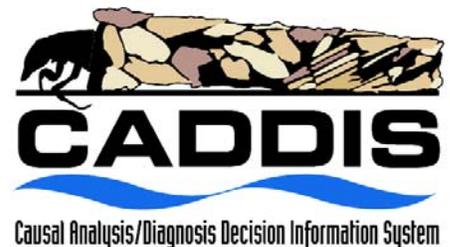


US EPA ARCHIVE DOCUMENT

Summary Report for the Workshop on the Causal Analysis/Diagnosis Decision Information System (CADDIS)

US EPA ARCHIVE DOCUMENT



Summary Report for the Workshop on the
Causal Analysis/Diagnosis Decision Information System
(CADDIS)

U.S. Environmental Protection Agency
Mt. Sterling, OH
August 26–28, 2002

National Center for Environmental Assessment–Washington Office
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

This document has been reviewed in accordance with U.S. Environmental Protection Agency (EPA) policy and has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

The drafts of this report were prepared by T.N. and Associates under an EPA contract (Contract No. 68-C-98-187, Work Assignment No 2002-04) as a general record of discussions during the Causal Analysis/Diagnosis Decision Information System Workshop. The drafts were reviewed by EPA personnel, and revised and finalized by the EPA Work Assignment Manager. Background materials for the workshop were prepared by T.N. and Associates under an EPA contract (Contract No. 68-C-98-187, Work Assignment Numbers 2001-09 and 2002-02) and were reviewed and approved for distribution by the EPA Work Assignment Manager. This workshop was co-sponsored by EPA's National Center for Environmental Assessment and National Exposure Research Laboratory. As requested by EPA, this report captures the main points and highlights of discussions held during plenary and break-out sessions. The report is not a complete record of all details discussed nor does it embellish, interpret, or enlarge upon matters that were incomplete or unclear. Statements represent the individual views of each workshop participant; none of the statements represent analyses by or positions of the EPA.

ACKNOWLEDGMENTS

The CADDIS Workshop Steering Committee thanks Seemeen Hashem from Aqueous Studios for the CADDIS logo. The design is inspired by caddisflies, whose occurrence in streams often indicates high quality, and the intricate cases they build from pebbles and sticks. We also thank David Bussard, Kate Smith, Tony Maciorowski, Don Brady, and Jonathon Herrmann for their support, and Donna Reid-Judkins, who co-chaired the original Stressor Identification workgroup.

CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	v
AUTHORS AND REVIEWERS	vi
LIST OF ACRONYMS	vii
1. EXECUTIVE SUMMARY	1-1
1.1. System Functions Should Focus on Guidance, a Knowledge Base and User Interaction	1-2
1.2 System Development Should Be User-Focused, Modular and Iterative	1-3
1.3 Conclusions	1-3
2. INTRODUCTION	2-1
2.1 Workshop Objectives	2-1
3. PLENARY SESSION	3-1
3.1 Stressor Identification History	3-1
3.2 Stressor Identification Process Overview	3-2
3.3 CADDIS	3-5
3.4 Design Options and Prototype	3-7
3.5 An Approach for Developing Innovative Software	3-8
4. BREAK-OUT GROUP REPORTS	4-1
4.1 Break-out Groups A and B	4-1
4.1.1 Session 1: Defining the Impairment and Listing Candidate Causes	4-1
4.1.2 Session 2: Analyzing the Evidence	4-7
4.1.3 Session 3: Characterizing the Causes	4-10
4.2 Break-out Group C	4-13
4.2.1 Session 1: Issues Surrounding System Development	4-13
4.2.2 Session 2: QA Requirements and Issues	4-22
4.2.3 Session 3: Interaction With and Among Users	4-26
5. FINAL PLENARY DISCUSSION	5-1
6. CONCLUSIONS	6-1
APPENDIX A: AGENDA, PARTICIPANTS AND BREAK-OUT GROUP ASSIGNMENTS	A-1
APPENDIX B: BACKGROUND MATERIALS	
Neptune Paper (CADDIS)	B-1
Menzie Paper No. 1 - U.S. EPA Sources	B-26
Menzie Paper No. 2 - Additional U.S. EPA Sources (Mid-Atlantic focus)	B-48

CONTENTS (continued)

APPENDIX C: CASE STUDY MATERIALS AND REFERENCES

Cabin John Case Study	C-1
Little Scioto Case Study	C-8

LIST OF TABLES

1. Issues to consider when choosing a web-based system and a local application 4-20

LIST OF FIGURES

1. The management context of the SI process 3-3
2. Conceptual diagram of the CADDIS project, research and user communities 3-6
3. Where CADDIS may aid the process of impaired waters 303(d) listing and TMDL development 4-16

AUTHORS AND REVIEWERS

The National Center for Environmental Assessment-Washington Office (NCEA–W) of EPA’s Office of Research and Development was responsible for the preparation of this document. The first and second drafts were prepared by T.N. and Associates under an EPA contract (Contract No. 68-C-98-187, Work Assignment No 2002-04). The drafts were reviewed by the workshop participants, and revised and finalized by the EPA Work Assignment Manager.

