of the sample sites in the province occur in Chesapeake Bay.

The probability-based sampling design allows the individual sample values to be statistically combined in a rigorous fashion so estimates of the distribution of indicator values can be reported, along with estimates of the uncertainty about the distribution. For example, Figure 4 shows a cumulative distribution function (CDF), and 95% confidence estimates, for bottom (1m above the sediments) dissolved oxygen (DO) in Chesapeake Bay for the late summer period (mid-July to end of August) in 1990. The mean is 5.9 mg/l, that is, 50% of the estuarine area of Chesapeake Bay had DO at or below this value. From this CDF, estimates can be made, for example, on the area of Chesapeake Bay below 5 mg/l ($29 \pm 15 \%$) and 2 mg/l ($18 \pm 14 \%$), two

DRAFT

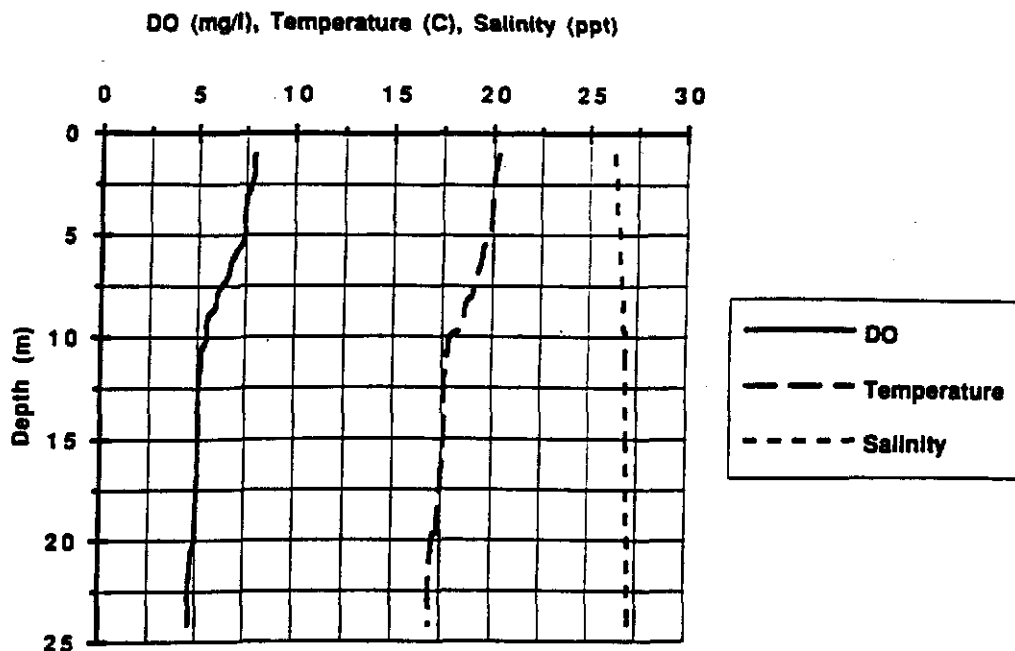


Figure 3. Example of EMAP-Estuaries station data from Long Island Sound ($41^{\circ} 0' 7.8''$ N, $73^{\circ} 27' 40.8''$ W) for dissolved oxygen, temperature, and salinity collected on 22 July 1990.

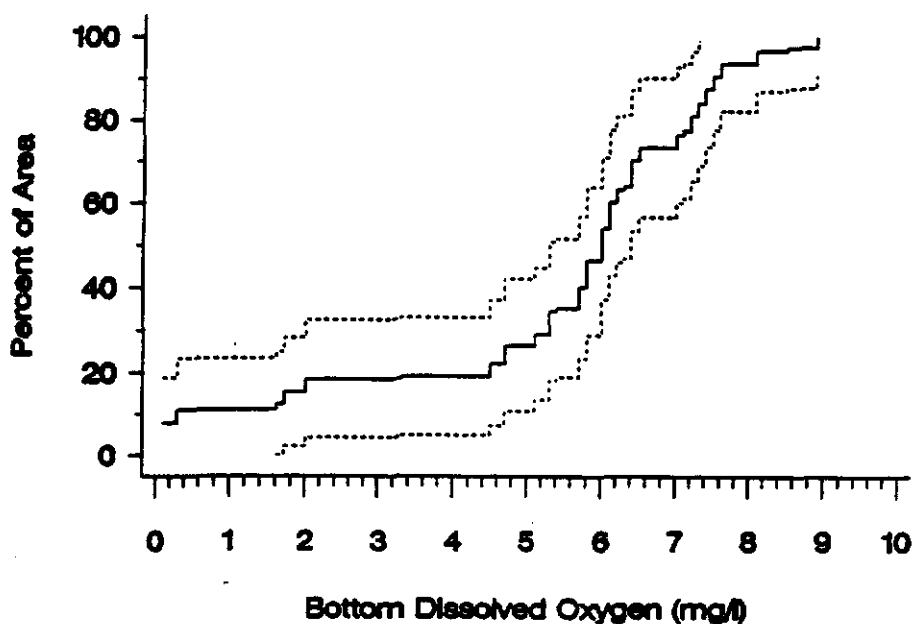


Figure 4. Cumulative distribution function for EMAP-Estuaries bottom dissolved oxygen data from Chesapeake Bay during late summer of 1990. Solid line is mean estimate, and dotted lines are 95% confidence intervals about the mean estimate.

DRAFT

commonly used values for estimating conditions of estuarine waters. The CDF in Figure 4 is based upon one year of data. The uncertainty estimates will decrease by approximately a factor of two as four years of data are used to construct the CDF.

The usefulness of this approach can be seen as follows. Environmental managers of the Chesapeake Bay watershed are instituting nutrient controls to alleviate the low DO conditions in the bay. The critical DO values of concern for the organisms residing in the bay waters are those typically below 2 mg/l. There is interest in decreasing that area of the Bay which has, say, $DO \leq 2$ mg/l. EMAP data are designed to provide information to determine how extensive these low DO conditions are, and how these conditions change through time. For example, simulations from the Chesapeake Bay water quality model (Cerco and Cole, 1992) might indicate that A% of the bottom waters have $DO \leq 2$ mg/l during the late summer period under existing nutrient loading conditions. If proposed loading reductions go into effect, then model simulations might indicate that these low DO waters are reduced to B% of the bay. Power analyses could be conducted to see if such reductions would be detectable (Weisberg *et al.*, 1993; EPA, 1992). The data provide for validation of the model predictions. Note that EMAP data would not provide detailed data on the annual dynamics in the bay, but rather would provide information during a critical period of the year, which can be tracked through time.

In summary, EMAP data provide estimates with known confidence for overall condition during a critical period (index period), that can be tracked through time. These data can provide long-term data sets suitable for model validation.

Because EMAP data are collected across geographic provinces, the full range of values that are expected to occur in the estuarine waters of the provinces are represented in the data. Therefore, the distributional information available with EMAP data provide a large-scale environmental "yardstick". This can be used to make comparisons with smaller-scale studies to see how they fit within the full range of data for the large geographic scale. Thus, EMAP data can be used to select "reference" conditions for environmental studies, or can be used in a subjective, judgmental fashion to indicate whether a particular study area is representative of the "best", "typical", or "worst" conditions that are expected to occur. Figure 5 shows the CDF of Pb concentrations in the surficial bottom sediments for the entire Virginian Province from the late summer period in 1990. Mean Pb values are approximately 19 ppm, while less than 10% of the bottom area had concentration below 7 ppm and 90% of the bottom area had concentration below 50 ppm.

In summary, EMAP data provide a comparative scale for conditions that are likely to be expected over large geographic scales, which place

DRAFT

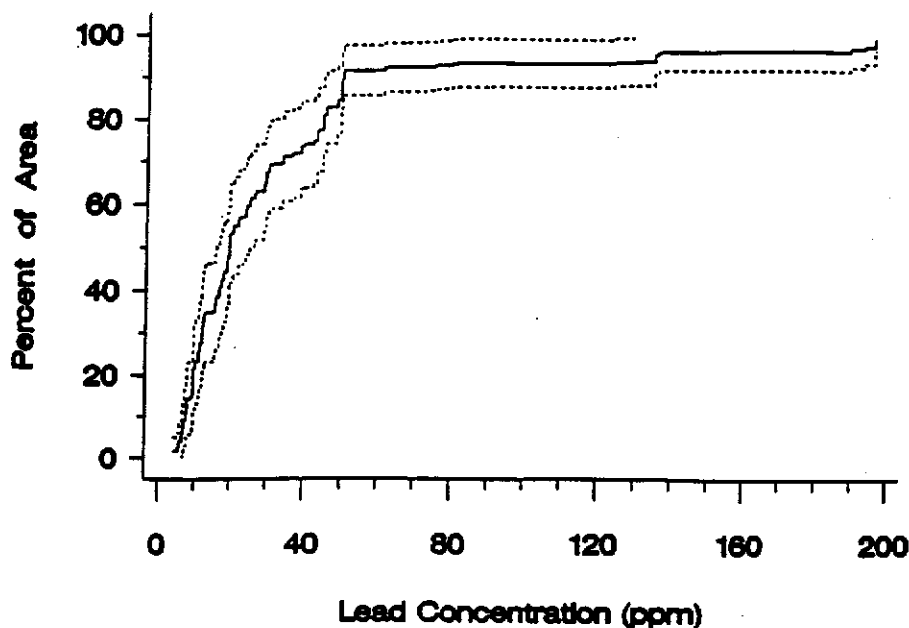


Figure 5. Cumulative distribution function for EMAP-Estuaries total lead concentration in surficial sediments for Virginian Province during late summer of 1990. Solid line is mean estimate, and dotted lines are 95% confidence intervals about the mean estimate.

environmental studies in a regional context.

Data Availability and Access

The type and quantity of EMAP-Estuaries data available for possible modeling applications are summarized in Table 1. Data are accessible through a variety of pathways, which are summarized in Figure 6. The EPA DEC WAN and can be used to access appropriate EMAP-Estuaries data interactively from the node on which the user has an account. An interface has been developed and tested which allows users to access data without knowing the location of the data.

Users who do not have access to the EPA network may have access to the data through a dial-in interface which is being developed and tested. The prototype interface will allow any user with a PC compatible computer and a modem to dial into the EMAP-Estuaries Information Center and access appropriate data. Users will not be given accounts on the EMAP-Estuaries system but will be provided mechanisms to access, download, and analyze the data.

Data requests can be sent electronically or on paper forms to the EMAP-Estuaries Data Librarian. Upon approval from the Director of Operations the data can be loaded onto electronic media or sent via electronic mail to the

DRAFT

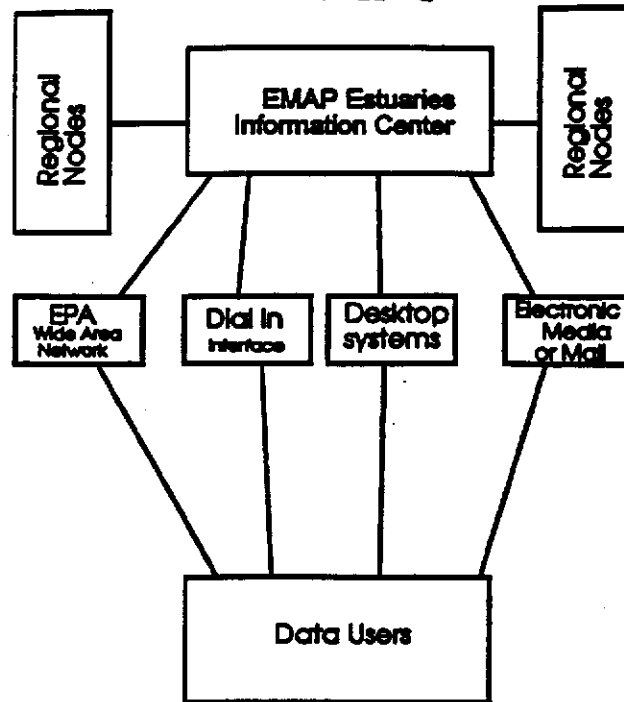


Figure 6. Diagrammatic representation of accessibility paths to the EMAP-Estuaries data.

Table 1. Summary of data availability through the EMAP-Estuaries Information Management Center¹.

Province Year	Water Quality Vertical Profiles	Water Quality Temporal Profiles	Sediment Grain Size	Sediment Total Organic Carbon	Sediment Contaminants
VA/90	106	21	101	104	104
VA/91	108	105	106	109	110
VA/92	116	0	108	108	108
VA/93	128	0	128	128	128
LA/91	100	85	100	98	100
LA/92	110	93	110	110	110
LA/93	122	109	122	122	122

¹Table entries represent the number of probability-based stations for which data are available. VA/90, VA/91, and LA/91 have completed QA. VA/92 and LA/92 have not completed final QA. VA/93 and LA/93 are projected. Other station data are available from EMAP special studies.

DRAFT

user. The EMAP-Estuaries Information Center has access to Internet electronic mail, therefore requests and delivery of the data can occur over Internet.

EMAP-Estuaries has collaborated with NOAA's Strategic Environmental Assessments Division to develop a stand alone desktop system which presents the EMAP-Estuaries data at the D3 level (NOAA, 1991). The system also incorporates data collected from other programs in with the EMAP data. The desktop system runs on a Macintosh computer and is designed to facilitate integrative analyses, and interface to mapping and other analytical tools.

All users are requested to get authorization from the EMAP-Estuaries Director of Operations. He can be reached at 401-782-3077. The Director of Operations grants appropriate access and facilitate the exchange of software and data. For technical information on the content of the data and hardware requirements contact the EMAP-Estuaries Data Librarian at 401-782-3184.

Summary

EMAP-Estuaries is a developing nationwide research, monitoring, and assessment program that relies on ecological indicators and a probability-based sampling design to estimate the current status, trends, and changes in the condition of estuarine resources. Large quantities of environmental data are managed through an information management system and are available to different user groups through a variety of pathways. EMAP-Estuaries data are collected and processed in a consistent manner, providing added-value for modeling applications. The data provide estimates with known confidence over long-time periods which make them suitable for model validation. And finally, EMAP-Estuaries data provide a comparative scale for conditions over large geographic scales, which can be used for placing environmental studies in a regional context.

Acknowledgements

The information in this paper has been funded by the U.S. Environmental Protection Agency, it does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. JSR was supported through EPA Contract No. 68-C1-0005. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. We want to acknowledge the assistance of the EMAP Virginian and Louisianian Province personnel in providing access to the data used in this paper. This is Contribution No. **** of the Environmental Research Laboratory-Narragansett.

DRAFT

Appendix - References

Cerco, C.F. and T.M. Cole. 1992. Thirty year simulation of Chesapeake Bay eutrophication. In: Estuarine and Coastal Modeling, Proceedings of the 2nd International Conference. M.L Spaulding, K. Bedford, A. Blumberg, R. Cheng, and C. Swanson (eds.), ASCE, New York, pp. 116-126.

Environmental Protection Agency. 1992. Monitoring Guidance for the National Estuary Program. EPA 842-B-92-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Environmental Protection Agency. 1993. Program Guide: Environmental Monitoring and Assessment Program. EPA/620/X-93/XXX. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Holland, A.F. (ed.) Near Coastal Program Plan for 1990: Estuaries. 1990. EPA/600/4-90/033. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI.

Hunsaker, C.T. and D.E. Carpenter. 1990. Environmental Monitoring and Assessment Program Ecological Indicators. EPA/600/3-90/060. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

National Oceanic and Atmospheric Administration. 1991. A Desktop Information and Mapping System for Near Coastal Areas of the Mid-Atlantic United States. Project Description. Strategic Environmental Assessments Division, Office of Resources Conservation and Assessment, NOAA, Rockville, MD.

National Research Council. 1990. Managing Troubled Waters: The Role of Marine Environmental Monitoring. National Academy Press, Washington, DC, 125 pp.

Overton, W.S., D. White, and D.L. Stevens. 1991. Design Report of the Environmental Monitoring and Assessment Program. EPA/600/3-91/053. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Corvallis, OR.

Paul, J.F., K.J. Scott, A.F. Holland, S.B. Weisberg, J.K. Summers, and A. Robertson. 1992. The Estuarine Component of the US E.P.A.'s Environmental Monitoring and Assessment Program (EMAP). *Chemistry and Ecology*, 7:93-116.

Rosen, J.S., H. Buffum, M. Adams, J. Copeland, J. Beaulieu, M. Hughes, D.

DRAFT

Battista, E. Petrocelli, M. Aitkenhead, C. Smith, R. Latimer, K. Summers, N. Rubinstein, and B. Melzian. 1992. Information Management Plan for the Estuaries Resources Group. Draft Report. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Protection Agency, Narragansett, RI.

SAS Institute Inc. 1990. SAS User's Guide: Basics, Version 6 Edition. SAS Institute Inc., Cary, NC.

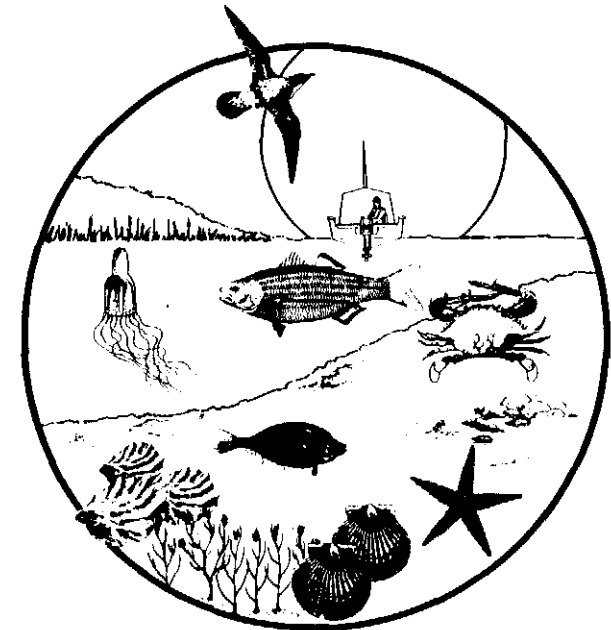
Science Advisory Board. 1988. Future Risk: Research Strategies of the 1990s. SAB-EC-88-040. U.S. Environmental Protection Agency, Science Advisory Board, Washington, DC.

Weisberg, S.B., J.B. Frithsen, A.F. Holland, J.F. Paul, K.J. Scott, J.K. Summers, H.T. Wilson, R.M. Valente, D.G. Heimbuch, J. Gerritsen, S.C. Schimmel, and R.W. Latimer. 1993. EMAP-Estuaries Virginian Province 1990 Demonstration Project Report. EPA/620/R-93/007. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

Wolfe, D.A., M.A. Champ, D.A. Flemer, and A.J. Mearns. 1987. Long-term biological data sets: Their role in research, monitoring, and management of estuarine and coastal marine systems. *Estuaries*, 10:181-193.

June 1992

Environmental Monitoring and Assessment Program



Near Coastal Component

condition of coastal waters. In 1990, in excess of 25,000 fish were collected, which were statistically representative of the entire Virginian Province. EMAP scientists found that less than 1% of these fish had visual external symptoms of disease (like tumors and lesions). This incidence of visual external symptoms was confirmed with the 1991 collection activities in the Virginian Province, and was also observed with the fish collected in the Louisianian Province in 1991.

To assist government managers, scientists, citizens, and others who will benefit from EMAP-NC, a user network is being developed. The network will provide for workshops, transfer of monitoring data, and the exchange of informational materials, and will help ensure that EMAP meets user needs.

Everyone will benefit from the program when it is in full operation. Scientists will have access to environmental data on a scale never before available. Managers and legislators will have the means to determine the effectiveness of pollution control measures. They will also have the information to determine which problems pose the greatest threat to the nation's coastal waters, allowing them to set priorities for pollution control activities. Every American, from fisherman to business person, will benefit because understanding the real health of our environment is critical to developing a sound, long-term plan to protect it. Only then can we assure an environment that future generations can enjoy.

Additional information can be obtained by writing to the following:

EMAP-Near Coastal
United States Environmental Protection Agency
27 Tarzwell Drive
Narragansett, RI 02882

National Status and Trends Program
National Oceanic and Atmospheric Administration
6001 Executive Blvd.
Rockville, MD 20852

Looking collectively at all indicators will provide information on the extent of problems and possible causes of those problems. For example, the consistent absence of any long-lived, bottom-dwelling animals would indicate a problem. Such a problem could be due to chemical pollutants in the sediments, low dissolved oxygen in the water, or other factors, such as disease or physical alteration. Information collected by EMAP-NC will allow scientists to explore these associations.

The long-term nature of EMAP-NC will allow environmental managers to determine if conditions are improving or getting worse and to evaluate if pollution control practices designed to protect our coastal waters are effective.

Eventually, the information gathered by EMAP for all ecosystems will be used to report on the well-being of the nation's environment. This information will also be used for more comprehensive studies of how the ecosystems interact and, therefore, how activities in one ecosystem may affect others.

During the summer of 1990, sampling began in estuaries throughout the Virginian Province. Results are available from the address given on the back of this brochure. Reviews by scientists and coastal managers have addressed assumptions, methodology, and findings from the 1990 and 1991 monitoring efforts. Comments and recommendations from these reviews are helping to determine future sampling methodology, indicators, and monitoring sites.

In 1991 sampling continued in the Virginian Province and began in estuaries of the Louisianian Biogeographic Province, which comprises the Gulf of Mexico from north of Tampa Bay to the Mexican border. Reports on these activities are available upon request.

In 1992 sampling will continue in the Virginian and Louisianian Provinces and a pilot project testing indicators will be conducted in Lake Michigan.

As an example of the types of information that EMAP-NC will be able to report, let's look at what the EMAP scientists have observed in fish disease. The occurrence of fish disease is often recognized as one measure of the

A USER NETWORK



GETTING MORE INFORMATION



PROGRESS REPORT



Environmental Protection Agency

National Oceanic and Atmospheric Administration





Each year the United States spends billions of dollars to reduce pollution to surface waters. Federal, state, and local government agencies monitor potential pollutants to ensure compliance with environmental regulations. Detailed studies are also conducted in systems such as Chesapeake Bay, Long Island Sound, Puget Sound, San Francisco Bay, and Galveston Bay. Yet, as a nation we have little comprehensive information about our success in protecting the overall health and integrity of the hundreds of bays, estuaries, and other bodies along our coast. We refer to water bodies along our coastal areas, including the Great Lakes, as near coastal waters.

An ecological resource is the combined population of plants and animals and the interactions among them and their non-living surroundings. The near coastal ecological resource includes the waters and bottom sediments of bays, estuaries, and coastal areas and their resident plants and animals, that range from microscopic plankton to large fish and marine mammals. Estuaries—the areas between the fresh water of inland rivers and salt water of the open ocean—have unique characteristics, which make them our most productive coastal ecological resources.

When near coastal ecological resources are healthy, they contain well-balanced populations of plants and animals and enjoy physical and chemical conditions that support their growth and reproduction. This ecological balance enables water bodies and their populations of living organisms to resist or recover from changes brought on by adverse weather conditions, disease, or man-induced stress.

The U.S. Environmental Protection Agency (EPA), in cooperation with other federal agencies that have research responsibilities, is implementing a long-term program that will provide the public, scientists, and Congress with information that can be used to evaluate the overall health of the nation's ecological resources. The EPA program is called the Environmental Monitoring and Assessment Program or EMAP.

EMAP is a nationwide, interagency, and environmental research, monitoring, and

assessment program, focusing on seven major ecological resource categories: estuaries; the Great Lakes; forests; inland waters, such as lakes and streams; agricultural lands; wetlands; and arid ecosystems, such as deserts and grasslands.

EMAP is conducting annual surveys to measure indicators of the health of plants and animals, the quality of their surroundings, and the presence of pollutants. The program is a long-term monitoring activity, operating into the next century, that will provide information on the overall health of our environment and the effectiveness of pollution prevention and measures, such as sewage treatment and industrial discharge control.

For the near coastal component of EMAP (EMAP-NC), ecosystem health is being addressed by investigating the regional distribution of fish and bottom-dwelling organisms. EMAP-NC is also determining what portions of estuaries and Great Lakes can support these plants and animals and finding out why certain areas do not support them.

EMAP is designed to look at ecosystem health on a regional scale. This means that information collected is addressing large regions such as the mid-Atlantic coast or Gulf of Mexico, rather than smaller systems like the New York Harbor or Galveston Bay. An intense study of every bay and estuary would be too costly.

A REGIONAL APPROACH



The EMAP-NC approach divides all the nation's coastal waters, bays, and estuaries into defined areas for study. The first area covers the mid-Atlantic coastal states, from Cape Cod, Massachusetts, to Cape Henry, Virginia—the Virginian Biogeographic Province. Studies began in the summer of 1990. Sampling of other regions will be phased in over the next several years, until all of the coastal areas of the country are sampled yearly.

Within each region, measurements will be made every year at randomly selected stations. Teams of scientists will be on the water during the summer, collecting important information that will help to characterize the ecological status of our coastal waters. Sampling will be done during summer months because that is when plants and animals are generally most

A TEAM EFFORT



COLLECTING AND USING DATA



active and when the effects of pollution are most severe.