AUTHORS

Barbara Brown, U.S. EPA, ORD

Susan Cormier, U.S. EPA, ORD

Doug Norton, U.S. EPA, OW

Susan Norton, U.S. EPA, ORD

Leela Rao, U.S. EPA, ORD

Bhagya Subramanian, U.S. EPA, ORD

Glenn Suter, U.S. EPA, ORD

William Swietlik, U.S. EPA, OW

Paul Black, Neptune and Company, subcontractor to T.N. and Associates

Cheri Butler, Menzie-Cura and Associates, subcontractor to T.N. and Associates

Jackie Little, T.N. and Associates

Charlie Menzie, Menzie-Cura and Associates, subcontractor to T.N. and Associates

Tom Stockton, Neptune and Company, subcontractor to T.N. and Associates

REVIEWERS

The workshop participants (Appendix A) were given the opportunity to review the draft workshop report for accuracy in the content captured from the plenary and break-out sessions. The comments received from the participants have been addressed and incorporated into the report. A second round of review was conducted on the revised draft by two U.S. EPA ORD workshop participants—Steve Fine, Scott Minamy—to ensure that the accuracy was maintained.

LIST OF ACRONYMS

BASINS	Better Assessment Science Integrating Point and Nonpoint Source
CADDIS	Causal Analysis/Diagnosis Decision Information System
CD	Compact Disc
CM	Conceptual Model
CWA	Clean Water Act
ECOTOX	Ecological Toxicity Database
EMAP	Environmental Monitoring and Assessment Program
U.S. EPA	U.S. Environmental Protection Agency
MIMS	Multimedia Integrated Modeling System
NCEA	National Center for Environmental Assessment
NGO	Non-government Organization
ORD	Office of Research and Development
OS	Operating System
OW	Office of Water
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
SI	Stressor Identification
SPRC	Strategic Planning and Research Coordination
STAR	Science to Achieve Results
STORET	STOrage and RETrieval database
TMDL	Total Maximum Daily Load
WATERS	Watershed Assessment Tracking and Environmental Results System

1. EXECUTIVE SUMMARY

Increasingly, the regulatory, remedial and restoration actions taken to manage impaired environments are based on measurement and analysis of the state of the biotic community. When an aquatic community has been identified as impaired, the cause of the impairment must be determined so that appropriate actions can be taken. The Stressor Identification (SI) Guidance, published in December 2000, describes a methodology for identifying the causes of observed impairments in aquatic systems. The ability to identify stressors accurately and defend those findings with supporting evidence is a critical step in developing strategies that will improve the quality of aquatic resources.

SI requires extensive knowledge on a variety of stressors as well as depth of knowledge on the mechanism, symptoms, and stressor-response relationships for specific stressors. The goal of the Causal Analysis/Diagnosis Decision Information System (CADDIS) project is to support investigators in Regions, States and Tribes in the determination of causes of aquatic impairments. We envision a decision support system that will help investigators find, access, organize and share information useful for causal evaluations in aquatic systems.

At a workshop in Mt. Sterling, OH, August 26–28, 2002, representatives from the Office of Research and Development (ORD), Office of Water (OW), U.S. Environmental Protection Agency (U.S. EPA) Regions and States came together to conceptualize CADDIS and to identify critical research needs for system implementation and population. This report summarizes the main points and highlights of the workshop. These results may be of interest to scientists involved with diagnosing impairments, as well as developers of decision support systems. The report will also be used as the basis of a strategic plan for system development and as a source of specific recommendations to be incorporated directly into the system's design, platform and architecture.

This report follows the general structure of the workshop. The workshop began with plenary presentations summarizing the history and vision for the project, after which the participants were divided into three break-out groups. Two groups used case studies to provide a context for identifying high-priority functions and research needs. The third group discussed issues of system development, quality assurance and the interaction with and among users in the context of the potential user groups. A final plenary session was used to summarize and synthesize participant views.

Overall, participants felt that:

1. Professional judgment could be aided, but not replaced by, a computer

2. CADDIS should focus on providing information, a guide through the SI process and support for the users, rather than actually making judgments (e.g., through statistical analysis).

In other words, CADDIS should support the SI process rather than try to perform the process. Useful feedback on high-priority system functions and considerations for system development that emerged during the workshop is described in greater detail next.

1.1. System Functions Should Focus on Guidance, a Knowledge Base and User Interaction

The functions that were identified as the greatest help to potential users included guidance and documentation, a knowledge base, and conceptual model (CM) development. Specifically, workshop participants stated that the most fundamental feature of CADDIS should be the capability to lead the user through the SI process and prompt him or her to provide information, make judgements, or determine the relevance of a step or consideration. This guidance could include providing the user with blank forms and tables to fill out at appropriate stages, general advice on avoiding common pitfalls, or suggestions on how best to perform the analyses and inferences. High-priority areas for more detailed guidance included developing the initial list of candidate causes, the appropriate use of statistics and the scores used for comparing strength of evidence. Once the user has completed the SI, CADDIS should then provide some sort of report that could be reviewed, revised and printed.

Workshop participants also identified the creation of a knowledge base as a high priority. The knowledge base should include stressor-response relationships, organism tolerances, probabilities of association for stressors and impairments and a searchable database of case studies. The need for States to have easy access to stressor-response information, particularly for the highest priority stressors, was frequently voiced. The highest priority stressors might be gleaned from the current knowledge of stressors influencing total maximum daily loads (TMDLs) and known information gaps (e.g., sediment, flow extremes, habitat modifications), supplemented by user input. In addition to making stressor-response data more accessible, the need for literature syntheses and shorter synopses was highlighted. Since the development of a comprehensive set of stressor-response relationships might require additional research, CADDIS should have a feedback mechanism to inform ORD research planners of user-identified information gaps.

Providing a database of case studies was seen as a way to enable the community as a whole to improve expertise in causal evaluations. Making the database searchable would allow users to find situations similar to the ones they are evaluating. All of the break-out groups recognized that CADDIS will need to have a mechanism for informing the users of the quality of the data and information in the knowledge base, as well as prompt the users to keep track of

quality assurance (QA) information for the data they themselves are using for the causal evaluation.

A third functionality that the workshop participants identified as a high priority was aiding the users in CM development. The initial help provided in this area should consist of a database of standard CMs for typical situations. The users could modify these examples to allow for customization to a given site. Future developments might include a graphics tool that allows the user to pick a source or stressor and connect it to effects. Such a tool would not only be useful for demonstrating the link from SI to TMDL efforts, but may also be attractive to other Agency and State programs.

1.2. System Development Should Be User-Focused, Modular and Iterative

CADDIS workshop participants felt strongly that early and ongoing input from user groups was essential to the success of the project. In addition to seeking a representative from the user community to serve on the development team, feedback could be solicited through the SI web page, regionally sponsored sessions and visits to scientists in the States and Tribes. However, it was also noted that the need to engage the user community does not preclude making initial progress on system development. In fact, it would be useful to have a prototype or a system with limited functionality available to potential users to help solicit feedback.

The workshop participants supported the concept of a phased, iterative, and modular development process. A phased and iterative development process would help solicit user and management advice, allow the development team to showcase successes quickly and often, and would help keep the system focused on, and relevant to, changing user needs. A modular CADDIS system would allow a broader user base to take advantage of individual components (e.g., CM builder, stressor-response databases) of the system. Modularity would also make it easier to match the strengths of different system platforms with functional needs.

The other major element of system design that participants discussed was communication/feedback mechanisms for users of CADDIS. A communication mechanism is important for facilitating both technical support and the sharing of user data and experiences. Participants expressed interest in dovetailing on existing lines of communication for initial user contact and marketing (e.g., existing mailing lists, newsletters, websites and conference calls), but felt that CADDIS should also have its own listserv and online request and submission forms to conduct technical support and user communication once the system is in place.

1.3. Conclusions

There was a general sense that the project needs to be on a fast track. States are moving forward on how to address impaired waters listing and TMDL-related issues. Emphasis on early

design considerations and outreach to user groups by workshop participants means that a CADDIS development team will need to be in place as soon as possible. This development team should include not only U.S. EPA technical staff, but also representatives from the user community and U.S. EPA management. Although early progress is essential, workshop participants also emphasized that CADDIS would be a long-term commitment for both developers and users. With the right input and support from all groups, CADDIS will become a system that can substantially help investigators identify causes of impairments in aquatic systems.