Although EMAP is being funded through EPA, many activities will be conducted jointly with other federal agencies and the states. The National Oceanic and Atmospheric Administration (NOAA), for example, is working closely with EPA on EMAP-NC. Since 1984, NOAA's National Status and Trends Program has been collecting information on the presence of chemical contaminants in mussels, oysters, fish, and sediments at almost 300 sites along the coast of the United States. It has been focusing primarily on chemical pollutants and their biological effects. EMAP-NC, on the other hand, is emphasizing indicators of ecosystem health. Through its Strategic Environmental Assessment Program, NOAA has been systematically characterizing the nation's estuarine and coastal areas, providing information essential for interpreting and understanding monitoring results. NOAA and EPA are jointly conducting these monitoring and assessment programs in coastal waters and sharing the information collected. Together, these programs are providing us with a better picture of the health of near coastal waters than either program could provide alone.

Each summer in EMAP regions throughout the coastal United States, EPA scientists will collect bottom sediments to determine the number and type of organisms present. Sediments will also be tested to see if they contain harmful chemical pollutants.

Fish collected through trawling will be counted, identified to determine the species present, and examined for disease. Some fish will be preserved for later chemical contaminant analysis. Sediment and fish tissue collected in the field will be shipped to specially equipped laboratories for analysis.

Specialized gear will be used to measure parameters like salinity, temperature, and dissolved oxygen, and other important indicators of water quality.

Information collected by the program will be used to determine the health of our coastal waters on regional and national scales.

DATA MANAGEMENT USING THE INTERNATIONAL DIRECTORY NETWORK

Presented to: The Gulf of Maine Data and Information Systems
Workshop
University of New Hampshire November 4, 1993

Gerald S. Barton
Director, Global Change Research Information Office
1825 K St. NW Suite 805
Washington DC 20006
202-775-6628 fax -6622
gbaron@mail.ciesin.org
g.barton on OMNET

U.S. GLOBAL CHANGE RESEARCH PROGRAM

A set of data and information management tools are in use by United States and international organizations. Activities initiated by the Interagency Working Group on Data & Management for Global Change (IWGDMGC) resulted in a world-wide system for locating and accessing data and information. An International Directory Network was implemented through the Committee on Earth Observation Satellites (CEOS).

The IWGDMGC established the Global Change Master Directory (GCMD) in 1990 to be the main service provided by U.S. government agencies to describe and locate data and information useful for the study of Climate and Global Change. The GCMD currently describes over 2200 data sets. A standard format, the Directory Interchange Format (DIF) is used to describe and exchange information about data sets. The CONTACT field of the DIF refers the user to the holder of the data. If the data system described is available on-line, the user can be connected immediately to the system using the LINK option.

INTERNATIONAL DIRECTORY NETWORK

CEOS is concerned with the coordination of satellite operations and data from satellites. Using the GCMD software and data base, CEOS sponsored the implementation of the International Directory Network (IDN) with major coordinating directory nodes in the United States, Italy, and Japan. The directories are connected by communication networks and can LINK to each other and to cooperating directory systems and on-line data and information systems.

An example of the world-wide interconnection capability is a user in the U.S. connecting to the Spot/Image satellite inventory system in Toulouse, France. A U.S. user can access the NOAA Environmental Services Data Directory, search for the description of the IDN directory at the European Space Agency (ESAPID) in Frascati, Italy, LINK to ESAPID and search it for the SPOT/IMAGE description, and LINK to it. The user from U.S. may then do an inventory search for satellite data. This all happens over an international system of computer communication networks.

GLOBAL CHANGE RESEARCH INFORMATION OFFICE

In 1990, Congress passed and the President signed Public Law 101-606, the Global Change Research Act of 1990. The purpose of the legislation was "to require the establishment of a United States Global Change Research Program aimed at understanding and responding to global change, including the cumulative effects of human activities and natural processes on the environment, to promote discussions towards international protocols in global change research, and for other purposes."

Under Title II (International Cooperation in Global Change Research) of the Act, Section 204 requires that a Global Change Research Information Office (GCRIO) be established. The stated purpose of the GCRIO is "to disseminate to foreign governments, businesses, and institutions, as well as citizens of foreign countries, scientific research information available in the United States which would be useful in preventing, mitigating, or adapting to the effects of global change."

In May 1992, the the Subcommittee on Global Change Research (SGCR) of the Committee on Earth and Environmental Sciences (CEES) designated that the GCRIO be implemented within the Consortium for International Earth Science Information Network (CIESIN).

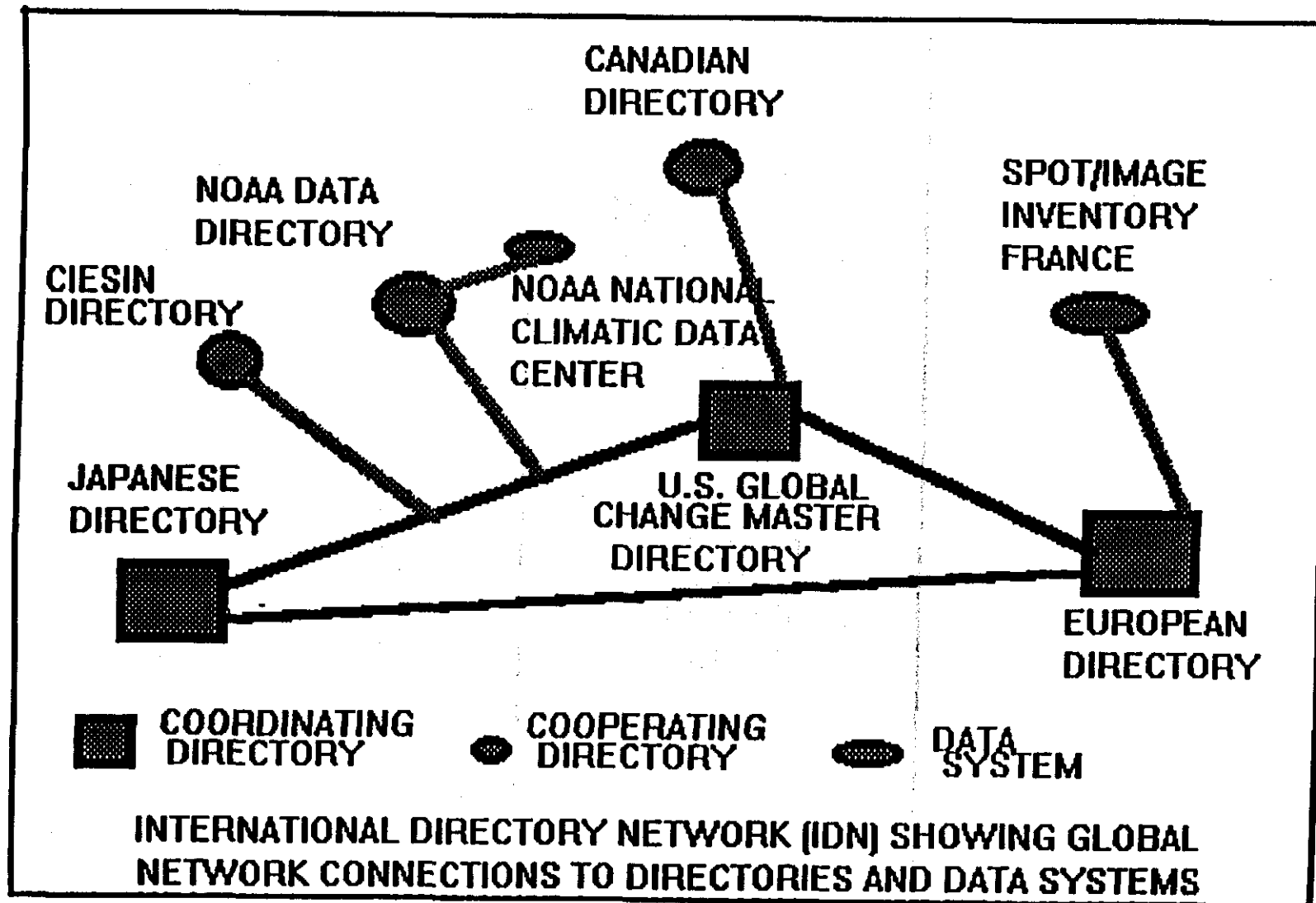
The services of the GCRIO are being established to allow users to access a variety of services:

- Telephone calls to GCRIO Customer Services staff
in Washington DC and Saginaw Michigan
- Computer access to on-line GCRIO information services
telephone dial-in access in Washington DC
and Saginaw Michigan
- Internet access
- Electronic Computer Mail
- Postal mail access
- User services staff.

The GCRIO services use the Global Change Data and Information System (GCDIS) under development for the U. S. Global Change Research Program, and other on-line data and information services including:

- | | |
|-----------------------|---------------------|
| bulletin boards | CD-ROM |
| data browsing | data directories |
| data inventories | information systems |
| libraries | on-line data |
| on line data ordering | on-line documents. |

The GCRIO is implementing a GOPHER menu system on the Internet which will identify these resources, and allow the user to connect to some of these on-line systems.



DATA MANAGEMENT USING THE INTERNATIONAL DIRECTORY NETWORK

PRESENTED TO THE GULF OF MAINE DATA AND INFORMATION SYSTEMS WORKSHOP

UNIVERSITY OF NEW HAMPSHIRE NOVEMBER 4, 1993

GERALD BARTON, DIRECTOR
GLOBAL CHANGE RESEARCH INFORMATION OFFICE



U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX -6622

GCRIO is operated by CIESIN for the U.S. Global Change Research Program

DATA MANAGEMENT USING ON-LINE DATA & INFORMATION TOOLS INTERNATIONAL DIRECTORY NETWORK (IDN)

COORDINATING NODES
GLOBAL CHANGE MASTER DIRECTORY
EUROPEAN DIRECTORY
JAPANESE DIRECTORY

COOPERATING NODES EXAMPLES:
NOAA ENVIRONMENTAL SERVICES DATA DIRECTORY
CIESIN CATALOG SERVICES
CANADIAN DIRECTORY

U.S. GLOBAL CHANGE RESEARCH INFORMATION OFFICE
USE OF ON-LINE DATA AND INFORMATION SERVICES



U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX -6622

GCRIO is operated by CIESIN for the U.S. Global Change Research Program

U.S. GLOBAL CHANGE RESEARCH PROGRAM (USGCRP)

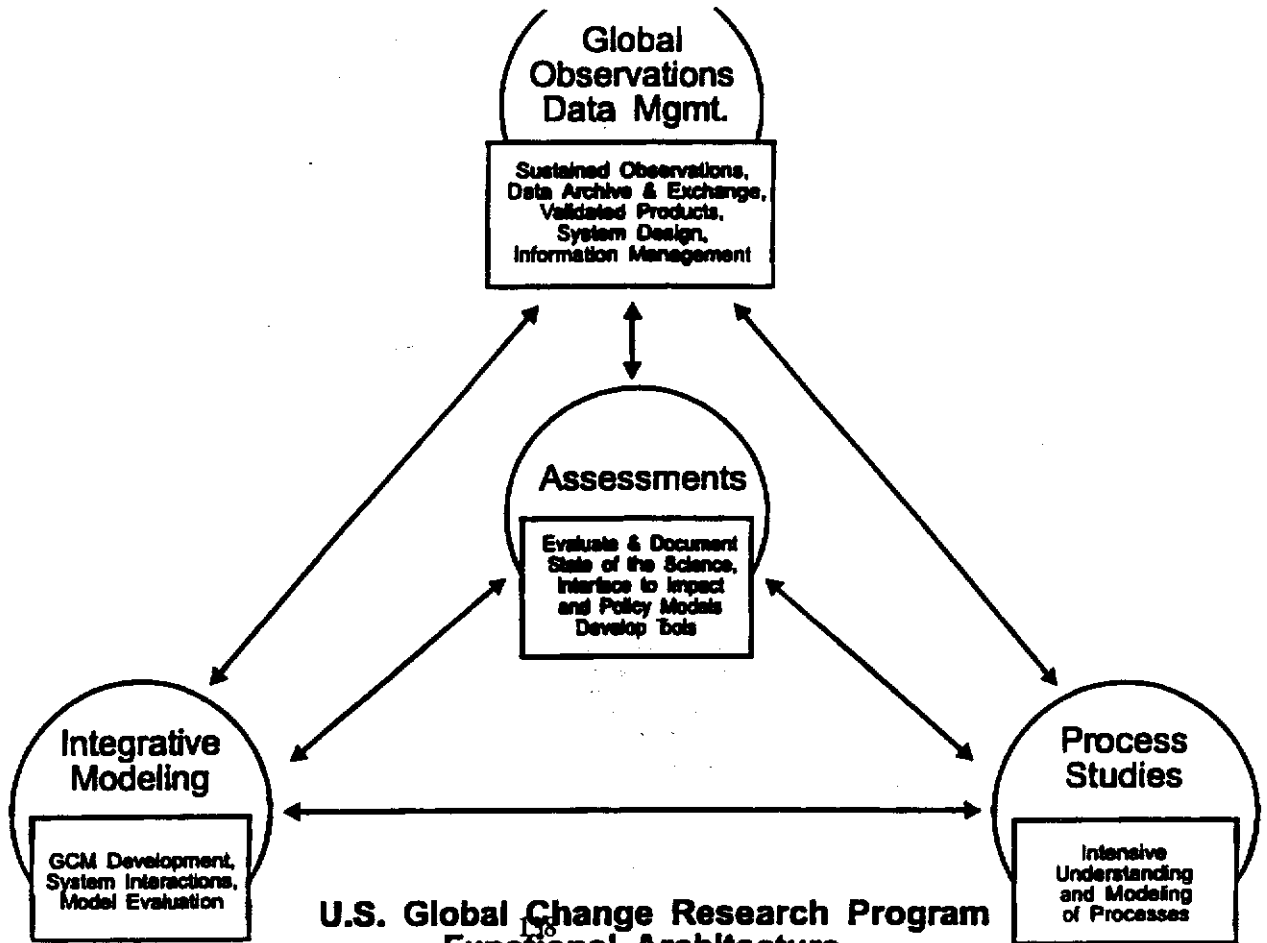
FEDERAL COORDINATING COMMITTEE ON SCIENCE
ENGINEERING AND TECHNOLOGY (FCCSET)

COMMITTEE ON EARTH AND ENVIRONMENTAL
SCIENCES (CEES)

SUBCOMMITTEE ON GLOBAL CHANGE RESEARCH (SCGR)

WORKING GROUP 1 OBSERVATIONS / DATA
AND INFORMATION
MANAGEMENT

3



WORKING GROUP 1

DATA AND INFORMATION MANAGEMENT BY THE
INTERAGENCY WORKING GROUP ON DATA
MANAGEMENT FOR GLOBAL CHANGE (IWGDMGC)

GLOBAL CHANGE MASTER DIRECTORY 1990

DIRECTORY INTERCHANGE FORMAT (DIF)
METADATA STANDARD FOR DATA DESCRIPTIONS

DIRECTS USER TO HOLDER OF DATA

CONNECTIONS TO ON-LINE DATA SYSTEMS

DIRECTORIES, INVENTORIES, DATA,
INFORMATION, BROWSE, ORDERING

5

9

Template for DIFs

```
Entry_ID:
Entry_Title:
Parameter:
Group: Data_Center
Data_Center_Name:
Dataset_ID:
Group: Data_Center_Contact
Last_name:
First_name:
Middle_name:
Email:
Phone:
Group: Address
End_Group
End_Group
End_Group
Originating_Center:
Group: Summary
End_Group
Start_date:
Stop_date:
Sensor_name:
Source_name:
Group: Investigator and/or Technical_contact and/or Author
Last_name:
First_name:
Middle_name:
Email:
Phone:
Group: Address
End_Group
End_Group
End_Group
Campaign: or Project:
Storage_medium:
Discipline:
Group: Coverage
Southernmost_Latitude:
Northernmost_Latitude:
Westernmost_Longitude:
Easternmost_Longitude:
End_Group
Location:
Keyword:
Revision_Date:
Science_Review_Date:
Future_Review_Date:
Group: Reference
End_Group
Group: Quality
End_Group
```

COMMITTEE ON EARTH OBSERVATIONS SATELLITES (CEOS)

WORKING GROUP ON DATA

CATALOG SUBGROUP

INTERNATIONAL DIRECTORY NETWORK (IDN)

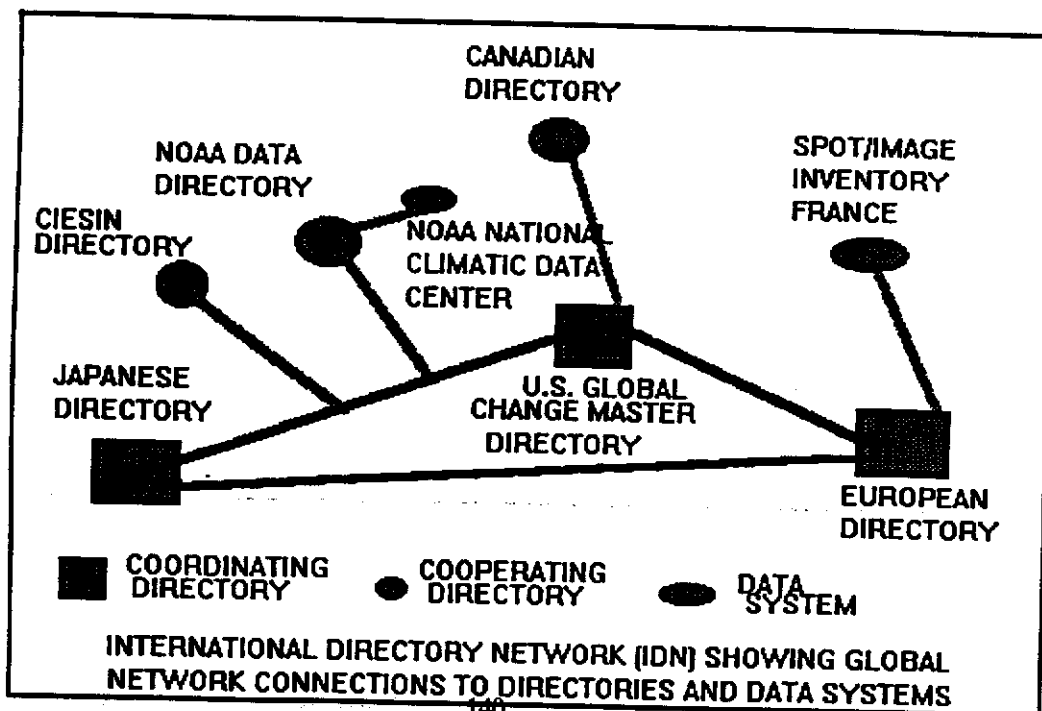
COORDINATING NODES

U.S., EUROPE, JAPAN

COOPERATING NODES

CIESIN, CANADA, NOAA, SPOT/IMAGE

7



NOAA DIRECTORY

USES DIF FOR DESCRIBING DATA

USES THE SAME VAX SOFTWARE AS THE GCMD
(TRANSFER TO WORKSTATION LATE 1992)

AVAILABLE VIA

NASA DECNET (SPAN)

INTERNET

DIAL-IN TELEPHONE INCLUDING 800 NUMBER

SERVES NOAA MANAGEMENT, NOAA AND OTHER USERS

CONTENTS AS OF SEPTEMBER 30, 1992

OVER 1200 DESCRIPTIONS OF NOAA DATA

49 ENTRIES IN THE GUIDE SECTION
15 CENTERS, 12 PROJECTS,
7 SOURCES, 15 SENSORS

9

Welcome to NOAA's Environmental Services Data Directory

1. Original Directory Interface (MD1)
(No special terminal emulation required)

SORRY. THIS OPTION NOT WORKING TODAY. PLEASE CHECK BACK LATER.

- [2. New Directory Interface (MD2) -- TEST VERSION]
[(Requires vt100 or higher terminal emulation)]

3. NEW !!! Manuals for using the NOAA Directory
4. Connections to Other Environmental Data Systems
5. Future Plans for the NOAA Directory
6. Leave us your Comments, Questions, Suggestions

Problems to Anne O'Donnell at (202) 606-5012, odonnell@esdim1.nodc.noaa.gov

Choice (0 = exit):

Connect to Other Environmental Data Systems

NOTE: Brackets [...] show future connections

NOAA Systems:

1. NCDC - National Climatic Data Center
- [2. NGDC - National Geophysical Data Center]
- [3. NODC - National Oceanographic Data Center]
- [4. NOAA Library]

Other Earth Science/Global Change Systems:

5. Canada's GCNet
6. CIESIN's gopher
7. CIESIN's Green Pages
8. ESA - European Space Agency
- [9. FedWorld - Federal Government Information]
10. GCMD - Global Change Master Directory
11. GLIS - Global Land Information System
12. KuDA - Kuwait Data Archive
13. NASDA - Japanese Space Agency

Choice (0 = exit):

17

SELECTED TITLES PERTAINING TO LAKES

Ice Analyses and Forecasts for the Arctic, Antarctic, and Great Lakes from the NAVY/NOAA Joint Ice Center

Great Lakes Water Level Program (NOAA), including hydrography of Monthly Mean Levels, Bench Mark Descriptions, and Great Lakes Datum Elevations

Bay and Stream Chemistry of Lake Michigan (1970-72)

* Climate Data for the USSR (1700-1891)

Hourly Observations and Daily Means of Weather Data for Ukraine Observatory in Russia (1924-1925)

Hourly Observations and Climate Data for Cities in the Russian Empire (1813-1816)

Snow, Frost, and Ice Data for Finnish Winters (1892-1899)

Hourly Observations and Daily and Monthly Climatic Data for Cities in Sweden (1858-1929)

Monthly, Seasonal, and Annual Surface Weather Data for Mexico City and Stations Throughout Mexico (1877-1927)

13

13

EXAMPLE OF A FOREIGN METEOROLOGICAL DATA SET
FROM THE NOAA LIBRARY COLLECTION

Entry_ID: NCL00038
Entry_Title: Climate Data for the USSR.(1700-1891)
Originating_Center: NOAA/NESDIS/NODC/LISD
Start_Date: 1700-01-01
Stop_Date: 1891-12-31
Sensor_name: ANEMOMETER
Sensor_name: BAROMETER
Sensor_name: RAIN GAUGE
Sensor_name: THERMOMETER
Source_name: GROUND STATIONS
Group: Author
First_name: PETER
Middle_name: L.
Last_name: GRIMM
Email: INTERNET > GRIMM@NODC2.NODC.NOAA.GOV
Email: NSI/DECnet > NODC2::GRIMM
Phone: (202) 606-5008
Group: Address
Environmental Information Services
NOAA/NESDIS/EIS Ex2
Room 506
1825 Connecticut Avenue, NW
Washington, DC 20235 USA
End_Group
End_Group
Group: Data Center
Data_center_name: NOAA/NESDIS/NODC/LISD > Library
Information Services Div.
Dataset_ID: C/DB V872C, V872

Group: Data Center Contact
Phone: 301-443-8330
Phone: FAX 301-443-0237
Group: Address
NOAA Central Library
6009 Executive Blvd.
Rockville, MD 20852

End_Group
End_Group
End_Group
Storage_Medium: Soft-cover and hard-cover books (4)
Parameter: ATMOSPHERIC COMPOSITION > CLOUDS
Parameter: ATMOSPHERIC DYNAMICS > ATMOSPHERIC
TEMPERATURE > Surface Temperature Frost
Parameter: ATMOSPHERIC DYNAMICS > ATMOSPHERIC
TEMPERATURE > Surface Temperature
Parameter: ATMOSPHERIC DYNAMICS > CLOUD TYPES
Parameter: ATMOSPHERIC DYNAMICS > PRECIPITATION
Parameter: ATMOSPHERIC DYNAMICS > PRESSURE > Surface
Pressure
Parameter: ATMOSPHERIC DYNAMICS > WINDS > Surface
Winds
Parameter: GEOGRAPHY AND LAND COVER > ICE
Parameter: GEOGRAPHY AND LAND COVER > LAKES
Parameter: GEOGRAPHY AND LAND COVER > RIVERS >
Streamflow
Parameter: GEOGRAPHY AND LAND COVER > RIVERS > Water
Level
Parameter: GEOGRAPHY AND LAND COVER > SURFACE
VEGETATION > Harvest
Parameter: GEOGRAPHY AND LAND COVER > SURFACE
VEGETATION > Growth

Parameter: HYDROLOGIC PARAMETERS > PRECIPITATION
Parameter: HYDROLOGIC PARAMETERS > RIVERS >
Streamflow
Parameter: HYDROLOGIC PARAMETERS > RIVERS > Water
Level
Parameter: HYDROLOGIC PARAMETERS > RUNOFF
Parameter: HYDROLOGIC PARAMETERS > TEMPERATURE >
Surface Temperature
Discipline: EARTH SCIENCE > ATMOSPHERE
Group: Coverage
 Minimum_Latitude: 30N
 Maximum_Latitude: 80N
 Minimum_Longitude: 10E
 Maximum_Longitude: 180E
End_Group

Location: ASIA > USSR
Location: ASIA > Russia
Location: MID-LATITUDE
Keyword: ANNUAL WEATHER DATA
Keyword: FROST
Keyword: GROWTH
Keyword: HARVEST
Keyword: MONTHLY WEATHER DATA
Keyword: RUSSIA
Keyword: SEASONAL WEATHER DATA
Keyword: STREAMFLOW
Keyword: SURFACE
Keyword: SURFACE PRESSURE
Keyword: SURFACE TEMPERATURE
Keyword: SURFACE WEATHER DATA
Keyword: SURFACE WINDS

Keyword: USSR
Keyword: WATER LEVEL
Revision_Date: 1992-05-10

Group: Summary
The title of these books are:
Radical differences in the Climate of Russia
Le Climat De L'Empire Russe et la Meteorologie en Russie et en
Siberie
(Written in French)

The call numbers are: C/db V872 C/db V872c
Number of books: 4

These books are part of the foreign meteorological data collection held
by the NOAA Central Library in Washington, DC. Information in this
collection dates back to the 18th century for daily, monthly, seasonal,
and annual tabular summaries, and the 19th century for weather maps.
These data are the result of foreign exchange agreements, but the
collection has not been updated since 1983.