2. INTRODUCTION

2.1 WORKSHOP OBJECTIVES

Increasingly, the regulatory, remedial and restoration actions that are taken to manage impaired environments are based on measurement and analysis of the state of the biotic community. When an aquatic community has been declared impaired, it becomes necessary to determine the cause of the impairment so that appropriate actions can be taken. In response to that need, ORD and OW developed the Stressor Identification (SI) Guidance, which was published in December 2000.¹ The Guidance describes a methodology for identifying the causes of observed impairments in aquatic systems. The ability to identify stressors accurately and defend those findings with supporting evidence is a critical step in developing strategies that will improve the quality of aquatic resources.

Use of the SI process is prompted by a biological impairment. The general SI process entails critically reviewing available information, forming possible causal scenarios that might explain the impairment, analyzing those scenarios, and producing conclusions about which stressor or stressors are causing the impairment. The SI process is flexible enough to support varying water management requirements and stringencies in identifying stressors.

Implementing the SI process requires extensive knowledge about a variety of stressors as well as depth of knowledge on the mechanism, symptoms and stressor-response relationships for specific stressors. The SI Workgroup acknowledged that making the guidance useful for users in the Regions and States will require (1) developing easier access to supporting information, (2) providing investigators with additional support in organizing and analyzing information, (3) assisting the users in performing logical inferences, and (4) facilitating the sharing of experience and knowledge gained in conducting causal evaluations. The need for an expert system to support causal evaluation was also highlighted by the OW/ORD Strategic Planning and Research Coordination (SPRC) Diagnostics Group. The Group advised that the developed system should help States and Tribes diagnose causes of biological impairment by linking them with information they would not otherwise have the resources to access. As such, the overall objective of the Causal Analysis/Diagnosis Decision Information System (CADDIS) project is to develop a decision support system that will help investigators find, access, organize, apply and share information useful for causal evaluations in aquatic systems.

The objectives of the CADDIS workshop were to conceptualize CADDIS and to identify critical needs necessary to develop and implement the system. This report summarizes the main points and highlights of the workshop. These results may be of interest to scientists involved

¹U.S. Environmental Protection Agency 2000. *Stressor Identification Guidance Document*. EPA 822-B-00-025.

with diagnosing impairments, as well as developers of decision support systems. The report will also be used as the basis of a strategic plan for system development and as a source of specific recommendations to be incorporated directly into the system's design, platform and architecture.

The workshop began with plenary sessions that were designed to provide participants with background on the SI process, project history, and system design options. Following plenary presentations, in-depth discussions were conducted in three break-out groups. Break-out Groups A and B addressed system functionality and research needs from the perspective of two case studies (the Little Scioto, OH; and the Cabin John Watershed, MD). Both of these groups walked through the SI process in three sessions: session one focused on identifying the impairment and listing candidate causes, session two presented information on analyzing evidence, and session three concluded the SI process with characterizing and identifying cause. Appendix C of this report presents background materials and references that were used in development and presentation of these two case studies.

During each of the workshop sessions, participants discussed issues of system functionality and research needs in the context of the case study and the information provided for a particular step in the SI process. When discussing system functions, participants were asked to think of functionality that would be useful in CADDIS and to prioritize that list of functions. The participants were also requested to assign priority to information database needs and general research needs.

Break-out Group C addressed overall system development issues from the perspective of potential users. The first session specifically considered system development, and the participants were requested to give feedback on who they envisioned as CADDIS users and how those users should be engaged in the development process. Once a user base was identified, Group C discussed possible customer use of the system and implications of various design options. The second session was devoted to discussing the issue of quality assurance (QA) and considerations for the development of a Quality Assurance Project Plan (QAPP) for CADDIS. The final session focused back on the users and how the system might function to facilitate communication between users, developers and researchers. The participants were asked to develop some specific design options for communication, as well as to take into consideration the risks of facilitating communication among these groups.

At the conclusion of the workshop, a final plenary session was held to summarize the workshop participants' views on approaches to developing the system, high-priority functions, and critical dates for system development and use. The participants were each given the opportunity to make one final comment expressing a key point for CADDIS development.

This workshop report generally follows the organization of the workshop: a summary of the opening plenary sessions, synthesis of the break-out group sessions, and conclusions from

the final plenary session. Since the conclusions from the two case study groups were similar, they have been combined and are reported together by session. The outcome from Group C is presented separately.

3. PLENARY SESSION

3.1 SI HISTORY

Presented by William Swietlik, U.S. EPA, OW, Office of Science and Technology

Use of biological assessments and criteria as a tool to measure the health of waterbodies began in the early 1970s. Throughout the 1980s and 1990s, the U.S. EPA published numerous technical and policy guidances to further the nationwide use of biological assessments and criteria in surface water management. Currently, active biological assessment programs are in place in all the States and at least two Tribal Nations. Biological criteria have been formally adopted into the water-quality standards programs of five States and one Tribal Nation, and another 28 States are working to do so. Most of these States are now using biological assessments and criteria, among other tools, to help determine the condition of their waterbodies under Clean Water Act (CWA) sections 305(b) and 303(d).

With the increased use of biological assessments and criteria to list waterbodies as impaired under 303(d), States urgently need new tools to help them identify the causes of those impairments and to isolate and control the sources. States find this task especially difficult when the cause of the biological impairment is unknown, or when complex multiple causes are involved. States have even claimed that the lack of scientific tools to properly diagnose causes of biological impairment was becoming a major impediment to the further use of biological assessments and criteria in their water-quality programs. This is especially critical given the large number of legal actions involving biological impairments under the total maximum daily load (TMDL) program.

As a result, in May 1999, OW and ORD initiated a series of research planning meetings to identify critical short- and long-term research needs to ensure that States and Tribes would continue to integrate biological assessments and criteria into their water-quality programs. As a first effort, the SI workgroup outlined, drafted, and published the *Stressor Identification Guidance Document*. This document is designed to lead water-resource managers through a rigorous scientific process that identifies stressors causing biological impairment in aquatic ecosystems and provides a structure for organizing the scientific evidence supporting the conclusions. In addition to this, other important new tools were identified for ORD development. These included a database system for States to help investigators access, organize and share information useful for causal evaluations (i.e., CADDIS); models for identifying sources; stressor-response relationships for pollutants; models to predict response of biota to stressors; and expanded guidance for complex situations. At this time, OW's highest priority next step for SI tools is the development of CADDIS.

3.2 SI PROCESS OVERVIEW

Presented by Glenn Suter, (Modified from the *Stressor Identification Guidance Document*, EPA 822-B-00-025) U.S. EPA, ORD, National Center for Environmental Assessment (NCEA)

The SI process may be applied to any level of biological organization (e.g., individuals, populations, communities) and to any type of waterbody (e.g., freshwater streams, estuaries, wetlands, etc.). Some of the criteria presented for evaluating evidence may be specific; however, to a waterbody type (e.g., references to upstream/downstream associations). Similarly, the logic of the SI process may be applied in straightforward, single-stressor situations or in complex situations with multiple stressors and cumulative impacts. Complex situations may require investigators to refine the definition of the study area, gather new data, or do multiple iterations of SI to identify all the important stressors.

The SI Process

Figure 1 provides an overview of the SI process within the context of water-quality management and data collection. The SI process is initiated by the observation of a biological impairment (shown in the top box). Decision-maker and stakeholder involvement is shown along the left-hand side; their involvement is particularly important in defining the scope of the investigation and listing candidate causes. At any point in the process of identifying stressors, a need for additional data may be identified; the acquisition of this data is shown by the box on the right-hand side of the diagram. The accurate characterization of the probable cause allows managers to identify appropriate management action to restore or protect biological condition. Once stressors are identified and management actions are in place to control them, the effectiveness of the SI process (as demonstrated by improved conditions) can be monitored using appropriate monitoring tools and designs.

The core of the SI process is shown within the bold line of Figure 1 and consists of three main steps:

1. Listing candidate causes of impairment
2. Analyzing new and previously existing data to generate evidence for each candidate cause
3. Producing a causal characterization using the evidence generated in Step 2 to draw conclusions about the stressors that are most likely to have caused the impairment.