These books contain monthly, seasonal, and annual weather data tables
for the USSR. These data are for the period 1700 to 1891. These
tables contain the following surface data averaged over the period:
mean monthly and seasonal average, maximum, and minimum
temperature, pressure, cloudiness, precipitation, and wind direction
frequencies; and mean annual temperature, rainfall, frost-free days, and
days for formation and breakup of the ice on the rivers.

In addition, there are tables containing data for mean harvest time and
annual days of growth for rye, wheat, and oats. The data for individual
years are seasonal and annual flow rates and levels of rivers and lakes.
End_Group



5



Possible links to COTAS
follow on in Data Review & Access

LIBRARY DATA DESCRIPTION SELECTED FROM THE NOAA DIRECTORY
G.BARTON 202-606-4548 JULY 16, 1993

TITLE >>>

Daily and Monthly Temperature for Penn State College (1886- 1913)

BRIEF >>>

The title of this book is:

Annual Report of the Penn State College Agricultural Experimental Station

The call number is: C/gf 385 S79 oe C2

Number of books: 1

This book is part of the foreign meteorological data collection held by the NOAA Central Library in Washington, DC. Information in this collection dates back to the 18th century for daily, monthly, seasonal, and annual tabular summaries, and the 19th century for weather maps.

These data are the result of foreign exchange agreements, but the collection has not been updated since 1983.

In these summaries, definitions are:

1. Hourly/daily/monthly/seasonal/annual data -- values determined for each consecutive period, e.g. monthly temperature data: Jan 1900, Feb 1900, etc.
2. Long-term data -- values determined over a period of years, e.g. long-term monthly temperature data using Jan 1901, Jan 1902, etc.

This book contains tables of daily and monthly surface temperature for Penn State College from 1886 to 1913.

Tables contain the following data:

- monthly mean (given hour) temperature
- monthly absolute maximum and minimum temperature
- long-term daily mean temperature.

ATTRIBUTES >>>

Entry_ID: NCL00564 (MD Identifier: 3693)

Temporal Coverage:

From: 1886-01-01 TO: 1913-12-31

Geographic Coverage:

Southwest Extent: 40N,78W Northeast Extent: 41N,77W

Source:

GROUND STATIONS

Sensor:

THERMOMETER

Storage Media:

Hard Cover Book

Discipline, Subdiscipline:

EARTH SCIENCE > ATMOSPHERE

Location Keyword:

MID-LATITUDE
NORTH AMERICA > United States > Pennsylvania > State College

Parameter Group, Parameter:

ATMOSPHERIC DYNAMICS > ATMOSPHERIC TEMPERATURE
SurfaceTemperature

General Keywords:

DAILY DATA
MONTHLY DATA
PENNSYLVANIA
STATE COLLEGE
SURFACE
SURFACE TEMPERATURE
UNITED STATES

Revision Date: 1993-06-11

Quality Statement:

DATACENTER >>>

Archive:
NOAA/NESDIS/NODC/LLSD > Library Information Services Div.
Data Set ID: C/GF 385 S79 OE C2 (Annual Report of the Penn State College
Agriculture Experimental Station)

Contact: NOAA Central Library,
6009 Executive Blvd.
Rockville, MD 20852 USA

Electronic Mail: NSI/DECNET > NODC::SERVICES
TELEMAIL > NODC.WDCA/OMNET

Phone: (301) 443-8330
FAX (301) 443-0237

PERSONNEL >>>

Entry Author: EEBLUM, DANIEL CHARLES
Scientific and Commercial Systems Corporation
4651 King Street, Suite 200
Alexandria, VA 22302 USA
Phone: (301) 210-0250

Information in this entry provided by NOAA/NESDIS/NODC/LLSD

REFERENCE >>>

SUP_MENU >>>

***** No Supplementary Information available for this selection *****

NOAA DIRECTORY CATALOG INTEROPERABILITY

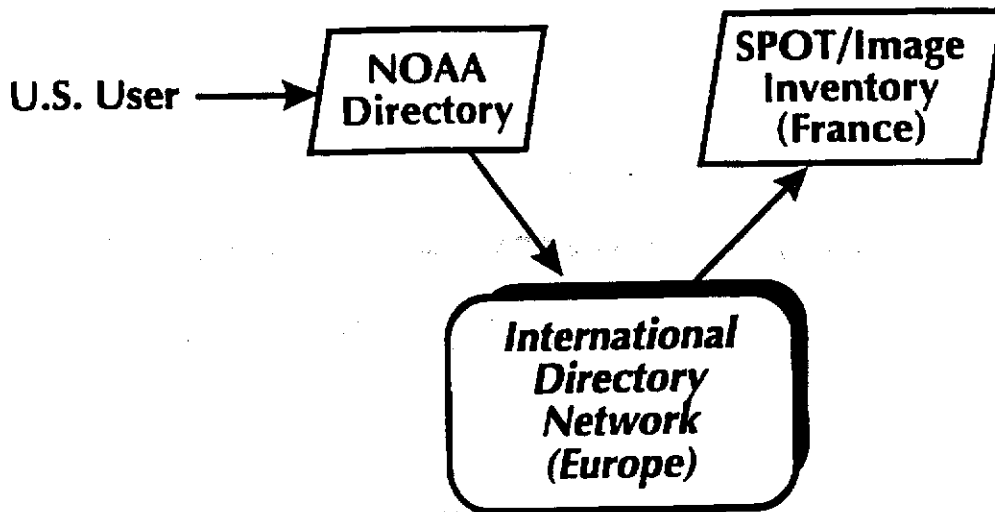
DESCRIPTIONS AND LINKS TO IDN NODES

GLOBAL CHANGE MASTER DIRECTORY
EUROPEAN SPACE AGENCY
JAPANESE SPACE AGENCY

LINKS TO NOAA DATA SYSTEMS

STORM INVENTORY
CLIMATE ANALYSIS CENTER
ERL DATA
NODC INVENTORY SYSTEM
KUWAIT DATA

Using the International Directory Network



20

GLOBAL CHANGE RESEARCH INFORMATION OFFICE

GCRIO

Gerald S. Barton, Director



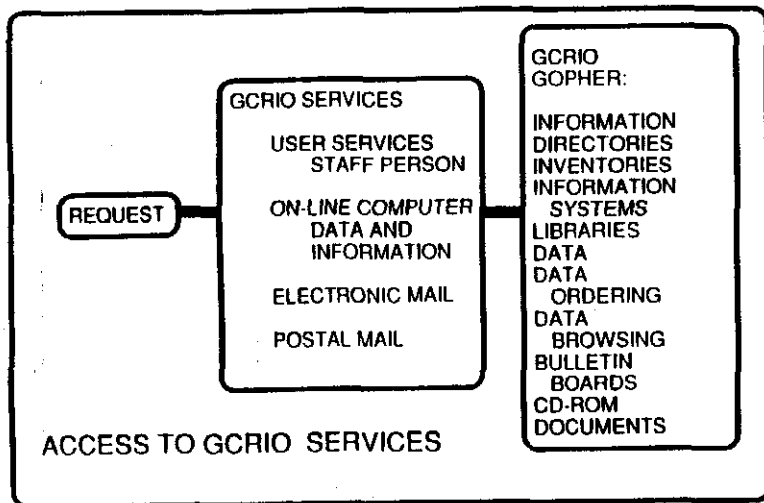
GCRIO

**U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE**

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX -6622

GCRIO is operated in CIRESIN for the CEES Subcommittee on Global Change Research

21



U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX 6622

GCRIO is operated by CIESIN for the U.S. Global Change Research Program

GCRIO ESTABLISHED BY 1990 PUBLIC LAW 101-606

THE GLOBAL CHANGE RESEARCH ACT OF 1990

Section 204 requires GCRIO

"to disseminate to foreign governments, businesses, and institutions, as well as citizens of foreign countries, scientific research information available in the United States which would be useful in preventing, mitigating, or adapting to the effects of global change."



U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX 6622

GCRIO is operated in CIESIN for the CEES Subcommittee on Global Change Research

AREAS OF EMPHASIS OF THE GLOBAL CHANGE RESEARCH ACT OF 1990

1. Reducing energy consumption
2. Solar energy and renewable energy resources
3. Replacements for CFC and other ozone-depleting substances
4. Conservation of forest resources
5. Ecological pest management, and proper use of agricultural and industrial chemicals
6. Recycling and source reduction of pollutants



U.S. GLOBAL CHANGE RESEARCH
INFORMATION OFFICE

1825 K STREET NW SUITE 805
WASHINGTON DC 20006
202-775-6628 FAX 6622

GCRIO is operated in CIESIN for the CEES Subcommittee on Global Change Research

GCRIO

U.S. Global Change Research Information Office
1825 K Street NW Suite 805
Washington DC 20006
202-775-6628 Fax 202-775-6622

GCRIO SERVICES

The services of the GCRIO are being established to allow users to access a variety of data and information sources:

Telephone calls to GCRIO User Services staff
in Washington DC 202-775-6607 fax -6622
in Saginaw Michigan 517-797-2727 fax -2622

Computer access to on-line GCRIO information services
Internet access gopher.gcrio.org

Electronic Computer Mail
help@gcrio.org
info@gcrio.org

Postal mail access

The GCRIO services use the Global Change Data and Information System (GCDIS) under development for the U.S. Global Change Research Program, and other on-line data and information services including:

bulletin boards	CD-ROM
data browsing	data directories
data inventories	information systems
libraries	on-line data
on line data ordering	on-line documents.

The GCRIO is implementing a GOPHER menu system on the Internet which will identify these resources, and allow the user to connect to some of these on-line systems.

GCRIO is operated by CIESIN for the U.S. Global Change Research Program

GCRIO

U.S. Global Change Research Information Office
1825 K Street NW Suite 805
Washington DC 20006
202-775-6628 Fax 202-775-6622

U.S. GLOBAL CHANGE RESEARCH INFORMATION OFFICE

In 1990, Congress passed and the President signed Public Law 101-606, the Global Change Research Act of 1990. The purpose of the legislation was "to require the establishment of a United States Global Change Research Program aimed at understanding and responding to global change, including the cumulative effects of human activities and natural processes on the environment, to promote discussions towards international protocols in global change research, and for other purposes."

Under Title II (International Cooperation in Global Change Research) of the Act, Section 204 requires that a Global Change Research Information Office (GCRIO) be established. The stated purpose of the GCRIO is "to disseminate to foreign governments, businesses, and institutions, as well as citizens of foreign countries, scientific research information available in the United States which would be useful in preventing, mitigating, or adapting to the effects of global change."

In May 1992, the Subcommittee on Global Change Research (SGCR) of the Committee on Earth and Environmental Sciences (CEES) designated that the GCRIO be implemented within the Consortium for International Earth Science Information Network (CIESIN).

GCRIO is operated by CIESIN for the U.S. Global Change Research Program

**BROCHURES DESCRIBING ACCESS TO SERVICES
ON THE INTERNATIONAL DIRECTORY NETWORK**

For information contact:

**Gerald S. Barton
Director, Global Change Research Information Office
1825 K St. NW Suite 805
Washington DC 20006
202-775-6628 fax -6622
gbarton@mail.ciesin.org
g.barton on OMNET**

November 3, 1993

NOAA Environmental Services Data Directory:

A link to worldwide information resources



The NOAA Environmental Services Data Directory (NOAADIR) provides a key to both managing and accessing global change data in NOAA. The Directory is being used to document NOAA data sets and data systems that support global change and other earth science studies.

To provide improved access and enhanced service to users, the NOAA Environmental Services Data Directory has been installed on a Sun SPARC-10 workstation at NOAA's Environmental Services Data and Information Manage-

ment (ESDIM) office. WAIS, the Wide Area Information Server, is installed on the workstation to provide Internet users with access to a full text version of the NOAA Directory, as well as the National Environmental Data Referral Service (NEDRES), and the NOAA Product Information Catalog. Users can access the Directory at no cost through national telecommunication systems such as Internet or through dial-in telephone lines including an 800 number.

The NOAA Directory is one of the directories in the Global Change Master

Directory System. This national and international system of directories uses the Directory Interchange Format (DIF) for the exchange of data descriptions between directories. The international Committee on Earth Observation Satellites sponsors the International Directory Network (IDN) that has three Coordinating Nodes: (1) the NASA Master Directory; (2) the European Space Agency Directory in Frascati, Italy; and (3) the Japanese Directory at the National Space Development Agency.

All NOAA data descriptions entered into the NOAA Directory are transferred to the Global Change Master Directory maintained at the National Space Science Data Center, NASA Goddard Space Flight Center, Greenbelt, Maryland. Through interconnections to the three IDN Coordinating Nodes, NOAA Directory users can access worldwide data directory information.

The NOAA Directory continues to grow and now contains descriptions of over 2,000 NOAA data sets. Recent additions include about 350 descriptions of foreign meteorological data reports in the collections of the NOAA Central Library, Rockville, Maryland.

*- Gerald Barton
Environmental Information Services
NOAA/NESDIS Ex2
1825 Connecticut Avenue, NW
Washington, DC 20235 ■*

Access to worldwide data directory information

NOAA ENVIRONMENTAL SERVICES DATA DIRECTORY

The NOAA Directory is a tool for locating NOAA data sets. You can access the Directory using your PC.

Terminal settings:

Full duplex, 8 bits, no parity, one stop bit, 1200 baud
Terminal type: VT-100 (preferred)

Via Internet

At the \$ prompt, enter: TELNET ESDIM1.NODC.NOAA.GOV
At the prompt USERNAME:, enter: noaadir (use lower case)

Via direct dial (1200 baud):

In the Washington, D.C. area
and outside the United States, dial: 202-606-4665, 202-606-4666, or
202-234-7123, -7143, or -7033

In other areas of the United States, dial: 1-800-722-5511

At the prompt XT_COMMAND, enter: C ESDIM1
(If you do not see the XT_COMMAND prompt, enter the break
key several times until it appears.)

At the prompt LOGIN:, enter: noaadir (use lower case)

When you end the session, enter the BREAK key.

At the XT_COMMAND, enter: DIS
This disconnects the link to the Sun workstation, and you can hang up.

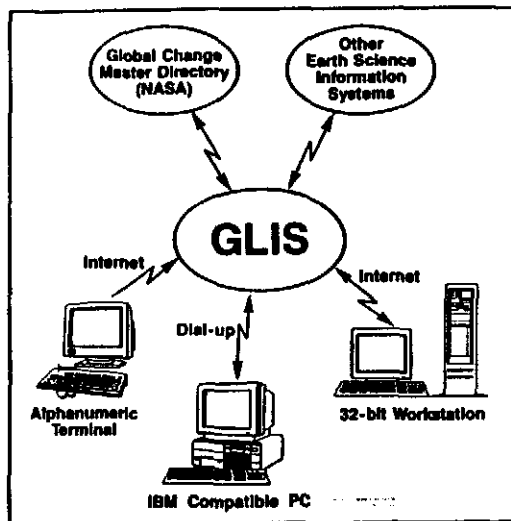
A system of menus and prompts guides users through each search session. HELP is available at any point in a session, or call Gerry Barton at 202-606-5012.

NOAA personnel who know of data sets that should be described in the Directory should contact their NOAA Directory LO Team Member or call Gerry Barton at 202-606-5012.

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service

GLIS System Access

Direct access to GLIS is through wide-area networks and through dial-up telecommunications interfaces. GLIS is one of the principal disciplinary nodes of the national Global Change Data and Information System (GCDIS). Access to GLIS is available through the Global Change Master Directory, an element of the GCDIS. GLIS also provides network linkages to and from other earth science information systems. For further system access information, please contact GLIS User Assistance.



Global Land Information System access

GLIS Access Procedures

Users wishing to perform graphic-based query functions can use an X terminal or X terminal emulator package on a PC.

From
INTERNET: \$TELNET xglis.cr.usgs.gov

Users wishing to perform text-based query functions can use an alphanumeric terminal or a terminal emulator package on a PC.

From
NSI/DECNET: \$SET HOST GLIS
USERNAME: GLIS

From
INTERNET: \$TELNET glis.cr.usgs.gov
or \$TELNET 152.61.192.54

Direct Dial: Set modem to 8 bits, no parity, 1
stop bit Dial: (605) 594-6888

GLIS User Assistance and Information on Obtaining Graphical Interface Software

Telephone: 1-800-252-GLIS (1-800-252-4547)
or commercial: (605) 594-6099

E-mail: NSI/DECNET GLIS:GLIS
GSFCMAIL EROS.DATA.CENTER
OMNET EROS.DATA.CENTER
INTERNET GLIS@GLIS.CR.USGS.GOV

Mail: U.S. Geological Survey
EROS Data Center
GLIS User Assistance
Sioux Falls, SD 57198 USA

The use of trade names does not imply an endorsement by the U.S. Government.

* U.S. GPO 1993-756-819

National Mapping Program

Global Land Information System



U.S. Department of the Interior
U.S. Geological Survey

What is GLIS?

The Global Land Information System (GLIS) is an interactive computer system developed by the U.S. Geological Survey (USGS) for scientists seeking sources of information about the Earth's land surfaces. GLIS contains "metadata," that is, descriptive information about data sets. Through GLIS, scientists can evaluate data sets, determine their availability, and place online requests for products. GLIS is more, however, than a mere list of products. It offers online samples of earth science data that may be ordered through the system.

Scientists can use GLIS to bring up outlines of the geographic areas covered by the data sets. Using digital browse functions to manipulate the data, they can determine such information as the amount of cloud coverage or the quality of individual scenes.

Online requests can be placed via GLIS for earth science data. The producing organization will receive the request and provide the researcher with price and ordering information.

What Kinds of Data are Available Through GLIS?

GLIS contains references to regional, continental, and global land information including land use, land cover, and soils data; cultural and topographic data; and remotely sensed satellite and aircraft data. Continual updates of information and the addition of new data set descriptions, as they are contributed by the global change scientific community, will allow GLIS to remain current.

Functional Overview of GLIS

GLIS data set descriptions are arranged in three levels of detail — directories, user guides, and inventories.

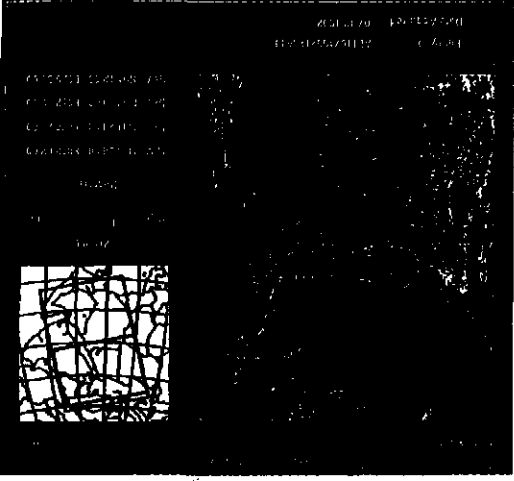
Directories contain summary information about each data set. Searches of a data set's directory may be based on any combination of keywords (discipline, location, geographical parameter), acquisition date, data source (spacecraft, sensor), geographic coverage, project, and investigator.

User Guides contain detailed information about the data set such as sensor specifications, extent of coverage, processing history, data quality, and product availability. User guide queries are by hypertext graphics and text retrieval.

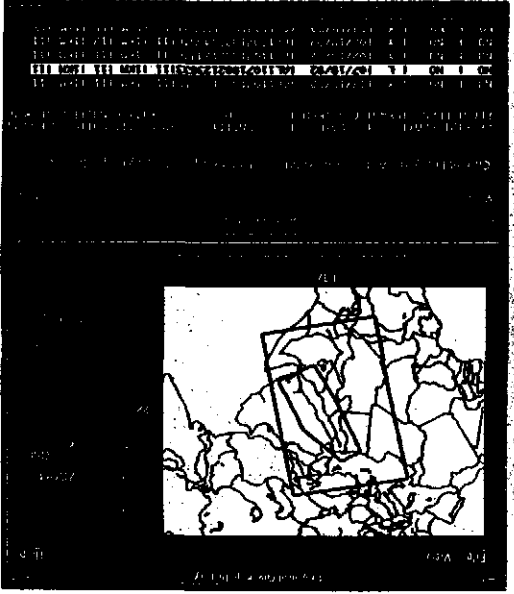
Inventories contain detailed information about individual data set entities such as Landsat scene or an Advanced Very High Resolution Radiometer (AVHRR) pass. Temporal, spatial, and parameter-oriented queries of inventories are possible.

In addition to the directories, user guides, and inventories, GLIS graphic services are available using the X-GLIS graphical user interface which supports the X Window System environment on 32-bit UNIX workstations. The services include user specification of geographic search parameters, geographic coverage outline, and digital image browse capabilities. Similar graphic extensions are available for IBM-compatible personal computers.

Interactive browse screen including an AVHRR image.



X-GLIS screen showing geographic coverage of an AVHRR image.





Gcnet

Quick Reference Guide



What is Gcnet ?

Gcnet (formerly the Global Change Network) was developed at the Canada Centre for Remote Sensing (CCRS) to serve as a single point of contact for global change researchers, scientists and users of remote sensing information. It is a free on-line system that directs users to pertinent international data sets and other up-to-date information. The following is now available:

IDN Directory Service

Users can access a centralized, multidisciplinary directory of scientific data sets which identifies Canadian and International data sets pertinent to global change research. Directory access is done through the "Master Directory" which is part of the "International Directory Network" (IDN). Information from the following participating Countries include: Antarctica, Canada, France, Germany, Italy, Japan, Russia, the United Kingdom and the United States. Future anticipated nodes include: Australia, Brazil, China and the Netherlands.

Data Centre Links

This choice permits users to link to other data centres world-wide and access their inventories, bulletin boards or information networks. This feature is part of our IDN directory service.

CCRS Image Inventory

This option invokes the CCRS QUERY program which permits searches of the LANDSAT, MOS, NOAA and SPOT satellites raw image inventories. A products catalogue of NOAA geocoded and composite products processed on the GEOCOMF system is also available. Results can be viewed on screen, mailed via Internet, NSI/DECnet, or sent by surface mail. An ERS-1 image inventory will be available soon.

Planet Bulletin Board Services

A directory of the products and services offered by the Surveys, Mapping and Remote Sensing Sector provides comprehensive price and contact details. Descriptions of CCRS projects, an up to date calendar of events and an image file transfer area are just a few of the features on the CCRS bulletin board option. An index of private sector companies and profiles of special interest groups complete an informative survey of the Canadian remote sensing community.

RESORS

This option provides users with information on how to get an account on the CCRS document retrieval service, RESORS. RESORS is a unique on-line bibliographic database that provides rapid and precise access to information on the technologies and applications of remote sensing world-wide.

Gcnet Access Procedures

International access procedures are as follows:
(The user input has been highlighted.)

From NSI/DECnet (formerly SPAN):
\$ SET HOST CCRS (or SET HOST 18732)
Username: Gcnet

From INTERNET:
\$ telnet gcnrl.ccrs.ensr.ca (or telnet 132.156.47.218)
Username: gcnrl

From DATAPAC (Canada only):

Set modem to 8 bits, NO parity, 1 stop bit (when using the downloading feature, please use 7 bits, EVEN parity, 1 stop bit and the Kermit protocol).
The CCRS DTB address is 20800336
Username: Gcnet (To connect to the Gcnet Computer.)
Username: Gcnet (To access the Gcnet Main Menu.)

Direct Dial:

Set modem to 8 bits, NO parity, 1 stop bit,
300, 1200, 2400 or 9600 bps (V.32 compliant).

Dial (613) 941-8434
Enter username > Gcnet
Local > CONNECT Gcnet
Username: Gcnet

Gcnet User Assistance and Information

On-line HELP is available, however, if you need further assistance or more detailed information about Gcnet, please contact:

Gcnet Database Coordinator

Canada Centre for Remote Sensing
588 Booth Street, 2nd Floor

Ottawa, Ontario
Canada

K1A 0Y7

Telephone: (613) 947-1298

Fax: (613) 947-1408

E-mail:

NSI/DECnet > CCRS::GCNETADMIN (or 18732::GCNETADMIN)
INTERNET > gcnrl@crrs.ensr.ca

The CCRS Image Inventory

The CCRS Image Inventory service provides users with current information on raw satellite images of Canada being archived at CCRS. The following satellites and sensors are being maintained in the catalogue:

LANDSAT MSS	(since 1972)
LANDSAT TM	(since 1984)
SPOT PLA	(since 1986)
SPOT MLA	(since 1986)
NOAA AVHRR	(since 1991)
MOS MESSR	(since 1988)
SEASAT SAR	(1978)

There is also an inventory of NOAA AVHRR composite products at 1 km and 4 km resolution generated on the GEOCOMP system at the Manitoba Remote Sensing Centre. An ERS-1 SAR raw imagery catalogue of data acquired over Canada will soon be accessible.

The Image Inventory service allows users to search the catalogues using a number of parameters, for example; satellite, sensor, acquisition date, latitude/longitude or track/frame. The search results contain additional details such as cloud coverage, synchronization losses, that are useful to assess the general quality of the scene. Results from the catalogues can be viewed on screen immediately. A copy can also be returned via mail; Canada Post, or E-mail; Internet, NSI-DECnet (SPAN).

Most imagery catalogues are updated within two days of reception.

To place an order for the images, users are directed to the SMRSS Products and Services option on GCNet for the satellite order desk details.

April 1993

The International Directory Network (IDN) Directory Service

The IDN provides directory-level information on scientific data sets available throughout the world. It is a valuable tool that enables researchers to quickly identify, locate and possibly order specific data sets. Information on socio-economic data sets are currently being added. This on-line service is free and no account is required.

The information in the IDN is divided into five main disciplines: Earth science (geoscience, hydrospheric, biospheric, satellite remote sensing, atmospheric sciences), Space Physics, Solar Physics, Planetary Science and Astronomy. Keywords, prepared by scientists/researchers within each discipline, have been used to allow users to quickly identify required data.

The following is a sample of the information retrieved in an IDN directory search: data set description, geographic and temporal coverage, physical parameters measured, data centre and contact where data set is archived, data set storage format, access procedures, technical references, and information about the platform and sensor used to acquire the data set. Some directory entries allow the user to electronically link to the archive data centre to access more information and possibly browse or order the data.

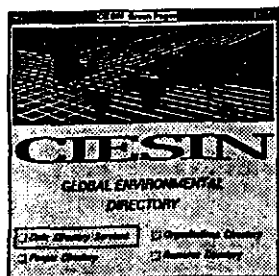
Users who have valuable data sets to be included into the IDN should contact the GCNet Database Coordinator to acquire the directory-entry submission format.

The IDN is sponsored by the Committee on Earth Observations Satellites (CEOS).

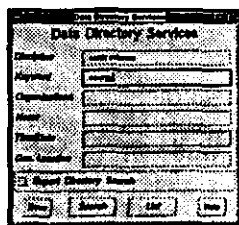
April 1993

Getting Started with the Catalog

X-Windows Interface:

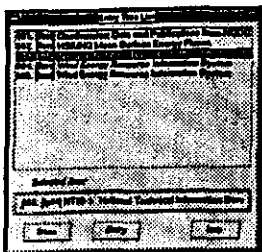


Main menu screen

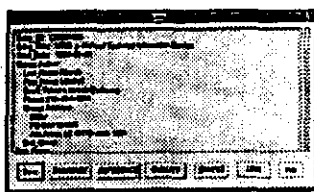


Basic search screen

To access the Catalog, select **Data Directory Services** from the CIESIN Main menu screen. The Basic search screen will appear on your display. From this screen you can enter search parameters for the attributes shown and press the **SEARCH** button to perform a directory search. Or, you can select **Expert Directory Search** to proceed to a screen where complex search specifications may be entered. Once the search is complete, a list of matches is presented, from which you can select a particular DIF entry to display.



Entry selection screen



DIF Entry display

Text-based Interface:

To perform a directory search, type **search** followed by a specification in the form **attribute = value**. Complex searches may be formed by using logical operators between pairs of search specifications: **&** (and); **|** (or); **~** (but not). Parentheses can be used to group several search specifications to produce more complex queries. The following are some example queries*:

```
gp> search discipline=earth science
gp> search discipline=health &
keyword=toxicity
gp> search (keyword=ozone | keyword=aurora) -
(location=europe)
```

When the search is complete, a list of matches is presented. Type **h** for help on how to scroll through the list; type **q** to return to the **gp>** prompt. The content of an entry is shown by typing the appropriate command followed by the entry number. Type **?** for a list of available commands; use **HELP** for detailed instructions.

* Queries must be entered on a single line

User Assistance and Information

If you need assistance using the Catalog or would like more information, please contact:

Polytechnic University	Telephone: (718) 260-3230
CIESIN Activities	FAX: (718) 260-3930
Five Metrotech Center	E-mail: INTERNET >
Brooklyn, NY 11201	info-gp@poly.edu
USA	



Data Catalog System

(previously known as "GREENpages")

Quick Reference Guide

Polytechnic University
CIESIN Activities

Revision 0.4
April, 1993

About the CIESIN Data Catalog System

The CIESIN Data Catalog System (hereafter referred to as "the Catalog") is a distributed system providing convenient access to cataloged information. It was developed by the Consortium for International Earth Science Information Network (CIESIN) Project at Polytechnic University. The Catalog aims to assist users in locating and retrieving information concerning global change.

Information currently accessible through the Catalog includes CIESIN created entries related to human dimensions of global change. The system is also seamlessly interoperable with the extensive catalog of the Global Change Master Directory (GCMD). Entries are currently stored in the *Directory Interchange Format* (DIF) standard or as plain text, although in the near future other standards (such as MARC) will also be available. The Catalog also provides samples of some directory entries to help the user locate and evaluate specific data of interest. Access to *Inventory level* metadata is being implemented.

The Catalog is a client/server based system. The client program runs on a computer connected to the Internet, communicating with various servers on the network. Access is available via Internet or modem dialup.

The distributed nature of the system allows for the linking of many locally managed data collections. Currently, the Catalog is composed of 5 "nodes" (servers). One node holds CIESIN's DIF entries; another node communicates directly with the GCMD (client/server based) database to retrieve its holdings; the remaining nodes are currently utilized for testing purposes.

Accessing the Catalog

There are currently three client programs:

- X-Windows/Motif Graphical User Interface (GUI)
- Microsoft Windows PC Interface
- Text-based query system (alpha-numeric interface)

All are available from any host connected to the Internet, and the Windows and text-based systems also provide direct dial access. [Contact CIESIN for information on how to obtain the Microsoft Windows client program.]

In the following instructions, user input is indicated in **bold** font, computer responses are in *typewriter* font (computer responses are abbreviated).

Connecting via the Internet:

- If you wish to use the X-Windows interface, open an *xterm* window and type **xhost +sirius.poly.edu** (or **xhost +128.238.32.62**)
- telnet sirius.poly.edu**
(or **telnet 128.238.32.62**)

Connecting via Direct Dial:

- Set modem to 8 bits, NO parity, 1 stop bit, Any baud rate up to 9600
- Dial (718) 260-4000

**** Welcome to the Polytechnic University Terminal Server! ****

Poly-TS> **sirius**
Trying SIRIUS.POLY.EDU (128.238.32.62)... Open

Upon successful connection, access as follows:

SunOS UNIX (sirius)

login: **gp**
SunOS Release 4.1.2 (CALC:2)

Welcome to the GreenPages Service at Polytechnic University

Please enter your e-mail address (optional):
<enter your electronic mail address>

X-Windows access:

If you're using X-Windows, the following is displayed:

For X window access, enter your DISPLAY name. Otherwise, if you do not wish to use X, enter 'none'.

DISPLAY (default=your.node.name:0.0)=

If the default display shown is correct, press **RETURN**. If not, enter the correct display (e.g., **pc.umich.edu:0**). The terminal connection will be closed, and the CIESIN title screen will soon appear on your display.

Text-based access:

The following is displayed:

Try: '?' for a list of commands
'report' to send a report to the
greenpages manager

gp running on host sirius
using isode 8.0 #18 (sirius)
gp>

Enter commands at this prompt.

New Master Director Interface Offers Many Enhancements

Patricia Bailey and Janis Shipe, Hughes STX

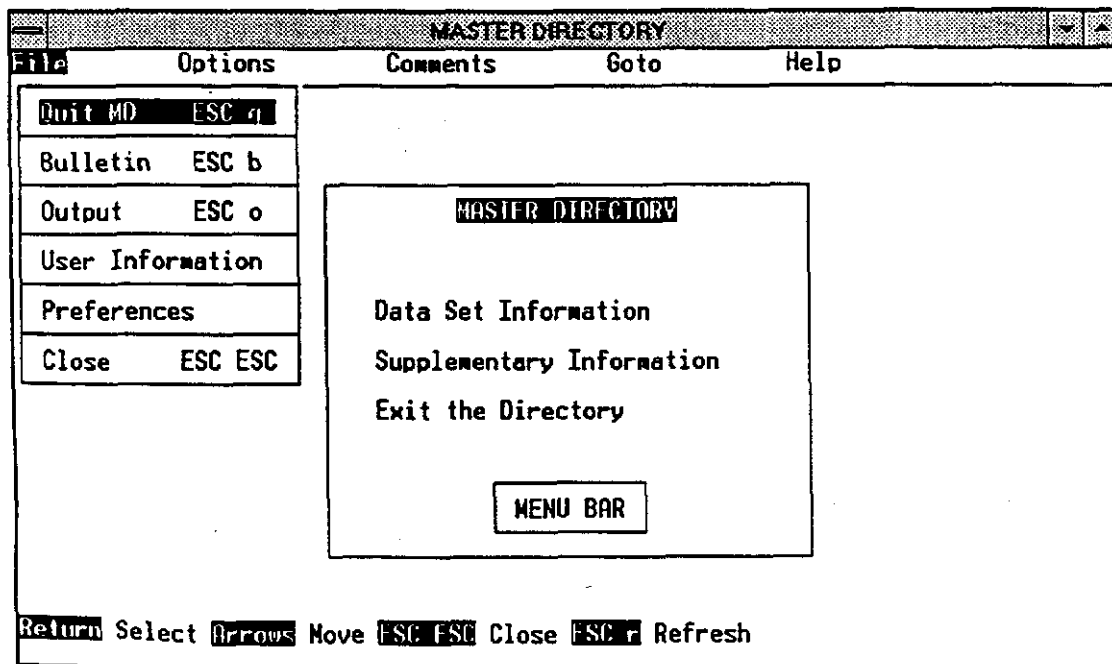
Throughout the lifecycle of the Master Directory (MD), the system and interface have evolved through interactive dialogue among system developers and science users. Initially, the MD system was developed based on requirements constructed by the Catalog Interoperability Working Group. This design and implementation included a database structure to store directory information, a data loader, and a user-friendly, menu-driven interface. Subsequent development tasks included a data extractor, data deletion utilities, database support utilities, and numerous reporting capabilities.

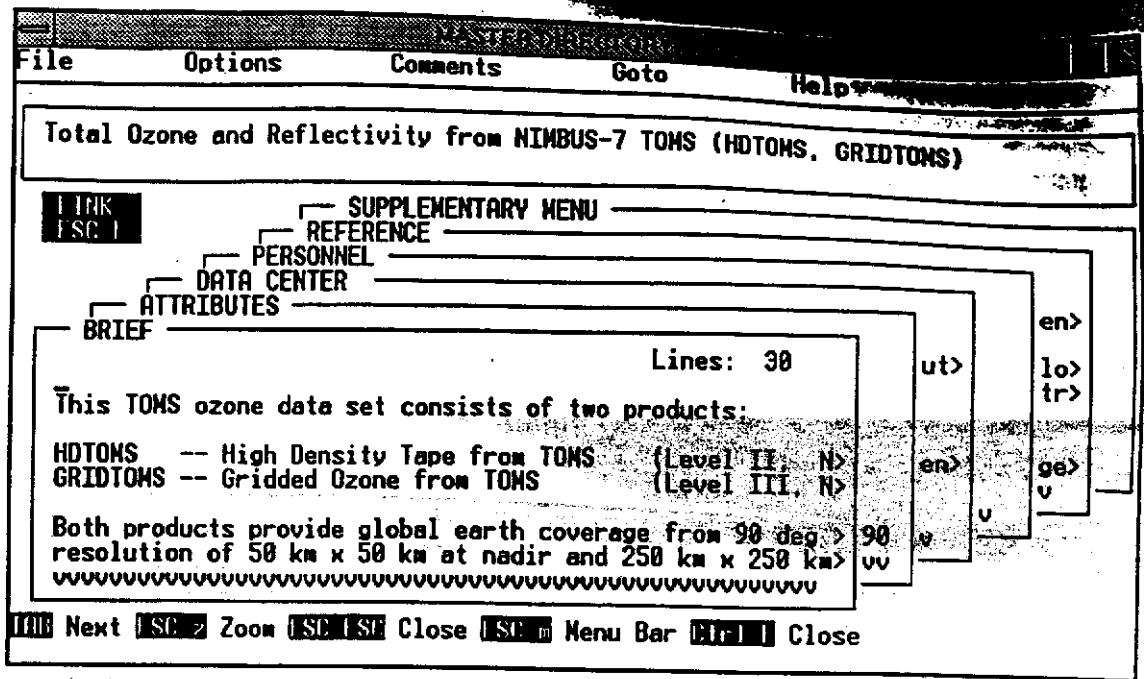
The introduction of the MD into the international science community altered the scope of the MD task. The new International Directory Network (IDN) required special tools for automatic transfer and loading of directory information between nodes. In addition,

technical training and support were provided to each of the IDN nodes. The IDN environment along with the changing needs of the science community led to a second design initiative and new implementation. The current MD system design includes a client/server architecture and many enhancements as suggested by the science community. Several client applications are being developed including an alphanumeric client using JYACC Application Manager (JAM) and XWindows. Also, the original MD system is being converted to take advantage of the new client/server design. Currently, both the MDI interface and the new the JAM client are available to the user community via the NODIS system at the NSSDC.

New features of the MD JAM client

The new MD client has many enhancements that offer extended capabilities over Version 1.





Pull-down menus provide access to automatic help and valids, session output, user information profile and session preferences, as well as a GOTO function. Titles resulting from a query may be sorted by one of several available attributes. The information display is presented on six "index cards" each containing a section of the directory entry. These cards may be zoomed to full screen display for easier viewing.

Accessing the Master Directory

NSI/DECnet

Set host: NSSDCA

Username: NODIS

Select option #1 from the menu

Answer "YES" to use Version 2

Internet

Telnet: 128.183.36.23

Username: NODIS

Select option #1 from the menu

Answer "YES" to use Version 2

Direct Dial

Set modem to 8 bits, no parity

1 stop bit for 300, 1200 or 2400 baud

Dial (301) 286-9000

For 9600 baud, dial (301) 286-4000

Prompt: Enter Number, you enter "MD"

Prompt: Call Complete, you enter <CR>

Username: NODIS

Answer "YES" to use Version 2

Where to get help

Contact the Master Directory User Support Office (MDUSO) for help on any aspect of the MD. The MDUSO will direct you to the appropriate MD team member who can assist you with questions about using the MD, about the specific contents of the MD, with preparing new data entries, and with establishing LINK connections. The MDUSO can be reached :
(301) 513-1687; NSI/DECnet: NCF::MDUSO;
Internet: MDUSO@nssdca.gsfc.nasa.gov

Global Change
 The MD has been adopted by the Interagency Working Group on Data Management for Global Change (IWGDMGC) as a centralized directory to facilitate global change research in the U.S.A. The IWGDMGC members include DOE, EPA, NASA, NOAA, NSF, USGS and other U.S. agencies.



International Directory Network (IDN)
 The MD is an international directory developed under the guidance of the Committee on Earth Observation Satellites (CEOS). IDN nodes exist at NASA Goddard Space Flight Center (GSFC) in Greenbelt, Maryland; Earthnet Program Office (EPO) in Frascati, Italy; National Space Development Agency (NASDA) in Japan; and at the Canadian Center for Remote Sensing (CCRS) in Ottawa, Canada.

Access Procedures
 The MD resides on a VAX computer and may be reached by several methods as describe below. The MD is Option #1 of the NSSDC's On-line Data Information Services (NODIS) account.

Via Direct Dial:
 Set modem to 8 bits, no parity, 1 stop bit, 300, 1200 (preferable), or 2400 baud, dial (301) 286-9000 or FTS 888-9000; for 9600 baud, dial (301) 286-4000 or FTS 888-4000.
 Enter Number: MD
 CALL COMPLETE <CR>
 USERNAME: NSSDC

Via NSI/DECnet (the old SPAN system):
 \$set Host NSSDC
 USERNAME: NSSDC

Via Internet:
 \$TELNET NSSDC.A.GSFC.NASA.GOV
 or \$TELNET 128.183.36.23
 USERNAME: NSSDC

From PINET:
 Select option 2 from the top menu (databases). The next menu option #1 will connect you to NSSDC Online Data & Information Service.

ASSISTANCE/INFORMATION
 Contact the MD User Support Office:

Master Directory
 NSSDC
 Code 633
 NASA Goddard Space Flight Center
 Greenbelt, Maryland 20771

(301) 513-1687
 NSI/DECnet: NCF:MDUSO
 Internet:

MDUSO@NSSDC.A.GSFC.NASA.GOV
 Telex: 248496 or 197640

December 1992



MD

A Scientific Data
 Information Service



Operated at NSSDC
 National Space Science Data Center
 NASA Goddard Space Flight Center
 Greenbelt, MD 20771 USA

Types of Services:
The MD provides five top level options to learn more about data.

Data Set Information
A search using selected keywords will provide you with a description of relevant data sets,

their attributes, spatial and temporal coverage, data center, names of persons to contact, and a bibliography. More detailed information, browse or data ordering capability is available through automated connections to other systems utilizing the LINK command.

Supplementary Information

Data Center Information
A description of the data center and its focus, contact information, access procedures, ordering and price policy, and available distribution media is provided. Direct access is provided to some data centers through the LINK command.

Campaign/Project Information
Scientific campaigns or projects, contact information, and their associated data bases are described.

Source/Spacecraft/Platform Descriptions
Information including orbital characteristics, mode of operation, attached instrumentation and dates of operation are described.

Sensor/Instrument Descriptions
The instruments and sensors used to obtain data are described. This includes the dates of operation, wavelength range, spatial and spectral resolution, and associated source/spacecraft.

The MD is a FREE on-line data information service.

The MD is MULTI-DISCIPLINARY:

data sets in Earth science (geology, oceanography, atmospheric science), space physics, solar physics, planetary science, astronomy/astrophysics are described

The MD is MULTI-AGENCY:
describes data held by NASA, NOAA, USGS, DOE, EPA and other agencies and universities.



MD Interconnections - Current	
ADP	Astronomical Data Center
AMPTE	Applied Physics Laboratory AMPTE System
ARIN	Aerospace Research Information Network
BATSE	Burst and Transient Source Experiment Solar Database
BR80	Big Bear Solar Observatory
BRUNET	UCLA Space Science Group Network
CBAT	Central Bureau for Astronomical Telegrams
CCRS	Canadian Center for Remote Sensing
DAU	CNES SPOT Image Catalog
EICS	Energy Ion Composition Spectrometer File
EINSTEIN	SAO Einstein Data Center
ENVIRONET	Space Environment Information Service
EOS	Earth Observation Center of Japan
ESA	European Space Agency
ESIS	European Space Information System
ESO-INFO	European Southern Observatory Information System
ESOSTARCAT	European Southern Observatory STARCAT Catalog System
EXOSAT	ESTEC Exosat Database System
GRIS	Coddard Institute for Space Studies
GLIS	Global Land Information System
HEASARC	High Energy Astrophysics Archive Research Center
INS	Incorporated Research Institutions for Seismology
IRPS	Image Retrieval and Processing System
IUE	International Ultraviolet Explorer
LANG	Langmuir Probe Catalog
LPI	Lunar Planetary Institute
MISO	Meer Solar Observatory
NACS	Neutral Atmosphere Composition Spectrometer File
NED	NASA/IPAC Extragalactic Database
NCDS	NASA Climate Data System
NDADS	NSDC Data Archive Distribution Service
NCDC	NOAA National Climate Data Center
SEL	NOAA Space Environmental Laboratory
NSF	National Science Foundation Information System
NSSDC	National Space Science Data Center Data Request System
OCEANIC	Ocean Network Information Center
OMNI	Interplanetary Medium Database
EOSDIS	Earth Observing System Data and Information System
CIRESIN	Consortium for International Earth Science Information Network
FLDS	Pilot Land Data System
FLDS-AMES	Pilot Land Data System at Ames Research Center
FLDS-CSRC	Pilot Land Data System at Coddard Space Flight Center
FLDS-JPL	Pilot Land Data System at Jet Propulsion Laboratory
PDS	Planetary Data System
PDS-ATMOS	Planetary Data System Atmospheres Node
PDS-CBO	Planetary Data System Conferences Node
PDS-PI	Planetary Data System Planetary Plasma Interactions Node
PDS-RINGS	Planetary Data System Rings Node
PDS-SBN	Planetary Data System Small Bodies Node
PNRA	National Program for Research in Antarctica
RESSTEC	Remote Sensing Technology Center of Japan
RCO	Royal Greenwich Observatory
SAL	Auroral Imaging File
SDAC	Solar Maximum Mission Data Analysis Center
SDCS	Synthetic Aperture Radar Data Catalog System
SDPAS	SWIR Data Display and Archival System
SEEDIN	Space Environment Realtime Data Intercommunication Network
SIMBAD	SIMBAD Database at CDS in France
SINRONIA	Global Observation Satellite Data Inventory Service
SAISS	Surveys Mapping and Remote Sensing Sector of Canada
SSEOP	Space Shuttle Earth Observation Program
STIS	Space Telescope Electronic Information System
STORM	Storm-scale Operations and Research Meteorology Program
SWRI	Southern Research Institute
TOMS	Total Ozone Mapping Spectrometer File
UA	University of Alaska Geodata Center
UARS	Upper Atmospheric Research Satellite Data Facility
UMI	University of Rhode Island AVHRR Archive
UTD	University of Texas at Dallas
VERI	Vector Electric Field Instrument Catalog
VILSPA	Villanova Satellite Tracking Station
VLA	NRAO Very Large Array Information System
WATSI	Wind and Temperature Spectrometer File
Dynamic Explorer Data Set Catalogs	

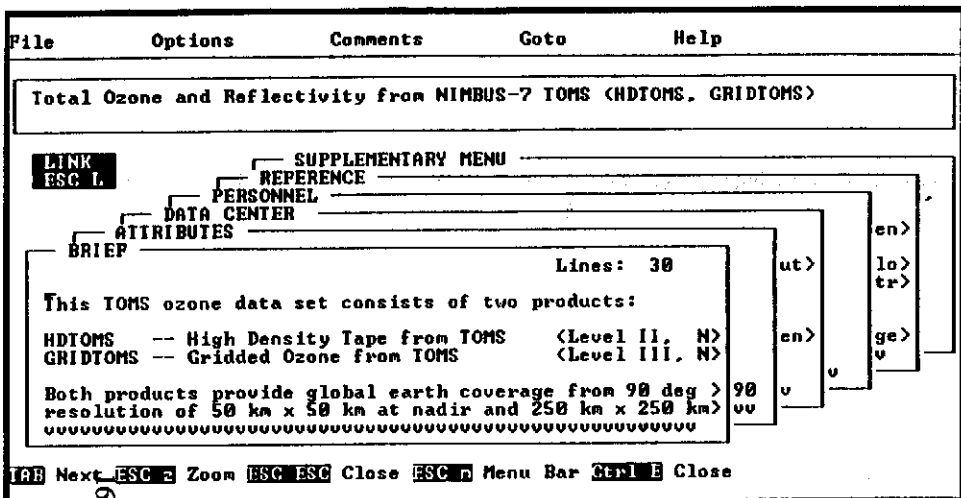
The Automatic LINK
The MD possesses the capability to LINK directly to other on-line systems. This immediate connection allows you to get more information on the directory entry, and in some cases browse, manipulate or order data of interest.

Future	
FLDS	Pilot Land Data System
FLDS-AMES	Pilot Land Data System at Ames Research Center
FLDS-CSRC	Pilot Land Data System at Coddard Space Flight Center
FLDS-JPL	Pilot Land Data System at Jet Propulsion Laboratory
PDS	Planetary Data System
PDS-ATMOS	Planetary Data System Atmospheres Node
PDS-CBO	Planetary Data System Conferences Node
PDS-PI	Planetary Data System Planetary Plasma Interactions Node
PDS-RINGS	Planetary Data System Rings Node
PDS-SBN	Planetary Data System Small Bodies Node
PNRA	National Program for Research in Antarctica
RESSTEC	Remote Sensing Technology Center of Japan
RCO	Royal Greenwich Observatory
SAL	Auroral Imaging File
SDAC	Solar Maximum Mission Data Analysis Center
SDCS	Synthetic Aperture Radar Data Catalog System
SDPAS	SWIR Data Display and Archival System
SEEDIN	Space Environment Realtime Data Intercommunication Network
SIMBAD	SIMBAD Database at CDS in France
SINRONIA	Global Observation Satellite Data Inventory Service
SAISS	Surveys Mapping and Remote Sensing Sector of Canada
SSEOP	Space Shuttle Earth Observation Program
STIS	Space Telescope Electronic Information System
STORM	Storm-scale Operations and Research Meteorology Program
SWRI	Southern Research Institute
TOMS	Total Ozone Mapping Spectrometer File
UA	University of Alaska Geodata Center
UARS	Upper Atmospheric Research Satellite Data Facility
UMI	University of Rhode Island AVHRR Archive
UTD	University of Texas at Dallas
VERI	Vector Electric Field Instrument Catalog
VILSPA	Villanova Satellite Tracking Station
VLA	NRAO Very Large Array Information System
WATSI	Wind and Temperature Spectrometer File
Dynamic Explorer Data Set Catalogs	

New Features of MD Version 2

Version 2 of the Master Directory has many enhancements which offer extended capabilities over Version 1. Pull-down menus provide access to automatic help and valids, session output, user information profile and session preferences, as well as a GOTO function. Titles resulting from a query may be sorted by one of several available attributes. The information display is presented on six "index cards" each containing a section of the directory entry. These cards may be zoomed to full screen display for easier viewing.