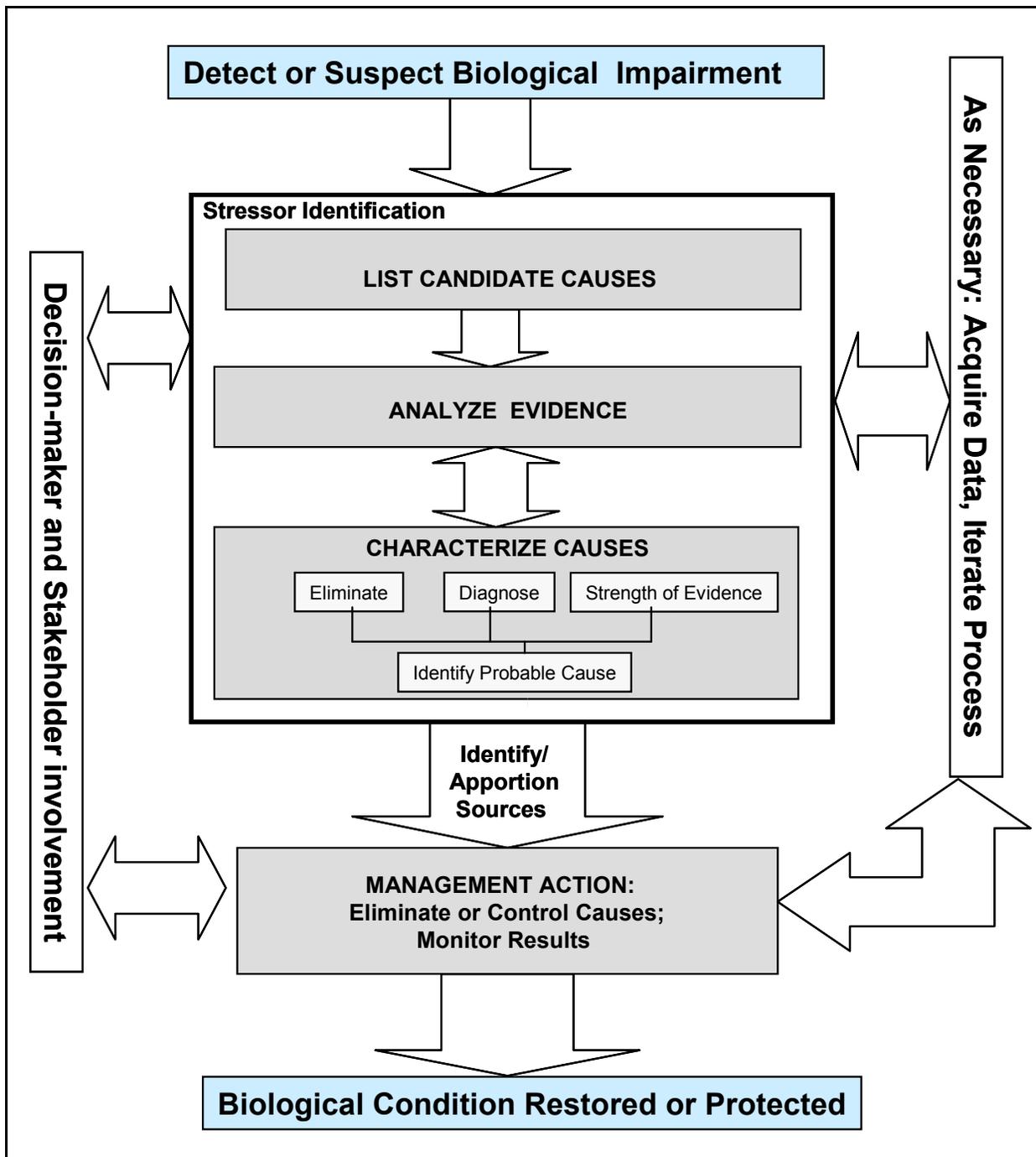


Figure 1. The management context of the SI process. (The SI process is shown in the center box with bold line. SI is initiated with the detection of a biological impairment. Decision-maker and stakeholder involvement is particularly important in defining the scope of the investigation and listing candidate causes. Data can be acquired at any time during the process. The accurate characterization of the probable cause allows managers to identify appropriate management action to restore or protect biological condition)

Source: Stressor Identification Guidance Document, EPA 822-B-00-025

Step 1 in the SI process is to develop a list of candidate causes, or stressors, that will be evaluated. This is accomplished by carefully describing the effect that is prompting the analysis (e.g., unexplained absence of brook trout) and gathering available information on the situation and potential causes. Evidence may come from the case at hand, other similar situations, or knowledge of biological processes or mechanisms. The outputs of this initial step are a list of candidate causes and a conceptual model (CM) that shows cause-and-effect relationships. Step 2, analyzing evidence, involves analyzing the information related to each of the potential causes. Virtually everything that is known about an impaired aquatic ecosystem is potentially useful in this step. For example, useful data may come from chemical analysis of effluents, organisms, ambient waters and sediments; toxicity tests of effluents, waters and sediments; necropsies; biotic surveys; habitat analyses; hydrologic records; and biomarker analyses. These data do not in themselves, however, constitute evidence of causation. The investigator performing the analysis must organize the data in terms of associations that could support or refute proposed causal scenarios. This includes considering levels of associations between:

- Measurements of the candidate causes and responses
- Measures of exposure at the site and measures of effects from laboratory studies
- Site measurements and intermediate steps in a chain of causal processes
- Cause and effect in deliberate manipulations of field situations or media

These associations comprise the body of evidence used to characterize the cause.