An MD Information Display Window



Catalog Interoperability

The Master Directory was created as part of the overall Catalog Interoperability Project whose objective is to enable researchers to quickly and efficiently identify, obtain information about, and access data. LINKs from the MD interface represent the first major step in interoperability.

Access Procedures

Using NSI/DECnet:

Set Host NSSDC
Username: NODIS
Select Option #1 from the menu
Answer "YES" to use Version 2

Using Internet:

Telnet 128.183.36.23
Username: NODIS
Select Option #1 from the menu
Answer "YES" to use Version 2

Via Direct Dial:

Set modem to 8 bits, no parity, 1 stop bit
For 300, 1200 or 2400 baud, dial 301-286-9000
For 9600 baud, dial 301-286-4000 or
FTS 888-4000
Prompt: Enter Number, you enter "MD"
Prompt: Call Complete, you enter <CR>
Username: NODIS
Answer "YES" to use Version 2

Where to get Help for the MD

Contact the Master Directory User Support Office (MDUSO) for help on any aspect of the MD. The MDUSO will direct you to the appropriate MD team member who can assist you with questions about using the MD, about the specific contents of the MD, with preparing new data entries, and with establishing LINK connections. The MDUSO can be reached by:

NSI/DECnet: NCF::MDUSO
Internet: MDUSO@nssdca.gsfc.nasa.gov
Phone: 301-513-1687 FAX: 301-513-1608

MD User Support Office:
Master Directory at NSSDC
c/o Hughes STX
7601 Ora Glen Drive, Suite 300
Greenbelt, Maryland 20771
USA

Master Directory

*A Scientific Data
Information Service*



QUICK REFERENCE GUIDE
Version 2 Alphanumeric Interface

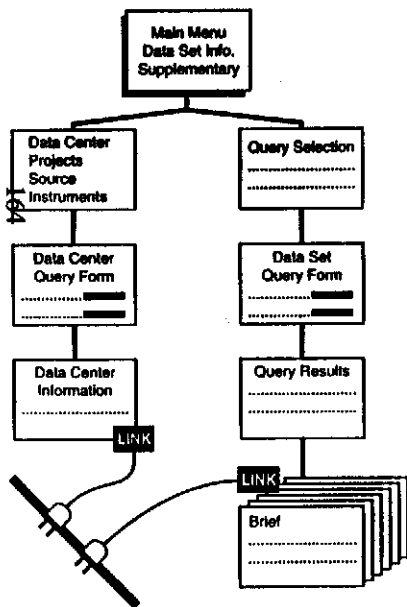


Operated at NSSDC
National Space Science Data Center
NASA Goddard Space Flight Center
Greenbelt, MD 20771 USA

What is the Master Directory?

The Master Directory (MD) is a free, on-line multidisciplinary database of information about Earth and space science data holdings of potential interest to the scientific research community. The MD contains high level descriptions of data set holdings of various agencies and institutions. It also contains supplementary descriptions about these data centers, as well as scientific campaigns and projects, sources (spacecraft, platforms), and sensors (instruments).

The MD also provides automatic connections, called LINKs, to many of the data centers or data systems holding the data sets. These connections are available from a simple command which may be activated from several places in the Directory.



Logging In

Enter your first and last names in the fields in the Log In window. If you are a new user, a window will open where you can enter user information; after you complete this, a User Tutorial window will open which explains how to use the Master Directory. If you are a repeat user, a window will open which asks you to confirm your identity. If you do not wish to leave your name, log in as "Guest".

How to Navigate in the MD

There are five basic tools used to navigate through an MD session.

Arrow keys move the cursor in the direction of the arrow and are used to:

- Move between items within a window
- Scroll through text in a window
- Move to a Command Button
- Move between Command Buttons
- Move off of a Command Button
- Move between Menu Bar functions
- Move between choices in a Menu Bar menu

The tab key is used to:

- Move to a Command Button
- Move between Command Buttons
- Move between Information Display windows

We suggest that the tab key be used to move to a Command Button as a general rule, even though the arrow keys may also be used. Use ESC tab to move backward between Information Display windows.

The carriage return <CR> is used to:

- Make a selection
- Undo a selection
- Enter a value
- Activate a command

Note: Menu or item choices available for selection are highlighted. A selected item will be preceded by an X. To undo a selection, move to the item selected and press the <CR> again. The item is now de-selected.

Command Buttons are located within a window and are used to access a function. Move the cursor to the Command Button and press the <CR> to activate its function.

MENU BAR: activates the Menu Bar

OK: acknowledges input as correct and activates subsequent functions

CANCEL: closes the current window and returns to the previous window; does not erase any inputs made to a value entry field

CLEAR: erases all inputs made to a value entry field and returns the cursor to the first value entry field; de-selects all choices made in a data set search selection

SEARCH: initiates the directory search

ADD CRITERIA: returns to the Data Set Search Key Selection window where you may select additional items by which to search

CLOSE: closes the current window and returns to the previous window

ESCAPE key sequences (Hot Keys) are combinations of the ESCAPE (ESC) key and another key used in sequence to activate frequently used functions in the MD. However, some Hot Keys are only available in certain windows. When using an ESC key sequence, first press the ESC key and then press the second key. Do not hold the ESC key down when using an ESC key sequence! Listed below are some frequently used Hot Keys:

- ESC c clears the input in the current field
- ESC C clears input fields on the current form
- ESC d page down
- ESC ESC closes current windows and returns to the previous window; in some specialized windows, moves to the CLOSE Command Button
- ESC h opens the HELP function menu located on the Menu Bar
- ESC L LINKs or connects to a remote data system
- ESC m toggles between the Menu Bar and the current window
- ESC q opens the QUIT MD window
- ESC r clears and redisplay the screen free of system noise or messages
- ESC s activates the directory search
- ESC u page up
- ESC v opens the valid values window
- ESC w opens a window containing help about the current window
- ESC z enlarges the current Information Display window to full screen size; when used a second time, returns the window to normal size

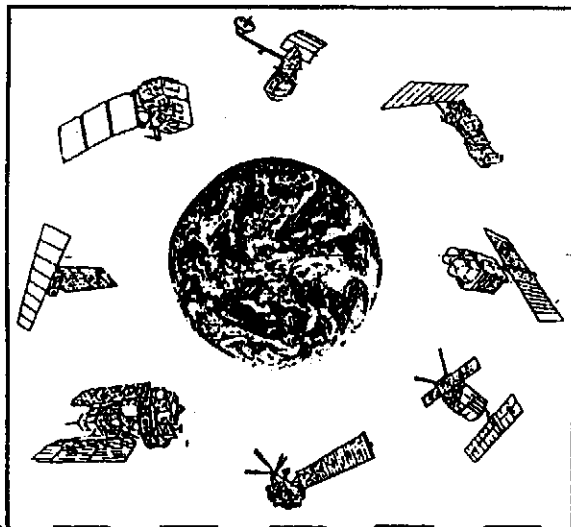
Use ESC hor ESC w to access help on using the MD at any time during your MD session. The User Tutorial window, accessed from the Menu Bar FILE function menu, gives you an overview of how to use the Master Directory.

4. BULLETIN BOARD SYSTEM (BBS)

The bulletin board system provides a place for communication among EOC staff and earth observation data users. The BBS contains five bulletin boards as shown below:

- IDN boards
 - New directory board
 - Update directory board
 - LINK board
 - System board
- Comment boards
 - IDN comment board
 - BBS comment board
 - SINFONIA comment board
 - Other comment board
- Notification boards
 - NASDA board
 - Event board
 - Free discussion board
 - System board
- Mission team boards
 - MOS board
 - ERS board
 - TRMM board
 - ADEOS board
- SINFONIA Boards
 - LANDSAT operation schedule board
 - MOS - 1 operation schedule board
 - System Operator board

165



5. HOW TO ACCESS

Earth scientists or application users can use EOC - IS through the following ways:

NSI/DECnet
\$ SET HOST 41950
USERNAME: NASDADIR

INTERNET
\$ Telnet 133.56.72.1
USERNAME: NASDADIR

Direct Dial
Set modem to 8 bit, no parity, 1 stop bit, X on/off, 1200 - 9600bps. (DEC Kanji)

TEL: 81 - 492 - 96 - 6400
USERNAME: NASDADIR

Packet Switching Digital Network
DTE Number 44014437216
USERNAME: NASDADIR

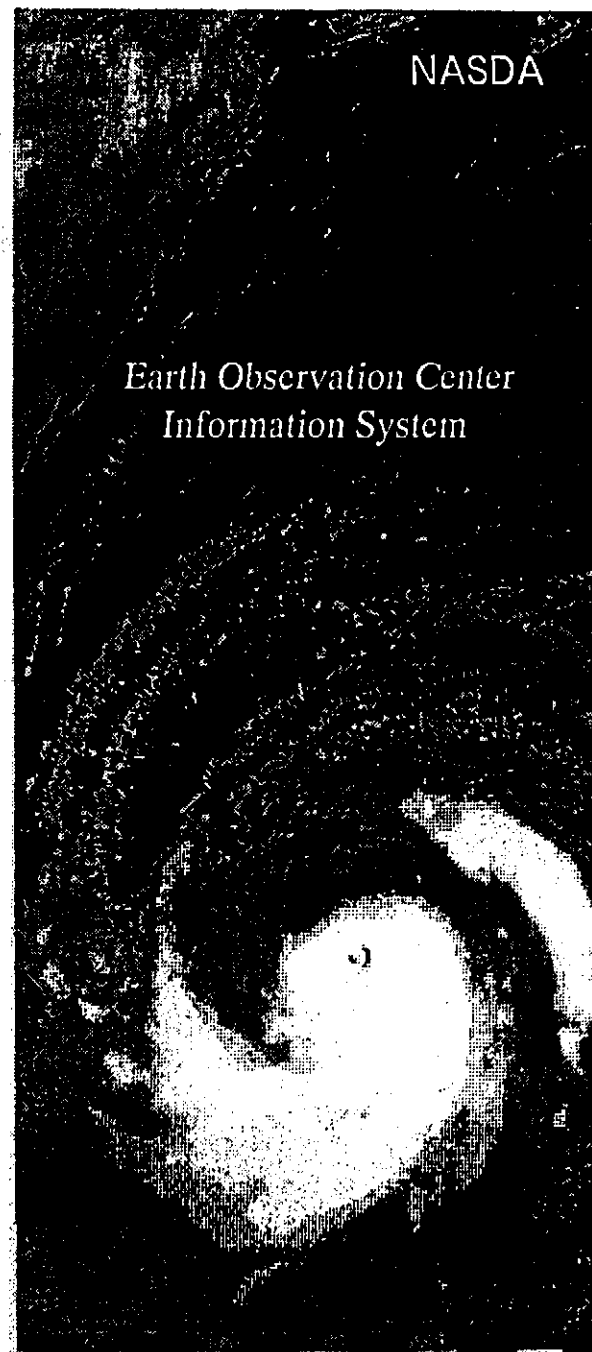
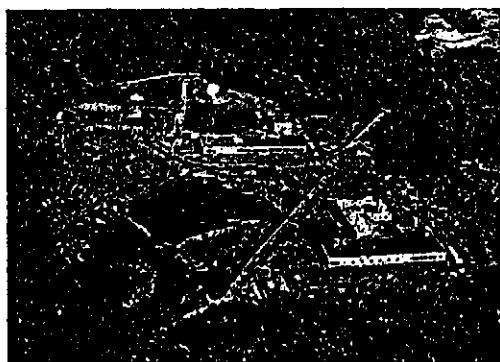
6. USER SUPPORT NUMBER

PHONE: 81 - 492 - 96 - 1611

FAX: 81 - 492 - 98 - 1001

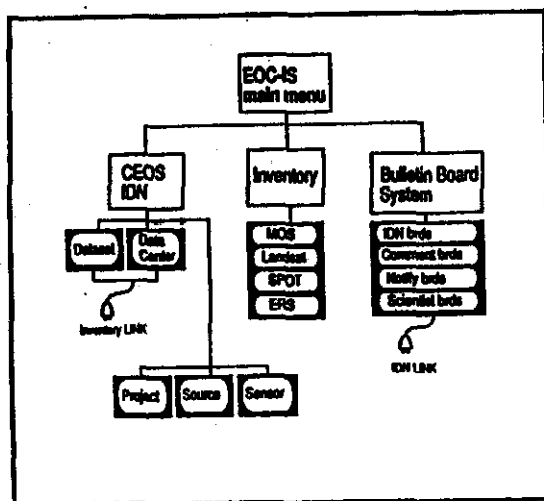
E - MAIL
NSI/DECnet: 41950 :: NASDAPROD
INTERNET: nasdaprod@nsaeoc.eoc.nasda.go.jp

MAIL: EOC/NASDA
1401 NUMANOUE, OHASHI,
HATOYAMA - MACHI, HIKI - GUN,
SAITAMA, 350 - 03 JAPAN



1.EOC - Information System (EOC - IS)

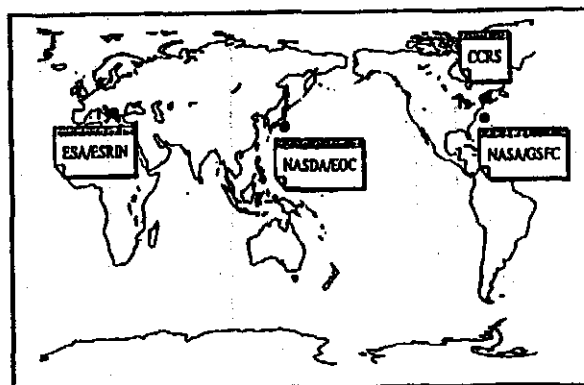
EOC-IS is an on-line information system developed by National Space Development Agency of Japan (NASDA) which provides an aid for earth scientists etc. to study various topics in the earth science including changes in the global environment. EOC-IS contains the earth science data directory system (CEOS IDN : Committee on Earth Observation Satellites International Directory Network), an inventory system and a bulletin board system (BBS). Through EOC-IS, earth scientists can retrieve valuable information on earth science data (especially location, contact point, data quality etc. of earth observation satellite data).



2.CEOS IDN

When earth scientists or application users want to know the data archive information that they need, it is sometimes difficult to determine. CEOS IDN is an International Directory Network promoted by the Committee on Earth Observation Satellites (CEOS). CEOS IDN major nodes exist at the NASDA Earth Observation Center, NASA/Goddard Space Flight Center and the ESA Earthnet Programme Office. IDN contains summarized earth science data information (directory information) and has the following features :

- Free of charge on-line data information service.
- No special training needed for use.
- Contains data center, project sensor and source descriptions.
- Automated connections (LINKS) to other information systems.

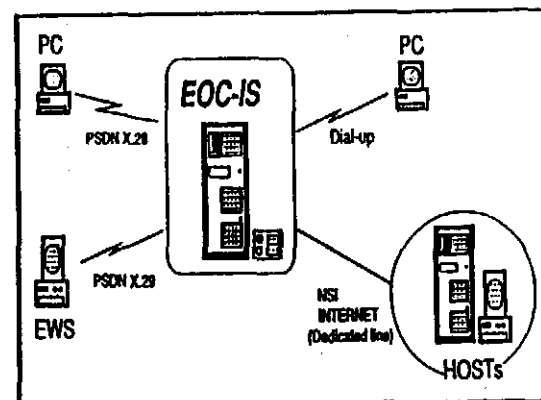


3.INVENTORY SYSTEM (SINFONIA)

When earth scientists or application users want to conduct their research or their activities using earth observation satellite data, they must consult with information such as temporal/spatial coverage, cloud cover, data quality etc. to determine the data they need.

EOC inventory system contains detailed information on earth observation satellites data archived at NASDA/EOC. Earth scientists or application users can retrieve them by using temporal, spatial and data quality keywords. The EOC inventory system has the following features :

- Full interactive system with no special training required
- Path/row, longitude/latitude or area name retrieving method
- Landsat 1,2,3 inventory information from all over the world
- Landsat, MOS, SPOT, ERS and JERS inventory of data received at EOC
- Near real time inventory information service



Information in ESA PID

ESA PID provides the following information:

Data Set Information

Searches in the directory - using selected keywords - provide descriptions of relevant data sets, their attributes, spatial and temporal coverage, data-providing centers, names of persons to contact there, and a related bibliography. More detailed information, as well as browse or data ordering services are accessible through automatic connections to information and catalogue systems using the "LINK" command.

167

Data Center Information

Additional, detailed information is provided about data centers, their activities and contact points, access procedures, ordering and price policies, and available distribution media.

Campaign and Project Information

Scientific data-collection campaigns/projects, and their associated data bases are described.

Source Information

Data source information (spaccraft, data acquisition system, ...) is provided including orbital characteristics, mode of operation, attached instrumentation and dates of operation.

Sensor Information

The sensors used to obtain data are described. This includes the dates of operation, wavelength, spatial and spectral resolution, and associated platform/spaccraft.

Access to ESA PID

Via SPAN

\$ set host EPOCAT (or: \$ set host 29628)
Username: ESAPID (no password required).

Via Internet

The address is 192.106.252.160
Username: ESAPID (no password required).

Via PSPDN

The Network User Address of ESA PID for international calls is : 02222650014761. In the following countries users can call (at local charges) the numbers below.

Austria	2618108061
Belgium	221044361
Canada	8580145861
Denmark	30106384161
Finland	20407661
France	17500039461
Ireland	36059922261
Italy	2650014761
The Netherlands	129017661
Norway	11062761
Spain	21406232161
Sweden	710401661
United Kingdom	2192115661

Via Direct Dial

Set modem to 8 bits, no parity, 1 stop bit, 300, 1200 or 2400 baud. Dial: +39 6 9417335.

Via ESA-Quest

under File 21. Further details can be obtained from Quest Help Desk (Phone:+39 6 94180300).

Further Information and Assistance

ESA PID User Support Office
European Space Agency ESRIN
Via G.Galilei,
00044 Frascati, Italy
Phone: +39 6 94180367
Fax: +39 6 94180361
Email: SPAN > 29628::PIDHELP



Prototype International Directory

an Earth science
information service



European Coordinating Node
of the International Directory
Network established by

Committee on Earth Observations Satellites

What is ESA PID?

The European Space Agency's Prototype International Directory (ESA PID) is a free, on-line information system to serve the European Earth science community which has research and operational data needs and requires access to data management systems.

ESA PID is a Directory

which contains more than 1200 Earth and space science data set descriptions, and summary information about Earth and space observation data centers, on-line systems, scientific campaigns, instruments and platforms.

ESA PID is International

It is the 'European Coordinating Node' of a world-wide network of directories, being sponsored by the Committee on Earth Observation Satellites, i.e.

- identical copies of the directory are maintained at NASA (USA) and NASDA (Japan);
- directory entries are reviewed by a panel of international reviewers;
- on-line access is provided via international networks and via European national access points;
- interfaces/links are provided to many on-line catalogues and archives.

ESA PID is a Prototype

for the study and development of future standards of directory functionality and catalogue interoperability, and for the collection of user requirements and feedback.

The Automatic "LINK"

ESA PID has the capability to "LINK" directly to various on-line systems. This immediate connection allows access to more information on a particular directory entry, and in some cases, to browse or order data of interest.

The following is the list of links which can be established via ESA PID with no telecommunication charges for the user.

Europe

EPOCAT	On-line ESA Earthnet data availability catalogue (formerly LEDA)
ESIS	European Space Information System
EUMETSAT	European Meteorological Satellite Organisation
QUEST	European Space Agency Information Retrieval Service

North America

ALI	University of Alaska Fairbanks/Geodata Center
ARIN	Aerospace Research Information Network
BBSO	Big Bear Solar Observatory
BRUNET	BRUNET Request Catalog
CCRS	Canadian Centre for Remote Sensing
DEWATS1	GSFC Dynamics Explorer Data, NASA
EICS	Energetic Ion Composition Spectrometer file
GLIS	Global Land Information System
GPLDS	Goddard Pilot Land Data System
IUE	International Ultraviolet Explorer Processing Facilities
LPI	Lunar and Planetary Institute
NCAROCEAN	National Center Atmospheric Research data availability catalog at OCEAN
NCDS	NASA Climate Data System
NODC	National Oceanographic Data Center
NODS	NASA Ocean Data System
NSSDC	National Space Science Data Center
OCEANIC	Ocean Network Information Center
OMNIFILE	Interplanetary Medium Database management System (OMNI) file
PDS	Planetary Data System
SAI	Dynamic Explorer Auroral Imaging Database
SEL	NOAA Space Environment Laboratory
SSEOP	Flight Science Support Office Photographic Database
SWRI	Southwest Research Institute
TOMS	Total Ozone Mapping Spectrometer Data System
UARS	Upper Atmosphere Research Satellite Central Data Handling Facility
URI	University of Rhode Island/AVHRR Inventory
UTD	University of Texas Dallas

Japan

JAPAN	World Data Center- C2
SERDIN	Space Environment Realtime Data Intercommunication Network

The National Geophysical Data Center

David Divins

The National Geophysical Data Center (NGDC) is composed of four divisions (solar-terrestrial physics, solid earth geophysics, information services, and marine geology and geophysics) and a paleoclimatology group. Also located at NGDC are the World Data Center A for each of these before mentioned disciplines.

The Marine Geology and Geophysics Division (MGG) has three primary data types, marine geology, marine geophysics, and bathymetry. Marine geology data bases include sediment thickness, grain size, geotechnical properties, physical properties of sediments, petrology, core and sediment sample descriptions, and well logs. Marine geophysics is one large data base with over 37 million records of underway geo[physical information. Contained in this data base are digital navigation, bathymetry, magnetics, and gravity. Bathymetric data are in the several data bases. The National Ocean Service (NOS) hydrographic data base has over 40 million soundings from the US EEZ. NGDC also maintains a multibeam/swath data bathymetric data base.

MGG is the national archive for all marine geology and geophysical data collected with US funds. As a result institutions are required to send copies of their data to NGDC. When a data set arrives at MGG it first is entered into a center-wide data ingest program. This program tracks the data set and keeps track of what institutions are send data. Once in the system the data set is run through quality control programs to check for consistency and that data values fall within reasonable and acceptable levels. If errors are detected the contributor of the data is contacted and corrections are made if not too labor intensive. If a large number of errors are detected the data set is returned to the contributor, with a list of known errors for the contributor to correct. The clean data are then entered into the appropriate inventory. In the case of marine geology data, the collecting institution, ship, cruise id, sample, id, date, water depth, position, sampling device, and what analysis were performed. No data are stored in the inventory. The inventory is then used to perform searches based on any one or combination of the above parameters. The results of this search point to the appropriate digital files, paper reports, or microfilm.

Distribution of the data to the user community is in three primary modes. The first, and perhaps the most popular, is CD-ROM. An entire data base is contained on a CD-ROM, along with inventory information and, most importantly, search and retrieval software. MGG now has the capability to provide a CD-ROM which is accessible, and usable on a PC, Macintosh, or UNIX platform/operating system. This development is extremely desirable since it allows MGG to make all of its digital data available to a majority of the user community. A second method of data distribution is on line. Currently this is limited to anonymous ftp transfers. The third method is custom searches and data products, for example custom color images.

NGDC is currently headed in the on-line direction, that is, NGDC will soon have a bulletin board system populated with meta data and inventory information. The goal of this system will be a bulletin board populated with fully searchable data inventories and the actual data.

processing, with replacements being made; the turn-around time must be short, since the data may be used for planning; data use must be coordinated to avoid any problems concerning ownership and publication rights.

Access to what?

The system needs to provide access to data at various levels, from raw, telemetered information to gridded, analyzed fields. It must include both historical data sets and new observations fresh from the ships. In addition to oceanographic data, we could see needs for information on river run-off, rainfall, etc. The data sets fell in four classes:

- Historical data
- Operational data
- RMRP-generated data
- Non-RMRP generated data [i.e. GLOBEC]

To get a sense of the range of information, we discussed some of the data sets we foresee including. These can be summarized as follows:

a) RMRP

Historical

nutrients, T, S, phytoplankton, contaminated sediments, streamflow, meteorology, wave climate

New

hydrography, nutrients, biology (net data & other forms), drifters, moorings (PI & agencies), AVHRR [processed], seawifs

Real time

AVHRR [2x daily], buoys

Models

analyzed fields, simulations, data assimilation runs

Bibliographic

b) Outside RMRP

Operational

Non-RMRP research data

GLOBEC

NODC

USGS streamflow / hydrology

NGDC

NCDC

MEDS

...