In Step 3, characterize causes, the investigator uses the evidence to eliminate, to diagnose, and to compare the strength of evidence to identify a probable cause. The input information includes a description of the effects to be explained, the set of potential causes, and the evidence relevant to the characterization. Evidence is brought in and analyzed as needed until sufficient confidence in the causal characterization is reached. In straightforward cases, the process may be completed in linear fashion. In more complex cases, the causal characterization may require additional data or analyses, and the investigator may iterate the process.

SI Process Iterations

The SI process may be iterative, beginning with retrospective analysis of available data. If the stressor is not adequately identified in the first attempt, the SI process continues using better data or testing other suspected stressors. The process repeats until the stressor is successfully identified. The certainty of the identification depends on the quality of information

used in the SI process. In some cases, additional data collection may be necessary to confidently identify the stressor(s). Although the SI process cannot accurately identify stressors without adequate data, completing the SI process is helpful even without adequate data because the exercise can help target future data collection efforts.

Using the Results of SI

SI is only one of several activities required to improve and protect biological condition (Figure 1). In some cases, the most effective management action will be obvious after the probable cause has been identified. In many cases, however, the investigation must identify sources and apportion responsibility among them. This can be even more difficult than identifying the stress in the first place (e.g., quantifying the sources of sediment in a large watershed), and may require environmental process models. The identification and implementation of management alternatives can also be a complex process that requires additional analyses (e.g., economic comparisons, engineering feasibility) and stakeholder involvement. Once a management alternative is selected and implemented, monitoring its effectiveness can ensure that biological goals are attained, and provides valuable feedback to the SI process. All of these important activities are outside the scope of the current document. However, accurate and defensible identification of the cause through the SI process is the key component that directs management efforts towards solutions that have the best chance of improving biological condition.

3.3 CADDIS

Presented by Susan Norton, U.S. EPA, ORD, NCEA

The CADDIS workshop's objective is to conceptualize a decision support system that will help investigators find, access, organize, apply and share information useful for causal evaluations in aquatic systems. In addition, critical research needs for system implementation and population will be identified.

CADDIS is needed because implementing the SI Guidance asks much of investigators. It requires wide knowledge about a variety of stressors as well as depth of knowledge on mechanism, symptoms, and stressor-response relationships for specific stressors. High-quality data from field and laboratory studies may not be readily at hand. Research results may need additional development to be directly useful for the process. Finally, documenting the process for presentation to others may be burdensome.

CADDIS will supplement the experiences of investigators by providing a guide to the SI process, facilitating access to relevant information and knowledge, providing tools to analyze and interpret evidence, and helping investigators organize, quantify, and share results (Figure 2). Close interactions with the research community will facilitate linkages to existing databases, construction of new databases, and the identification of research needs. Close interaction with the user community will facilitate the production of case studies, the application to new investigations, and the prioritization of system functions and research needs. CADDIS will help make investigations defensible, transparent, and repeatable. It will allow investigators to learn from each other by creating opportunities for users to submit information and examples and share experiences.