In addition to RMRP investigators' datasets, links to other groups - operational data streams, non-RMRP research programs, and large archives - will be important.

In the course of this discussion, it became clear that it is important in some cases to include multiple versions of the same data. The AVHRR temperature field in its real-time form is useful for planning and for a quick impression of the situation at the time of other observations. On the other hand the more carefully analyzed and corrected SST products are important for modeling and final analyses of the environmental state. We need access to data at various stages of processing, with quality changing as the information becomes more refined. To be useful, these multiple data sets must be sufficiently documented that the user can understand the relationship between them and decide which to use. The raw data may not be on-line. There is a clear role for large archives of such information, perhaps on tape or optical media, so that the numbers can be reprocessed if someone desires. At the same time, the volume precludes storing

the information in an on-line archive; rather, the data system should indicate that such information exists and where it can be obtained.

The system must provide mechanisms for indicating data quality; these include data flags, a "pedigree" indicating the processing steps, and a bulletin board for feedback from other users. Some data sets may have quite detailed quality information, while it may be sketchy in other cases - the system should encourage including such information, but not require it.

From this, we see the need for the program to establish clear responsibility for archives, for maintenance of data sets, and for updating. An agreed-to and enforced data policy will be necessary, both to encourage early submission of data for others in the program and to protect the providers (including appropriate procedures for referencing data sources in papers). It should lay out the obligation to contribute data, time frames for doing so, and limitations on access. While we want data in the system prior to publication, the PI's and users must understand fully their responsibilities and rights. The system should provide various levels of access - within the program, general scientific community, other users.

These requirements indicate some of the role which must be played by a centralized part of the data management effort. Directories / catalogs / inventories must be maintained and must reflect the status of various versions of the data set, expected times when new versions will become available, and commentary on the changes.

When?

We discussed briefly the time scales for development of the IMS. With some observational programs underway, there is clearly a need for immediate action. Over a several year time scale, linkage with other programs such as GLOBEC will become important; thus, the design should factor in ways in which interchange of information with other programs can be best accomplished. The TOS effort will also have a prototype during that time.

With What?

Distributed data systems seem well-suited to the task, since we want to encourage exchange of preliminary data and information which the PI's are still refining and modifying. The only way to accomplish this is to have the data system providing data from the PI's own machine. Internet is the only widely-based network with adequate bandwidth that can provide the "highway" for the desired information. Many researchers are already connected to the network. On the other hand, it will also be necessary to provide points through which non-networked systems can gain access to the information, presumably by dialup.

Attributes and Technology Summary (Thurs.)

From the discussion of our group and the other working group, we can summarize the attributes of the IMS as :

Attributes / functions

- Distributed
 - network based/non-network access
 - Role for central data management effort
 - directory, standards, data policy
 - Accessible
 - ease of use, access control/rights, multiple platforms
 - Multiple data types

historical	profile	metadata
new obs.	time series	data
real time	images	
models	2 and 3D grids	
bibliographic	text	
- Full of data
 - easy to incorporate existing data systems
 - easy to add new information
- Quality information
 - documentation, flags, user feedback,
 - referenceable
- Multiple versions
 - documented
- Links to other systems
- Archive
 - long term stability, on-line and off-line archives
- Analysis procedures
 - extensible

From the computer science view, the IMS can be described as follows:

- Architecture: distributed (db & users)
 - node (directories, tools?)
- Access link: network
 - pt. of access
 - link to other IMS
- Client-server

While the objects managed by the IMS have these characteristics:

- Type
- Characteristics for incorporation/ modification
- Quality
- Archival status
- Suitable for analysis
- Level of refinement

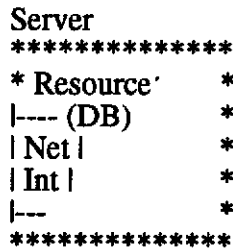
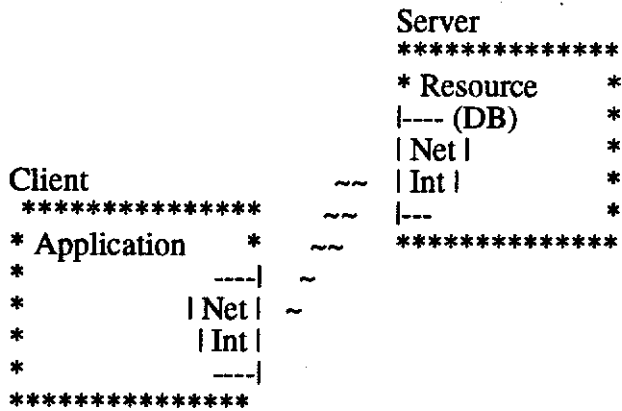
The tools envisioned are user-oriented and include filters, common analysis and display routines (e.g. browse), and individual analysis and display tools (outside the domain of the IMS).

Attributes

- Client-server
- Comprehensive
- Long life

Technology

- Internet



Technology

- Internet: evolving medium of choice
- CD-ROM: an important distribution mechanism
- UNIX: Evolving OS of choice [data servers]
- X: Graphics standard

Discussion: (Fri.)

We discussed briefly the operating system characteristics. The servers must be multitasking to satisfy requests without serious impact on the other uses of the machine. For servers, then, UNIX seems to be the current best choice. Clients, on the other hand, could operate on other systems, as long as they have appropriate Internet or phone access.

People:

We considered the GLOBEC plans for a data management office. Many of the roles they have envisioned for a data manager and a programmer seem also applicable to the RMRP program. For the data manager, we envision the responsibility to:

- Track data: submission - archive
- Provide information on submission procedures
- Accept and verify "publication" of data to rest of program
- Check conventions for naming, units, ...
- Work with PI's on QA
- Ensure documentation
- Own and manage directory
- Disseminate info (Bulletin board)
- Facilitate archival (timely / lawful)
- Facilitate linkage
- Publicize new objects / info exchange

It is possible that these functions could be handled by some of the existing facilities, though it is important that the oversight remain with the groups described below. Programmer(s) would work at various sites to:

- Aid in the decision whether to make or buy software for particular functions detailed below
- Interfaces to data objects (raw, model, derived)
 - ORACLE
 - AFAP
- Programs to transform/ process data
- Network access
 - PC, MAC
- System interfaces: other IMS's
- Data import to established programs
- Software for
 - Data providers
 - Data users

Whenever possible, existing software should be adapted to fit the needs. Again, the virtues of collaboration with other efforts seem obvious here. In addition, the system must have the flexibility that programmers can extend the individual data objects, the filters and transformation routines, and the interfaces. Scientists are often not interested in extensive database queries (and the updating problem so important in business is not a major difficulty, since science data sets are rarely updated but often referenced) - the important problem in getting the information into some program, whether commercial or user-written, where it can be examined and analyzed.

Oversight:

As the RMRP-IMS is developed and put into use, and as the community of both providers and users grow, it will be important to maintain oversight to ensure that the data manager efforts are serving the researchers and that the data is flowing up the triangle properly. We discussed the concerned parties which might make up an Oversight Board for the GOM-IMS:

- Members of the A., B., C.
- PI's (science, IMS)
- RMR board (EPA/ NOAA)
- GLOBEC Georges Bank SC
- Sea Grant
- AFAP
- CME
- RARGOM
- Users/ funders

In addition, a Technical Steering Committee should be established immediately to recommend a direction for the IMS and to guide the funding agencies. In the short run, we would see this as some combination of PI's, IMS experts, and members of the Board(?). In the longer run, presumably, the data manager and programmers would become part of the technical committee.

Funding:

While we did not spend a great deal of time on funding, we did suggest that much of the funding should go on getting the right people in place to ensure that the IMS system is developed in a timely way and will meet the needs. We also thought that there could be considerable sharing of funds among the various programs in the region. Some possibilities for leveraging may exist, as well.

Draft Recommendations --- RMRP group

- 1) The agencies and scientists interested in the Gulf of Maine should commit themselves to constructing and maintaining a comprehensive regional information management system. The system needs in its final form to meet the needs of the whole range of users, from the researchers to the managers. It should facilitate transfer of information among and between groups. [5 years]
- 2) A distributed system with a client--server architecture is most likely to meet the requirements for data management. The system should be based on the Internet, but have good access for people not on the network. Data servers will need to be multitasking machines (and therefore will initially be running UNIX), but clients may very well run on other systems. [Immediate]
- 3) It must be simple to add existing data systems to the GOM IMS, and links to AFAP and the developing GLOBEC system will be important because of the connections between the data sets. [Immediate]
- 4) The system should merge smoothly with the distributed data system being developed under the aegis of TOS. That system is expected to satisfy many of the technical requirements placed on the GOM system, and, in addition, will be connected closely to national oceanographic archives. [2 years]
- 5) The system should be compatible with, and ultimately merge with, operational systems used by resource management, monitoring, and enforcement agencies in the region. [5 years]
- 6) We need to form a Technical Committee immediately. The committee needs to investigate the possible ways of satisfying the requirements and what needs to be done in addition. When possible, existing efforts should be capitalized upon. Funds should be expended on value--added efforts if that is possible. The technical committee should track the system as it evolves and additional people added or subtracted as required. [Immediate]
- 7) A broader GOM--IMS Oversight Group should be formed to ensure that the technical committee is aware of the needs of various groups and that plans are developed to meet those requirements in a timely fashion. [6 months]

**Gulf of Maine IMS
Report on Working Group 2
Gulf of Maine Data and Information Management Workshop
November 3-5, 1994**

by
Paul Boudreau
Department of Fisheries and Oceans, Canada
P.O. Box 1006
Dartmouth, N.S. B2Y 4A2
Internet: Boudreau@biome.bio.ns.ca
and
John Sowles
Maine Department of Environmental Protection
State House Station 17
Augusta, ME 04333-0017
Internet: Sowles@biome.bio.ns.ca

The Charge: The working group on an Information Management System (IMS) for the Gulf of Maine was asked to identify attributes of an IMS that provides broad access to and utility of data and information relevant to a diverse Gulf of Maine community. The community includes researchers, resource managers, educators, non-government agencies and the public. This working group would make specific recommendations on the establishment of an IMS that facilitates the linkage of science and management and would complement the issues being addressed by the more focused Working Group 1 that was considering the specific needs of the research community.

The Composition: The makeup of the group was in keeping with the charge. It included a diverse representation of 30 individuals from Federal and State government agencies, Universities, other Regional Marine Research Boards and private sector organizations. This eclectic collection of interests provided a broad spectrum of experience and knowledge in data and information management. As an example of the range of experience, some members of the group were very experienced in the use of Internet and were involved in the development of new and powerful tools for handling distributed data base systems. Other members of the group were being exposed to the power of Internet for the first time. This made for some very interesting and challenging discussions.

Day 1: The deliberations of the group began with a round table introduction and brainstorming session. In addition to presenting themselves to the group, everyone was asked to comment on their view of a IMS; with respect to the kind of things that would characterize an IMS and what they would do with such a System. The intent of this round table was several fold: 1) to develop a working relation among the group by ensuring that everyone felt comfortable with saying at least a few sentences, 2) to develop some common definitions and understanding regarding our charge and 3) to begin defining an IMS that would later become the recommendations to the larger group. The list of items to come out of the round table included both qualities of the IMS such as user friendliness, and ease of access and functions such as archival storage and direct access to raw data (Table 1).

Table 1.
Attributes of a Gulf of Maine Data and Information System
Identified by Working Group #2

Qualities	Built-in Functionalities
Responsive	Analytical Capacity
User Friendly	Storage/Access Aggregation
Rugged	Acronym Dictionary
Robust	Student Access
Encourages Standardization	Multiple Format/Platform Access
Promotes/Facilitates Publication	Data Referenceability
High Quality	Library
Handles Multiple Data Types	Traceability/Trackability
Easy to Enter Information/Data	Ease of Information/Data Entry
Affordable	Dynamic Directory/Inventory
Sustainable	Communications
Efficient	User-Supplier Feedback Loops
Expandable Capacity	QA Documentation
Accessible	Access
Non-intrusive	Standards
Allows for Documentation	Transparent Network Access
Promotes Information Sharing	Handles Metadata
Provides Transparent Network Access	Preservation of Institutional Memory
Bifocalism (?)	Domain Analysis (?)

In addition to the above attributes, it was emphasized that for any system to succeed it must be administratively supported and maintained.

During the discussions, it became clear that disagreements around semantics were a major stumbling block. One member of the group, not totally facetiously, even proposed that an acronym dictionary be developed. Other disparities involved the extent or scope of coverage by a term. For example, to some, *access* simply meant ability to have permission to use one another's data while to others *access* meant electronic connectivity to data and information through a user friendly system that actually processed and analyzed the data immediately upon request. The word *standardization* evoked differing interpretations; from a single detailed sample collection, analysis, and reporting protocol (command and control approach) which enabled transparent comparability to all other data, to a minimal burden level of simply adhering to existing protocol(s) when appropriate and possible. Interestingly, it appeared that in general, meaning to a user of data and information differed from that of a supplier.

Consensus was nevertheless achieved the first day on three topics:

1. The system should be a **distributed data and information system**. From the morning presentations and our discussions it was obvious that the amount and diversity of data and information made it prohibitive for any single organization to consider amassing all data and information into a single location. There was also concern for the need to exercise some degree of control over some types of data and that a distributed database facilitated data referenceability.

2. The system should **take advantage of existing systems**. Again, the morning plenary sessions convincingly presented the existing state of information management in the Gulf of Maine and beyond, as having a number of viable pieces already developed or under development, such that it would not be reasonable or responsible to throw out the existing work.
3. The most immediate need for and use of an IMS was for an **electronic inventory**. Such an inventory would be widely distributed in a variety of formats and would evolve into an on-line product which could eventually provide connectivity to some of the actual databases.

Day 2: The second day began with the presentation of a "development plan" (Table 2.) that described the evolution of a Gulf of Maine IMS, from the present status quo through to the establishment of an electronic database to a system with real time connectivity.

Table 2.
Continuum of IMS Functions and Requirements
Ranging from Current Situation to Fully Staffed Data Center
 (from left to right, features are cumulative and increase in development time and cost)

	Status Quo	→	→	→	→	Long Term Goal
	Paper Lists & Internet	Unified Paper Lists & Internet	Electronic Directory	Electronic Directory with Connectivity	Electronic Directory with Resolution of Differences	Data Center with Standardized Centralized Archives
Administration	Steering committee		System Support	System Development		Steering Committee with Staff
Responsibility	Marketing Coordination	Compile Lists Updates	Maintenance	Connectivity with User Support	Libraries	Data Entry, Aggregation, Interpretation, User Support, Archival
Documentation	Descriptive, QA/QC, Data Documentation	Metafile Structure		Referencibility Pedigree		Results & Processes
Standards	Documentation, Geo-Referencing, Temporal Specification		Standard Directory Structure	Directory std's, Communications, Directories or Pointers	Concordance (mapping)	Standard Codes, Field Names, Structures
Policies & Enforcement				Data Policy		Adaptability, with Ability to Evolve
Data Types	Current Research			Historical		Operational, Real Time
Features		Bulletin Board		Feedback Loops		Std. Codes, Field Names, CME Policy & Mgt., RMRP (Research)
Access			Access	Raw Data		Aggregation

Unfortunately, as a result of the diverse experience of the working group to review existing technology and decide on the system to be used for the electronic data base and those who felt that such a decision was already indicated by an existing system, the discussion carried on until coffee break.

Following coffee and a number of small informal discussions, the group reconvened with better common understanding and a more relaxed mood. The remaining time proved very fruitful in arriving at consensus. Of note, the following were agreed to by the group:

1. **A working group should be set up** to promote the implementation of an IMS. This group should be small, possibly 6 people, representing the Council, the Board and the Association. This group would have the mandate to recruit the necessary technical expertise in the Region as it sees fit. The first task of this group would be implement the electronic inventory. The group should meet as soon as possible, no later than a couple of months from the time of this meeting.
2. **A startup group should be set up** to report to the meetings of the Council, Board and Association in the coming two months. These reports would ensure that the organizations are well appraised of the recommendations of this workshop and solicit commitment for establishment of a GOM IMS that would jointly serve the goals and mandates of all three organizations.
3. Attendees of this workshop and others in the Region should **become familiar with Internet** and the role it might play in an IMS. This would start with the establishment of electronic communications among the group. Such efforts would provide an educational benefit in the region to the potential of an IMS. This would help to maintain support for the initiative from all agencies involved. Internet addresses of all attendees should be distributed.
4. Attempts should be made to **connect with a few select high priority data bases**. Such a list might include the following databases:

NOAA Master Directory	NOAA Meteorological Data	State/Provincial Coastlines
Redtide Database	Bathymetry	CME-Gulfwatch
Shellfish Sanitation Surveys	Swimming Beach Sanitation Surveys	Bottom Type
Circulation	USGS Coastline	

- 5.) There was interest in convening another workshop similar to this as a good means of promoting and maintaining communications and awareness of information management in and around the Gulf of Maine.

Recommendations of Working Group #2

Ultimately, Working Group #2 endorsed the concept of and need for a Gulf of Maine Information Management system. The group reached consensus on a series of specific recommendations designed to make scientific information accessible to both scientists and resource managers in a timely and meaningful manner. Working Group #2, therefore recommends to the Gulf of Maine Steering Committee that the following set of principles be adopted and actions be initiated:

Principles to Adopt

- 1.) **Refine a long-range vision statement** toward which we will deliberately proceed Working Group #2 settled on the establishment of a Gulf of Maine Information Management System that enhances our ability to answer scientific and management questions by facilitating the integration and synthesis of complex and diverse types of data and information.
- 2.) **Develop a system of distributed databases** that recognizes the diversity of existing specialized database types.
- 3.) **Take advantage of and if necessary expand existing systems.** Where an existing system was found to be deficient, the group supported sponsoring its modification and/or expansion.
- 4.) **Adopt Internet** as the communications network of choice.

Actions to Initiate

- 1.) **Form a "start-up" committee** that presents the findings and recommendations of the overall workshop to the RMRB, CME, and RARGOM. The "start-up" committee should consolidate opinions expressed from each sub-workshop and recommend an overall implementation workplan inclusive of schedule, costs and responsibilities. This workplan will be completed by December 1993.
- 2.) **Form an ongoing GOM-IMS Working Group** to develop and present a long range workplan which will move the Gulf of Maine community toward the larger goal of a gulf-wide information management system.
- 3.) **Develop Request for Proposals to establish an electronic Gulf of Maine Database Directory.**
- 4.) **Immediately exercise Internet** by
 - ...publishing a list of GOM Internet addresses in these proceedings.
 - ...Set up a GOM Bulletin Board
 - ...distribute the DIMC Directory
 - ...identify and develop connectivity to high priority databases.
- 5.) **Host another workshop** annually

Appendix A: Workshop Agenda

Gulf of Maine Data and Information Systems Workshop
University of New Hampshire

AGENDA

Wednesday, 3 November 1993

- 4:00-6:00pm • Early Registration [New England Center (NEC) Hotel Lobby]
6:00-10:00pm • Poster Set-Up [NEC Berkshire Room]
- 6:30pm • Dinner for Steering Committee Members/Working Group Chairs and
Rapporteurs (SC/WGC&R) [NEC Dining Room]
- 7:30pm • Pre-Workshop SC/WGC&R Meeting [NEC Narragansett Room]

Thursday, 4 November 1993

- 7:30am • Continental Breakfast [NEC Gallery]
- 8:00-10:00am • Conference Registration [NEC Gallery]
- 8:15am PLENARY SESSION 1: BACKGROUND [Berkshire Room]
CHAIR - WENDELL BROWN
- Welcome (Berrien Moore III)
Introduction to Workshop Objectives and Activities (Don Phelps)
- The Gulf of Maine Regional Marine Research Program (RMRP): History, Goals and Status - *Bob Wall, Chair, Regional Marine Research Board*
 - Role of a Data and Information System in the Gulf of Maine (GOM) RMRP - *Don Phelps, Regional Marine Research Board Member*
 - A Context for Information Management Systems and User Needs: A User-Data Matrix - *Jeff Rosen, Database Developer*
 - The User Community for a Gulf of Maine Information Management System - *Maxine Schmidt, Data and Information Management Committee of the Council on the Marine Environment*

6:00pm • Buffet Dinner [NEC Dining Room]

7:30pm **PLENARY SESSION 4: WHERE DO WE STAND?** [Berkshire Room]
CHAIR - WENDELL BROWN

Working Group Chairs present summaries of discussions highlighting principal issues.

9:00pm • Post-Plenary SC/WGC&R Meeting

Friday, 5 November 1993

7:30am • Continental Breakfast [NEC Gallery]

8:00am • Reconvene Working Group Discussions

10:15am • Coffee Break

10:30am • Continuation of Working Group Discussions

12:30pm • Buffet Lunch Break [NEC Gallery]

Friday Afternoon

1:00pm **PLENARY SESSION 5: Summation** [Berkshire Room]
CHAIR - BOB WALL

Summation of Working Group Deliberations

- Working Group Recommendations
- Discussion of Next Steps

3:00pm • Workshop Adjourned

Friday, Late Afternoon

- 3:30pm
- Post-Workshop SC/WGC&R Meeting.
 - Assignment of workshop proceedings responsibilities.

Appendix B: Participants

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Dr. Ivar Babb
National Undersea Research Center
University of Connecticut at Avery Point
1084 Shennecossett Road
Groton, CT 06340
telephone 203-445-3451
fax 203-445-2969
e mail nurp.ucap (omnet)
working group #2:GoM

Dr. Ronald Baird
Worcester Polytechnic Institution
100 Institute Road
Worcester, MA 01609
telephone
fax
e mail
working group not indicated

Mr. James Balestrieri
SYSCON
10 John Clark Road
Middletown, RI 01840-5634
telephone
fax
e mail jamesbi@ids.net
working group #2:GoM

Ms. Ellie Baptiste
New England Marine Research Lab
Batelle-Duxbury
397 Washington Street
Duxbury, MA 02332
telephone
fax
e mail
working group #2:GoM

Dr. Gerry Barton
Global Change Research Info. Office
NMOA Suite 805
1825 K Street, N.W.
Washington, D.C. 20006
telephone
fax
e mail gbarton@mail.ciesin.org
working group #2:GoM

Dr. James Bisangi
NOAA
NMFS - Narragansett
28 Tarzwell Drive
Narragansett, RI 02882
telephone 401-782-3313
fax
e mail j.bisagni (Omnet)
working group #1:RMRP

Dr. Paul Boudreau
Dept. of Fisheries & Oceans
Biological Fisheries Branch
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia B2Y 4A2
CANADA
telephone 902-426-7564
fax 902-426-8484
e mail boudreau@biome.bio.ns.ca
working group #2:GoM

Ms. Eugenia Braasch
Dartmouth College
Thayer School of Engineering
Hanover, NH 03755
telephone 603-646-3480
fax 603-646-3856
e mail braasch@dartmouth.edu
working group #2:GoM

Dr. Wendell Brown
University of New Hampshire
OPAL
Morse Hall
Durham, NH 03824
telephone 603-862-3153
fax 603-862-0243
e mail wsb@panthr.unh.edu
working group #1:RMRP

Ms. Barbara Brown
EPA Region 1
Regional Scientist
JFK Federal Building
Boston, MA 02203
telephone
fax
e mail
working group #2:GoM

Dr. Marilyn Buchholtz ten Brink
U.S. Geological Survey
Quisett Campus
Woods Hole, MA 02543
telephone 508-457-2392
fax 508-457-2310
e mail mtenbrin@nobska.er.usgs.gov
working group #2:GoM

Mr. Mark Cameron
SYSCON
10 John Clark Road
Middletown, RI 01840
telephone
fax
e mail c/ojbalestrieni@meteor.syscon.hii.com
working group #2:GoM

Dr. Janet Campbell
University of New Hampshire
OPAL
Durham, NH 03824
telephone
fax
e mail
working group not indicated

Dr. Walton Campbell
NOAA Ocean Products Center
5200 Auth Road
Room 100
Camp Springs, MD 20746
telephone
fax
e mail
working group #2:GoM

Dr. Michael Collins
University of Maine
Dept. of Surveying and Engineering
Orono, ME 04469
telephone
fax
e mail collins@raven.umesve.maine.edu
working group #2:GoM

Dr. Peter Cornillon
University of Rhode Island
Graduate School of Oceanography
Narragansett, RI 02882
telephone
fax
e mail
working group #2:GoM

Dr. David Divins
NOAA - NGDC
University of Colorado
Boulder, CO 80302
telephone
fax
e mail dld@mail.ngdc.noaa.gov
working group #2:GoM

Mr. John Evans
MIT
Urban Studies and Planning
Room 9-514
Cambridge, MA 02139
telephone
fax
e mail jdevans@mit.edu
working group #2:GoM

Dr. Kenyon Fairey
University of Massachusetts
203 Godell
Amherst, MA 01003
telephone
fax
e mail kfairey@ucs.umass.edu
working group #2:GoM

Dr. Glenn Flierl
MIT
Center for Meteorology & Oceanography
Building 54
Cambridge, MA 02139
telephone 617-253-4692
fax 617-253-4464
e mail glenn@lake.mit.edu
working group #1:RMRP

Dr. Christopher Garside
Bigelow Laboratory for Ocean Sciences
McKown Point
W. Boothbay Harbor, ME 04575
telephone 207-633-9600
fax
e mail c.garside (Omnet)
working group #1:RMRP

Mr. Michael Gennert
Worcester Polytechnic Institute
Department of Computer Sciences
Worcester, MA 01609
telephone 508--831-5476
fax
e mail michaelg@cs.wpi.edu
working group #1:RMRP

Dr. Stephen Grilli
University of Rhode Island
Dept. of Ocean Engineering
Graduate School of Oceanography
Box 40
South Ferry Road
Narragansett, RI 02882
telephone 401-792-6636
fax
e mail grilli@mistral.oce.uri.edu
working group #1:RMRP

Mr. Nabil Hachem
Worcester Polytechnic Institute
Department of Computer Sciences
Worcester, MA 01609
telephone
fax
e mail
working group not indicated

Dr. James Irish
WHOI
Room 307
Smith Lab
Woods Hole, MA 02543
telephone
fax
e mail jirish@whoi.edu
j.irish (omnet)
working group #2:GoM

Dr. Dan Jacobs
University of Maryland
Sea Grant College Program
0112 Skinner Hall
College Park, MD 20742-7640
telephone
fax
e mail jacobs@mbimail.umd.edu
working group #2:GoM

Mr. David H. Keeley
Maine State Planning Office
Station 38
Augusta, ME 04333
telephone 207-287-3261
fax 207-287-5756
e mail d.keeley (omnet)
working group #2:GoM

Dr. Deidre Kimball
University of Rhode Island
Graduate School of Oceanography
Box 47
South Ferry Road
Narragansett, RI 02882
telephone
fax
e mail
working group #2:GoM

Dr. Christopher Kroot
Maine Dept. of Environmental Protection
State House Station #17
Augusta, ME 04333
telephone
fax
e mail
working group #2:GoM

Dr. Matthew Liebman
EPA Region 1
WQE
JFK Federal Building
Boston, MA 02203-2211
telephone
fax
e mail
working group #2:GoM

Dr. Daniel Lynch
Dartmouth College
Thayer School of Engineering
Hanover, NH 03755
telephone 603-646-2308
fax 603-646-3856
e mail dan_lynch@dartmouth.edu
working group #1:RMRP

Dr. Michael MacDougall
US EPA Information Management
1 Congress Street
Boston, MA 02203
telephone
fax
e mail
working group #2:GoM

Dr. Joseph Macknis
EPA Region 3
Chesapeake Bay Program
410 Severn Avenue, Suite 109
Annapolis, MD 21403
telephone 800-523-2281
fax 401-267-0282
e mail
working group #2:GoM

Dr. David G. Mountain
National Marine Fisheries Service
Northeast Fisheries Science Center
NOAA
Woods Hole Laboratory
166 Water Street
Woods Hole, MA 02543-1097
telephone 508-548-5123
fax 508-548-5124
e mail
working group #1:RMRP

Peggy Murray

telephone
fax
e mail pmurray@saicnpt.com
working group #2:GoM

Dr. Joan Palmer
National Marine Fisheries Service
Woods Hole Oceanographic Institute
Woods Hole, MA 02543
telephone
fax
e mail
working group #2:GoM

Dr. John Paul
EPA
Narragansett Marine Lab
27 Tarzwell Drive
Narragansett, RI 02882
telephone 401-782-
fax 401-782-3030
e mail paul.john@epamail.epa.gov
working group #2:GoM

Dr. Judith Pederson
Massachusetts Coastal Zone Management
Executive Office of Environmental Affairs
Leverett Saltonstall Building
Room 2006
100 Cambridge Street
Boston, MA 02202-0001
telephone 617-727-9530 ext.413
fax
e mail j.benoit (omnet)
working group #1:RMRP

Dr. Joseph Pelczarski
Massachusetts Coastal Zone Management
Executive Office of Environmental Affairs
Leverett Saltonstall Building
100 Cambridge Street
Boston, MA 02202
telephone
fax
e mail geministar@aol.com (america on line)
working group #2:GoM

Dr. Donald Phelps
EPA
Narragansett Marine Lab
27 Tarzwell Drive
Narragansett, RI 02882
telephone 401-782-3004
fax 401-782-3030
e mail dphelps%narvax@epavax.rtpnc.epa.gov
working group #1:RMRP

Dr. Jeffrey S. Rosen
American Mangement Systems, Inc.
54 Old Oaken Bucket Road
Scituate, MA 02066
telephone 617-544-3085
fax 617-544-3086
e mail jeff_rosen@mail.amsinc.com
working group #2:GoM

Dr. Maxine Schmidt
43 Fairview Avenue
Northampton, MA 01060
telephone 413-585-0644
fax
e mail mgschmidt@aol.com
working group #1:RMRP

Dr. Dylan Scott
Massachusetts Coastal Zone Management
Executive Office of Environmental Affairs
Leverett Saltonstall Building
100 Cambridge Street
Boston, MA 02202
telephone
fax
e mail
working group #1:RMRP

Dr. Peter C. Smith
Bedford Institute of Oceanography
Coastal Oceanography, PCS
P.O. Box 1006
Dartmouth, Nova Scotia, B2Y 4A2
CANADA
telephone 902-426-3474
fax
e mail pc_smith@bionet.bio.dfo.ca
working group #1:RMRP

Ms. Nancy Soderberg
USGS
Woods Hole, MA 02543
telephone
fax
e mail nsoderberg@usgs.er.gov
working group #2:GoM

Mr. John Sowles
Maine Dept. of Environmental Protection
State House Station #17
Augusta, ME 04333
telephone 207--287-7779
fax 207-287-7826
e mail
working group #2:GoM

Dr. Jennifer Steele
North Carolina State University
South Atlantic RMR Program
Box 8605
Raleigh, NC 27695
telephone
fax
e mail tuck@unity.ncsu.edu
working group #2:GoM

Dr. Richard Taupier
Massachusetts Coastal Zone Management
Executive Office of Environmental Affairs
Leverett Saltonstall Building
100 Cambridge Street
Boston, MA 02202
telephone
fax
e mail rtaupier@mis@eoea
working group #2:GoM

Dr. David Townsend
GoM Regional Marine Research Program
University of Maine
Associate Research Professor of Oceanography
Department of Oceanography
Orono, ME 04469
telephone 207-581-4367
fax 207-581-1426
e mail davidt@maine.maine.edu
d.townsend (omnet)
working group #1:RMRP

Mr. Bruce W. Tripp
Woods Hole Oceanographic Institution
Assistant Director, Coastal Research Center
Woods Hole, MA 02543
telephone 508-457-2000 ext. 2900
fax 508-457-2172
e mail btripp@whoi.edu
b.tripp (omnet)
working group #1:RMRP

Dr. John Unger
USGS
MS-922
Reston, VA 22092-0001
telephone
fax
e mail junger@mtn.er.usgs.gov
working group #2:GoM

Dr. Robert Wall
GoM Regional Marine Research Program
University of Maine
Center for Marine Studies
ME Sea Grant College Program
5715 Coburn Hall
Orono, ME 04469
telephone 207-581-1435
fax 207-581-1426
e mail r.wall (omnet)
working group not indicated

Dr. Gordon Wallace
University of Massachusetts @ Boston
Environmental Sciences Program
Boston Harbor Campus
100 Morrissey Boulevard
Boston, MA 02125
telephone 617-287-7440
fax 617-287-7474
e mail wallace@umb.edu
working group #1:RMRP

Dr. Peter H. Weibe
WHOI
Biology Department
Woods Hole, MA 02543
telephone 508-548-1400
fax
e mail pwiebe@whoi.edu
working group #1:RMRP

Dr. Stephan Zeeman
University of New England
Department of Life Sciences
11 Hills Beach Road
Biddeford, ME 04005
telephone 207-283-0171
fax
e mail siz@aip.org
working group #1:RMRP

Appendix C: Posters

The Gaea Project: An Overview

Nabil I. Hachem
Michael A. Gennert
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609-2280

{hachem,michaelg}@cs.wpi.edu

Abstract

This paper describes the Gaea System, a Spatio-Temporal Database Management System under development at Worcester Polytechnic Institute. The long-term goal of the Gaea project is to develop an extensible, object-oriented data management and analysis system to be used by researchers in the field of global change. The current goal is to develop a prototype, which can be used by geographers in a user-friendly manner, yet permits integration of heterogeneous and complex data types, and interactive development of sophisticated methods for data analysis, prediction, and display. Focus is on the object manipulation and analysis aspects of the system, and the management of “meta-data,” that is, data about the data.

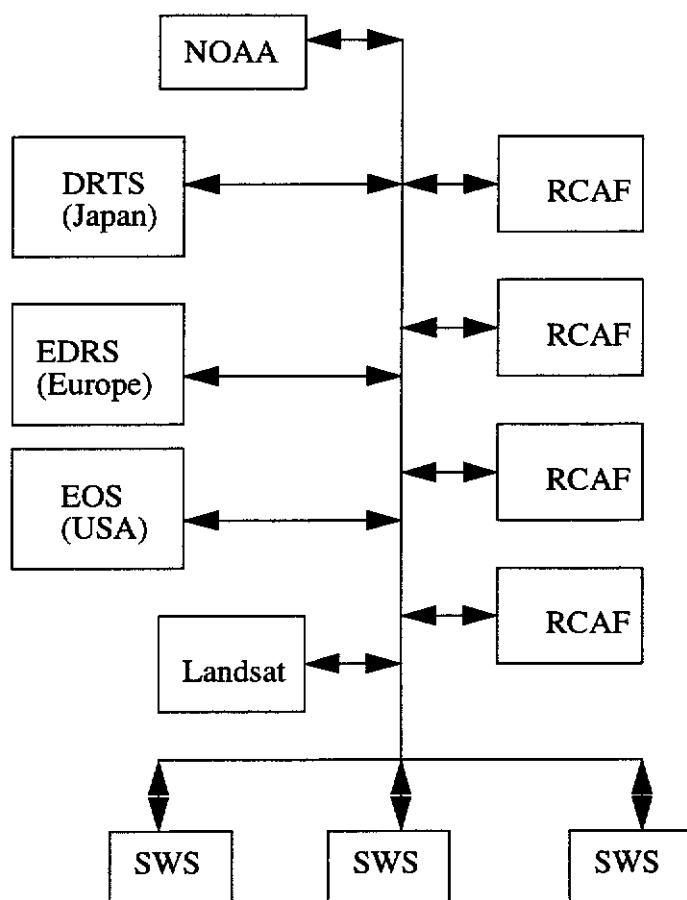
Scientific Databases

- Manage data for scientific investigation
- Common problems transcending disciplines
 - Computational biology, chemistry, physics
 - Computer-aided design and manufacturing
 - Earth and space sciences
 - Geographic information systems
 - Medical imaging
- Scientists distracted by bits and bytes—prefer to deal with scientific concepts
- Current database systems designed for business—inappropriate for scientific users

General Abilities of Gaea

- Deriving information about spatio-temporal objects
- Maintaining information on the evolution of data objects
- Providing derivation semantics for data objects
- Providing a visual environment for browsing, querying, and analysis
- Providing user-extensible data and operator types

Seamless Integration

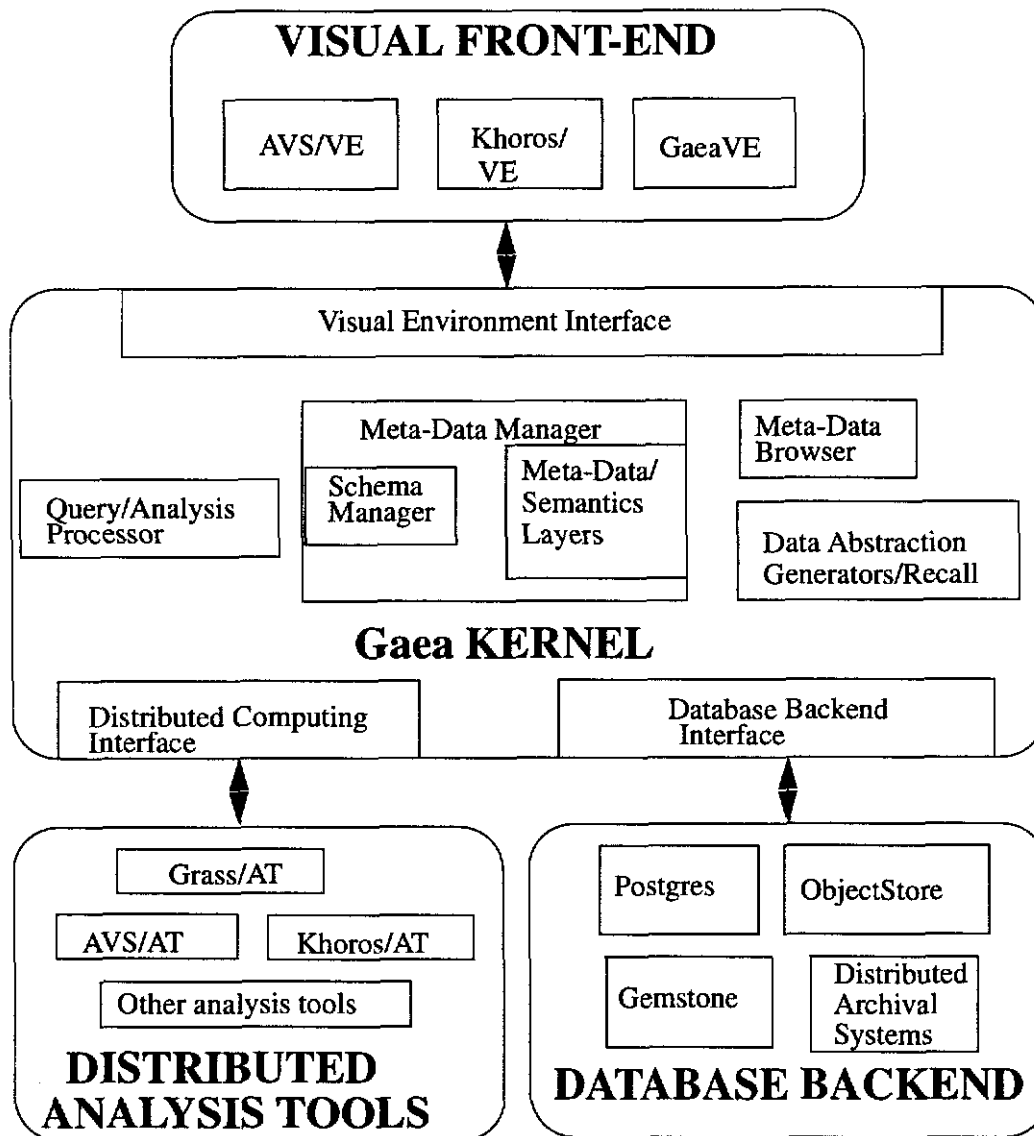


SWS: Scientific Workstation
RCAF: Remote Computing and Analysis Facility
EOS: Earth Observation System
EDRS: European Data Relay Satellite
DRTS: Data Relay and Tracking Satellite
NOAA: National Oceanographic and
Atmospheric Administration

Example Use

- Define focus problem – dataset independent
- Select spatio-temporal region
- Browse available data
- Data may be distributed
- Derive new from acquired or derived data
- Capture system interaction automatically
- Visualize results

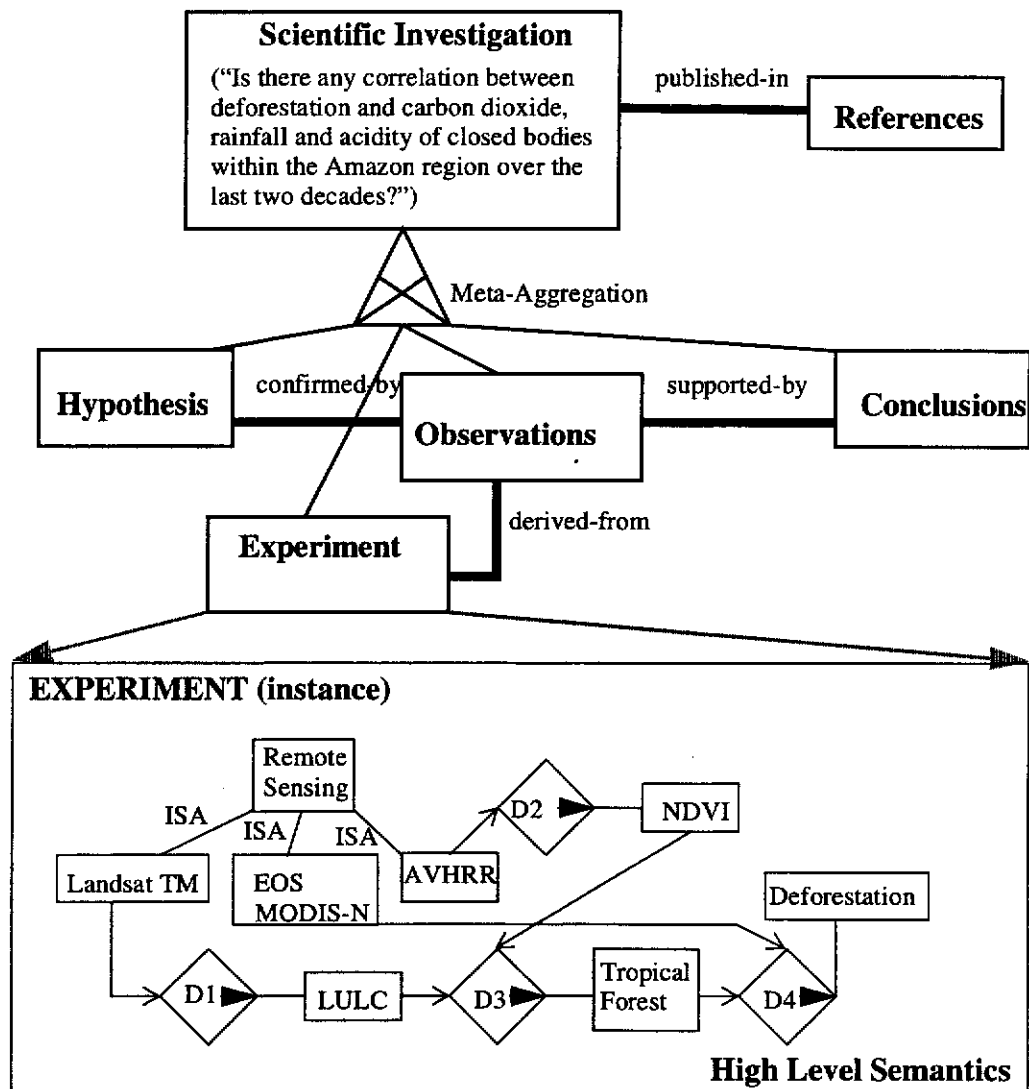
Long Term Vision of the Gaea Architecture



Visual Frontend

- Allow visual queries
- Primarily point-and-click operation
- Incorporate different interfaces
- Convert queries to general query language
- Display results

Semantic Model

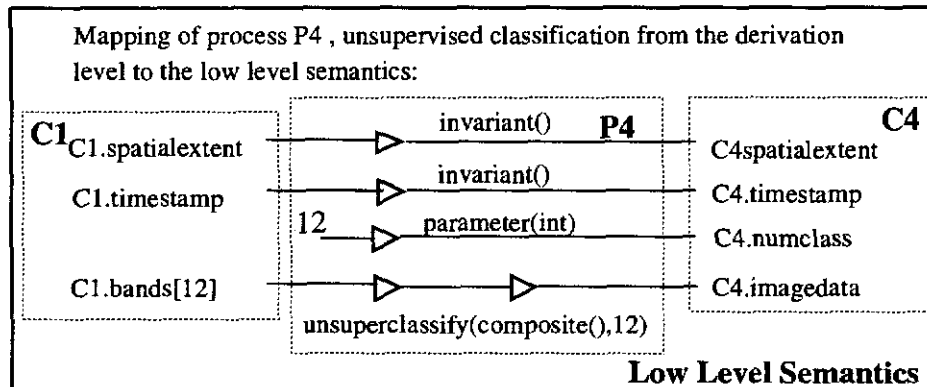
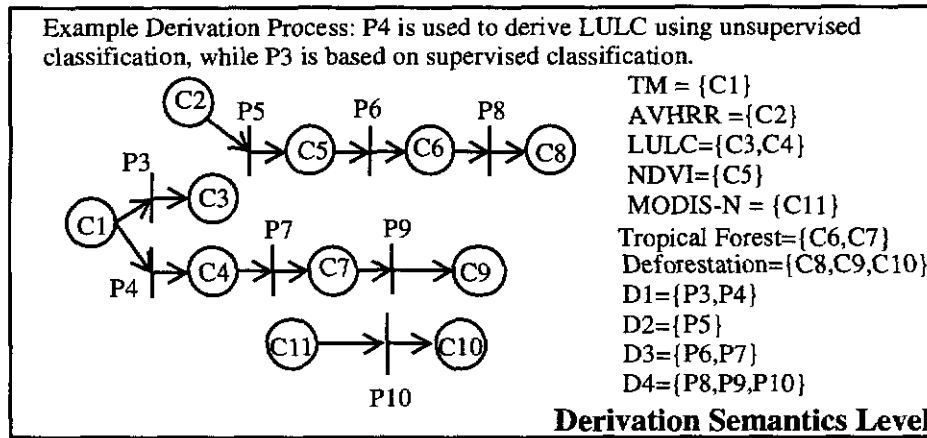
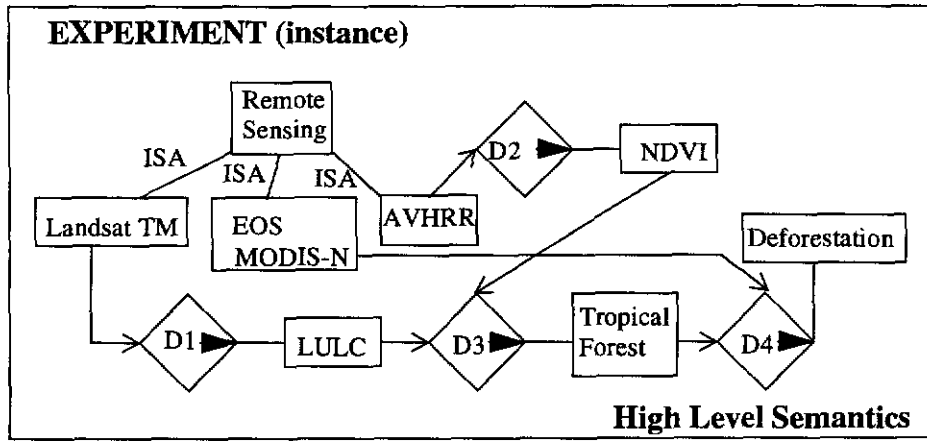
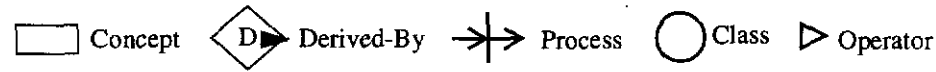


Gaea Kernel

- Query processing
- Manage metadata

- Understand query semantics at 3 levels
 - Concept level—where user primarily works
 - Process level—network of processes
 - Systems level—lowest level of operators, includes implementation of ADTs
- Optimizations possible

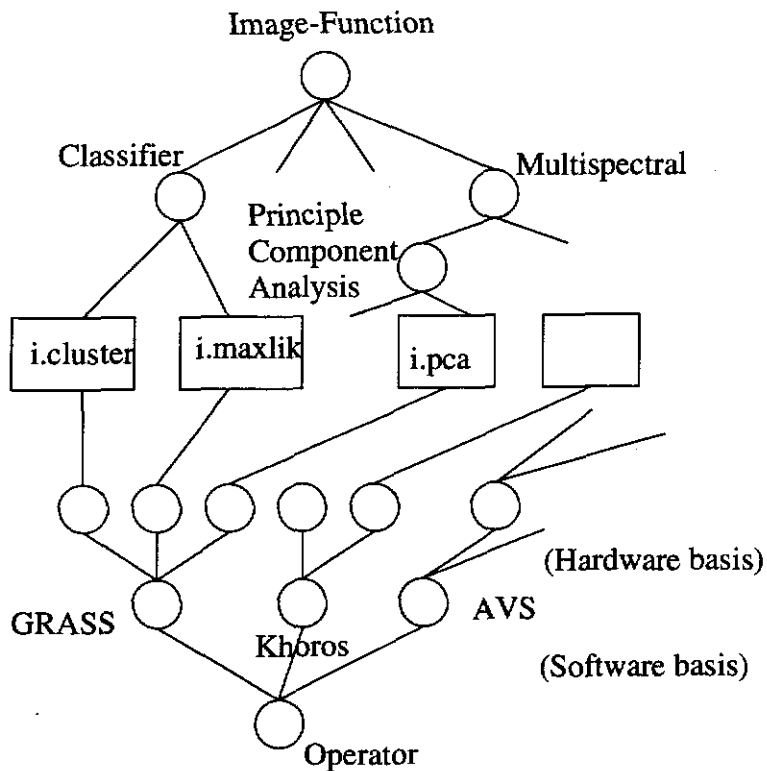
Derivation Management Layers



Analysis Tools

- Incorporate different software packages—AVS, Khoros, IDL, PV-WAVE
- DO NOT REWRITE!
- Encapsulate routines
- Describe tool with meta-data

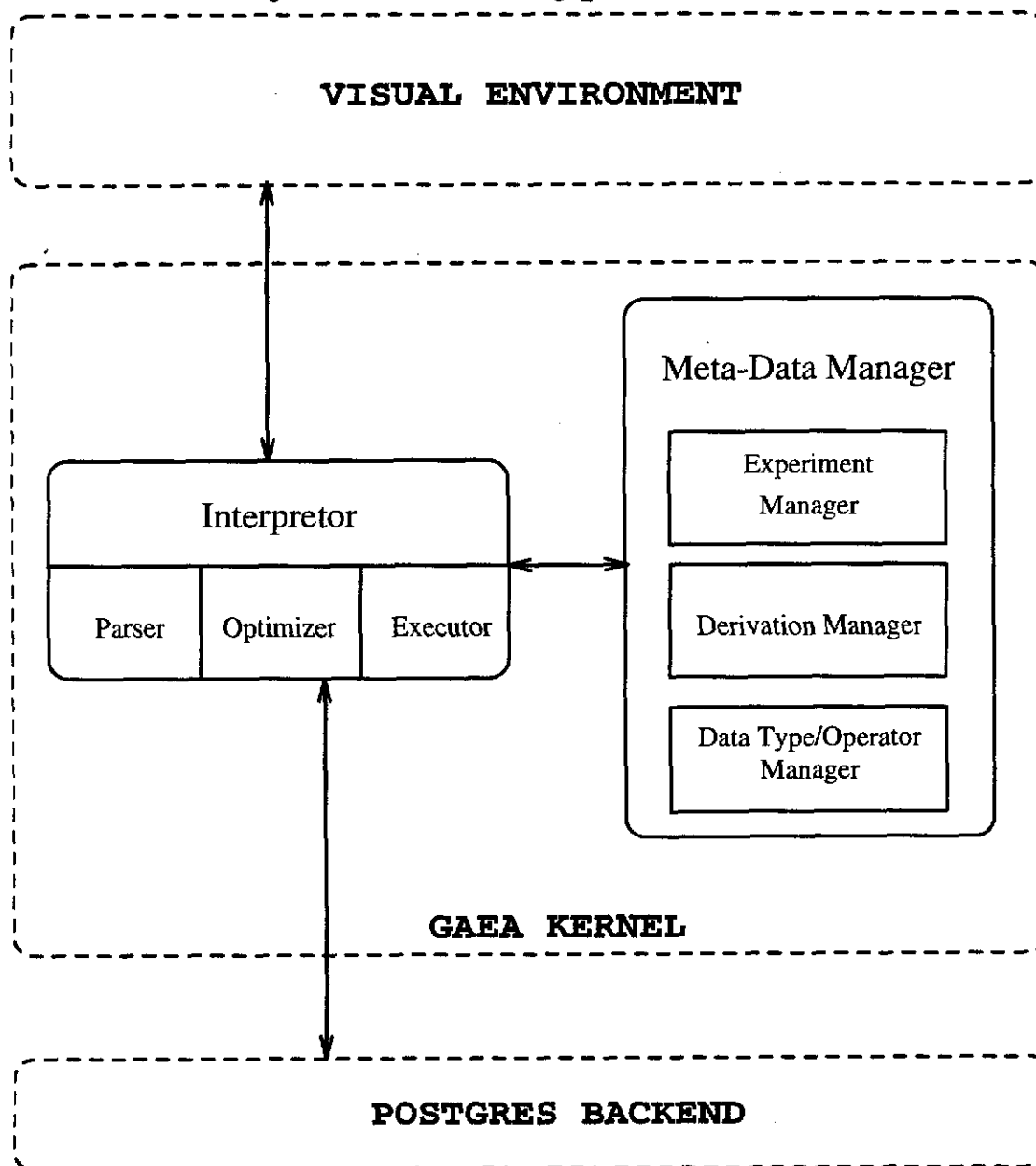
Operator Taxonomies



Current Status

- Identified global change needs
- Designed extensible Gaea architecture: VE, Kernel, DB
- Developed model for temporal data
- Prototype nearing completion
 - data and operator models
 - data language
 - Gaea VE, including browser
 - Gaea kernel
 - Postgres backend

The Gaea System Prototype



Future Work

- Object editor
- New visual query language
- History mechanism
- Include operators from GIS, image processing, etc.
- Network access
- Optimization
- Applications

Atlantic Fisheries Adjustment Program

(AFAP)

Players: Ken Drinkwater, Don Lawrence, John Loder,
Brian Petrie, Peter Smith (PCS; BIO)
Fred Page, Steve Smith (MFD; STAB, BIO)

Motivation: Need to predict long-term variability of coastal ocean environment in order to conserve and rebuild groundfish stocks (Haché, 1989)

Goals:

- 1) determine spatial and temporal scales of T,S variability on the continental shelf,
- 2) understand the climatological variability on the continental shelf and its relationship to the fisheries, and
- 3) establish a long-term monitoring program on the continental shelf.

Tools:

- 1) hydrographic (T,S) data base for the Gulf of Maine and Scotian Shelf, and
- 2) climate indices data base

Climate Indices:

- 1) coastal SST, eg. St.Andrews, Boothbay Harbor,
- 2) offshore SST from ships-of-opportunity,
- 3) selected subsurface T,S eg. Prince 5, Station 27, lightships,
- 4) river discharge,
- 5) shelf/slope water boundary, Gulf Stream "north wall" positions,
- 6) sea ice indices,
- 7) barometric pressure and air temperature,
- 8) large-scale atmospheric indices, eg. NAO

Hydrographic Data Base:

- data types: CTD, bottle, SST, BATFISH, TESAC, moored instrument
- sources: MEDS
NODC
other (eg. UNH '86-'87)
- rigorous QC at MEDS, then BIO
- size:

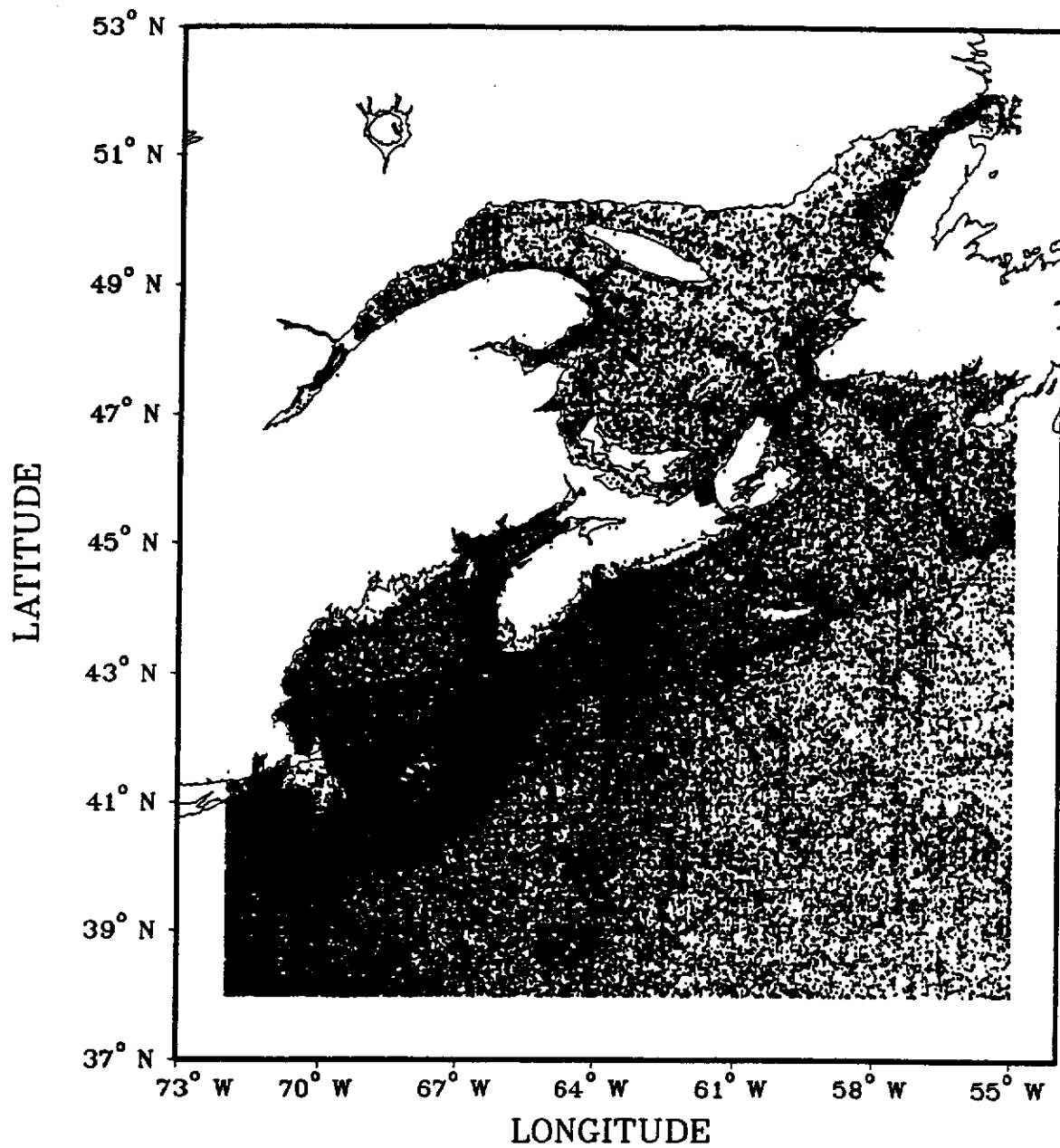
	GoM	S.S.
stns.	88,988	82,566
recs.	1,650,776	1,970,500
- updates: monthly new data files;
annual updates to data base
- future expansion: GoSl (MLI)
Newfoundland Shelves (NAFC)

Emerald Basin Studies:

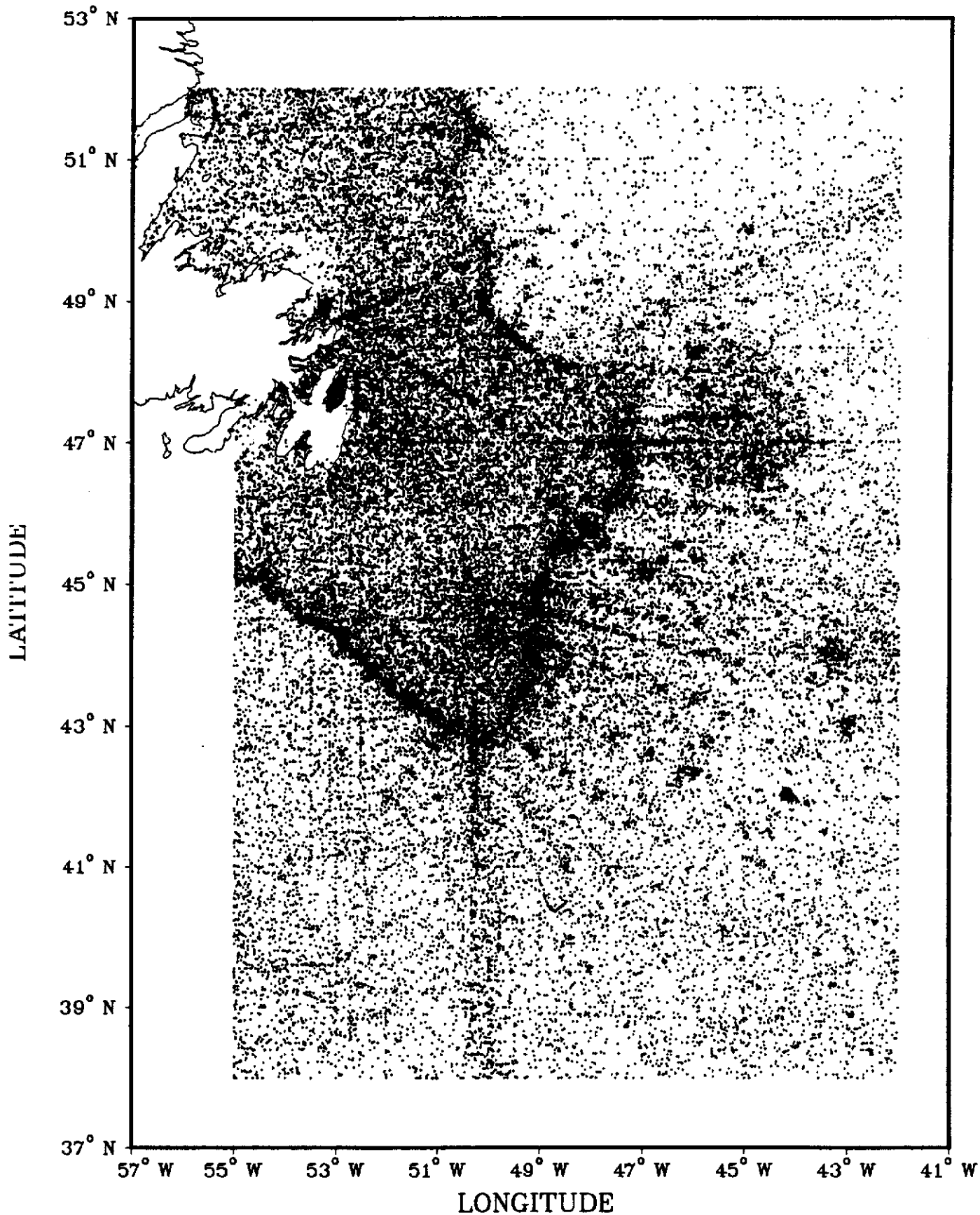
- analysis of monthly mean temperature anomalies: 1945-88
- 10-year cycle identified
 - 1) similar at all depths,
 - 2) regional coherence: MAB --> LC,
 - 3) maximum variance at 100-150 m suggests an offshore source,

Monitoring Proposal: opportunistic sampling of Emerald Basin hydrographic properties

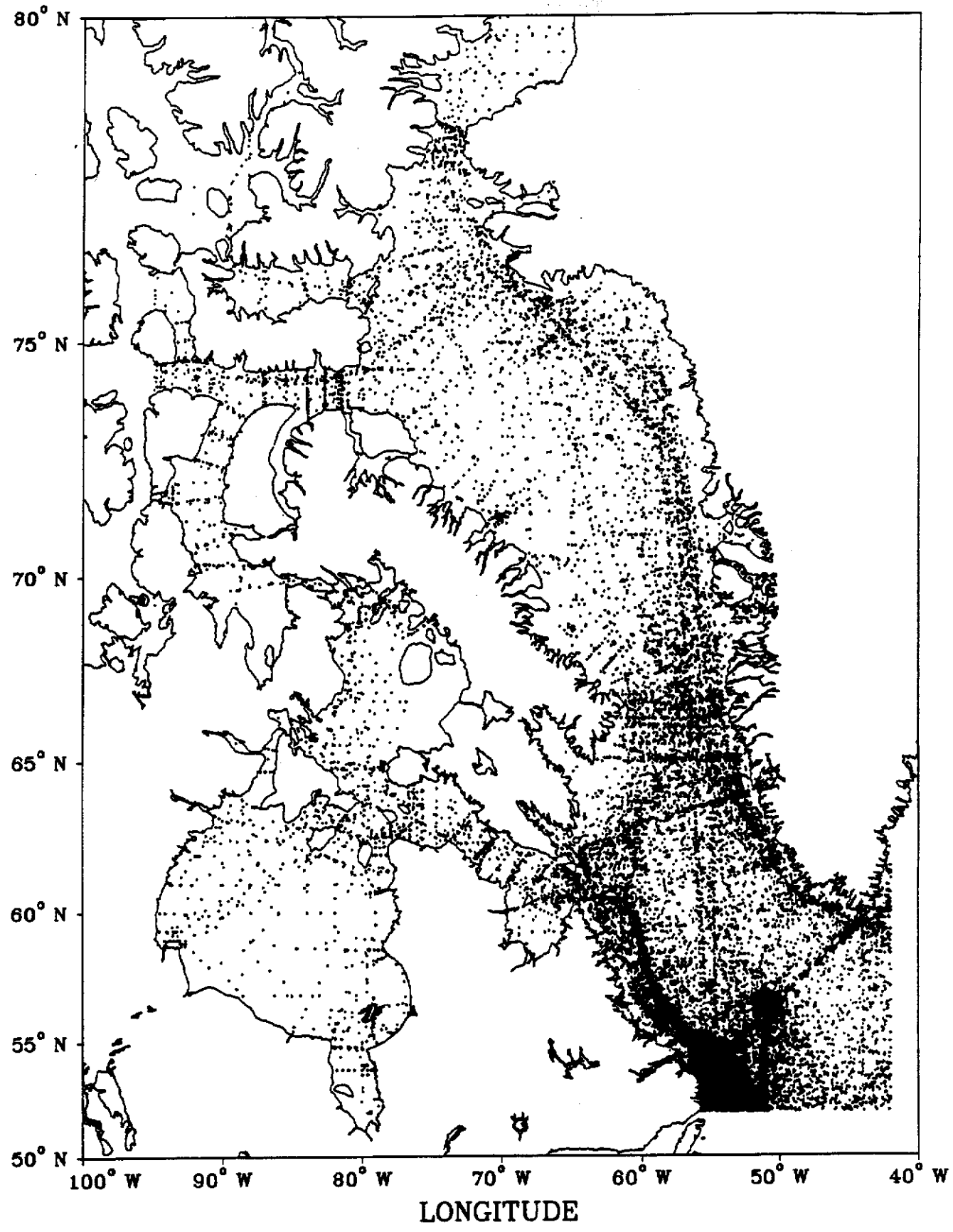
AFAP Database Station Positions Nov.1992



AFAP Database Station Positions (May. 1993)

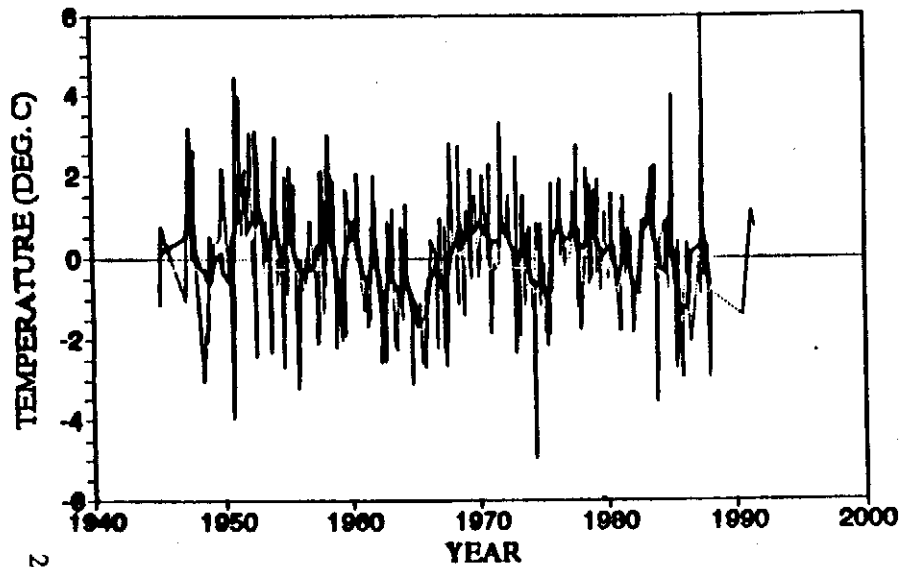


AFAP Database Station Positions (May. 1993)

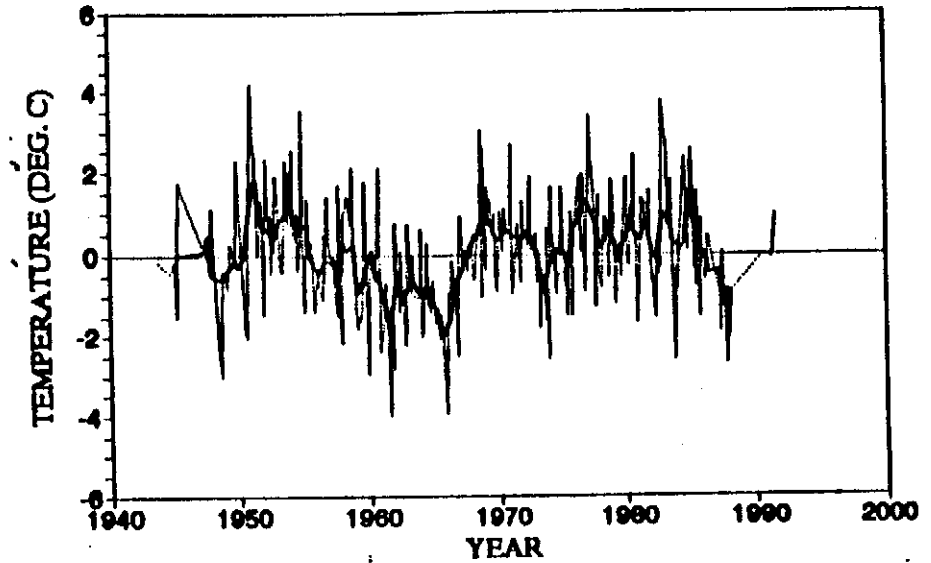


EMERALD BASIN

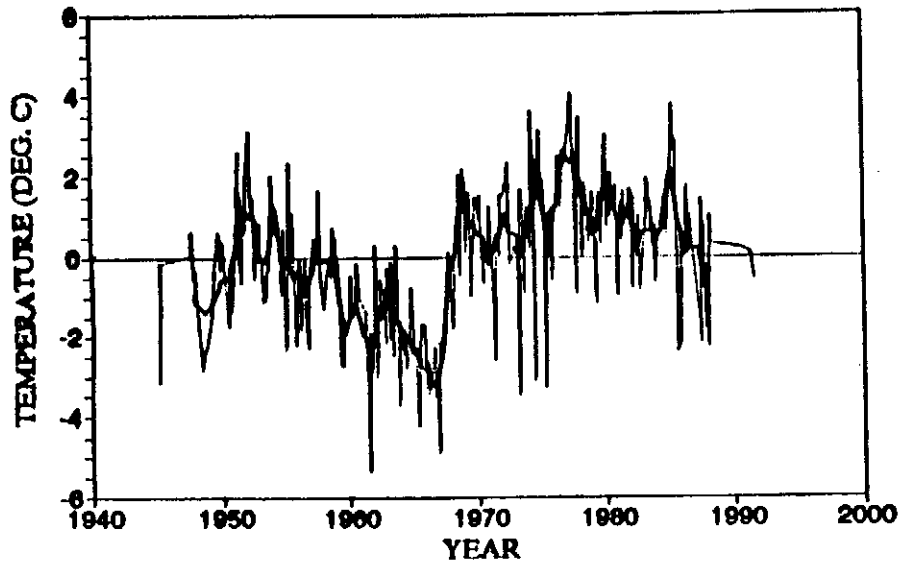
TEMPERATURE ANOMALY 0 M



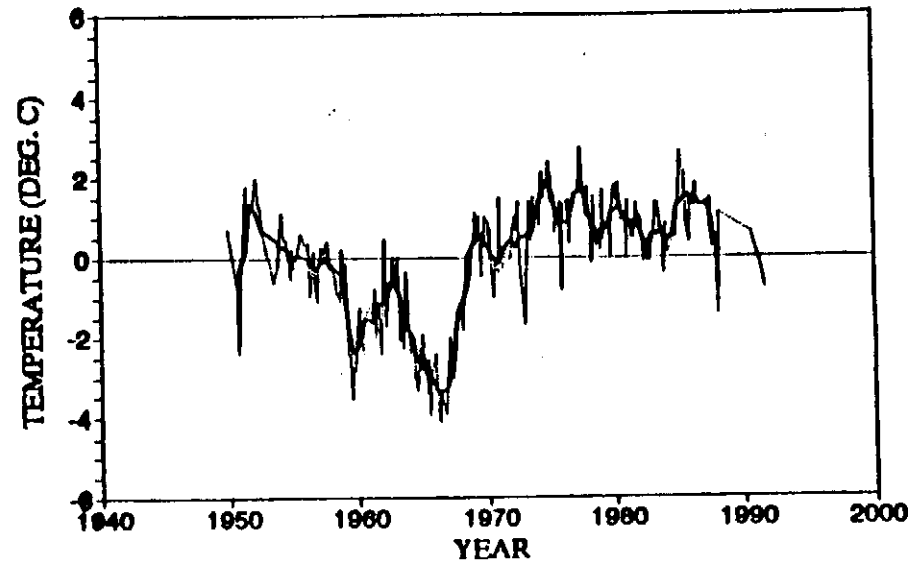
TEMPERATURE ANOMALY 50 M



TEMPERATURE ANOMALY 100 M



TEMPERATURE ANOMALY 150 M



Monthly means (light line) and filtered values (heavy line) of the temperature anomalies in Emerald Basin at 0, 50, 100 and 250 m.