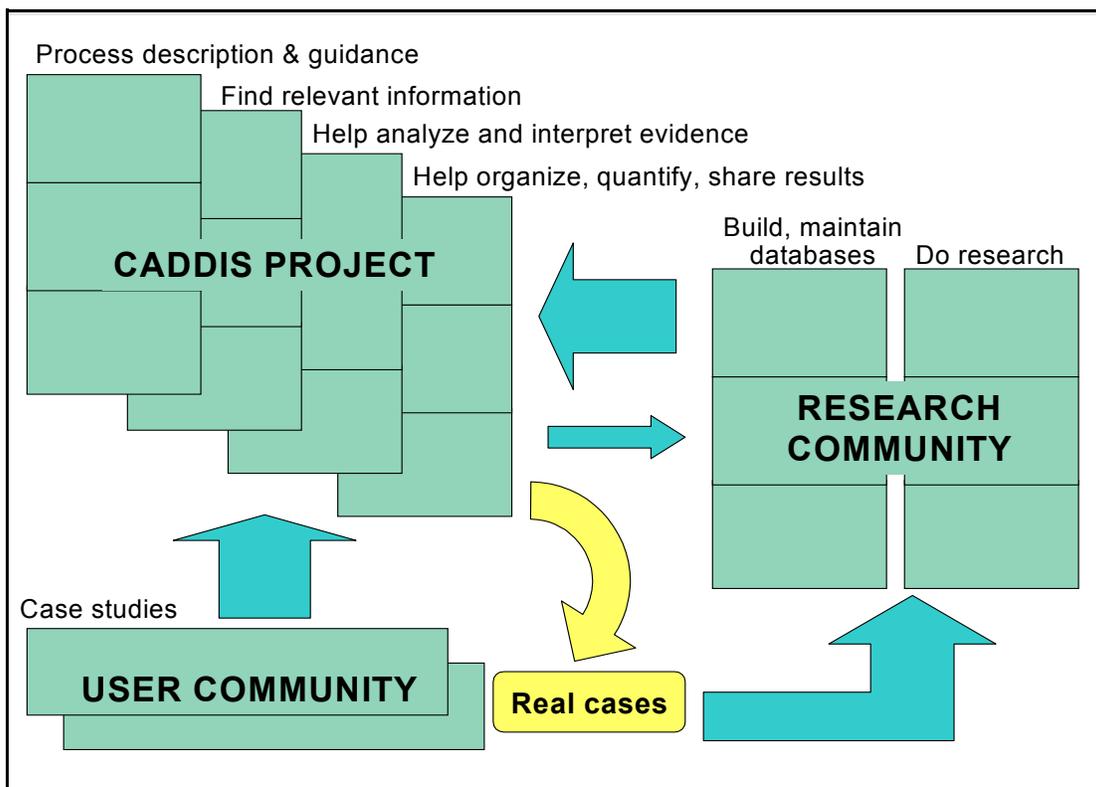


Figure 2: Conceptual Diagram of the CADDIS Project, Research and User Communities. The three stacked boxes under CADDIS Project and Research Community represent the three phases of the SI Process: list candidate causes, analyze evidence, and characterize causes.

3.4 DESIGN OPTIONS AND PROTOTYPE

Presented by Tom Stockton, Neptune and Company, Inc.

Another intent of the workshop is to provide input on the design of CADDIS. To facilitate workshop participation in CADDIS design, a white paper that describes three possible design options was distributed to workshop participants (Appendix B). These options provide examples of the types of material that could be included in CADDIS and were used to initiate discussion on these and other possibilities. The three options present visions of CADDIS at increasing levels of complexity.

The three CADDIS system design options, CADDIS¹, CADDIS² and CADDIS³, vary in complexity and resource levels needed to develop and implement. These options build on each other sequentially to define a resource-efficient path forward for developing CADDIS. The move to more complex approaches is in large part a move from streamlining access to information in a static environment towards an adaptive decision-making framework wherein uncertainties and decision consequences are taken into account.

CADDIS¹ is primarily a static system in which information is made available through web-based links, and the user is stepped through the SI process as described in the guidance. In CADDIS¹, SI worksheets can be completed externally following guidance provided by the system. CADDIS² builds on CADDIS¹ and includes options for analysis of SI data and completion of the SI process interactively. This is a more dynamic system in which analysis capabilities are included to allow synthesis of site-specific information with relevant database and reference information. CADDIS³ involves more dynamic functionality in terms of implementing the SI process. This option allows for building conceptual and numerical models and designing the numerical models to derive causal effects in a probabilistic setting.

A table of the functionality CADDIS could provide was also presented for discussion at the workshop (Appendix B-1). For each function or feature, this table indicates which of the three option levels applies, as well as the anticipated level of effort and resources needed for implementation.

A prototype web-based version of CADDIS was developed in association with the white paper so that the workshop audience could share some of the vision of what could be included in CADDIS. This is presented on a “localhost” web server. The look, feel, and technical approach for implementing the desired functionality are presented using this prototype. The prototype CADDIS takes the user through the SI guidance on a path that includes:

- Features and placeholders for all CADDIS¹ options (access to information sources)
- Examples of how CADDIS² could be implemented, including passage through the SI process, access to mapping and analytical support, and completion of the process of identifying causes for the stressors using the Little Scioto case study
- An overview of the possibilities for the CADDIS³ option

The prototype CADDIS website is intended as a tool for exploring how a user might be guided through the SI process and which technical tools are needed to implement CADDIS. This site is not intended to be a model for the aesthetics of CADDIS or to provide complete functionality. In general, the web tools used in this prototype are Open Source and freely available. The prototype serves as a point of reference and a point of departure for consideration in the workshop. Along with the web-based prototype, a table of the functionality CADDIS could provide is also presented as a point of departure for the workshop. For each function or feature, this table indicates which of the three option levels applies, as well as the anticipated level of effort and resources needed to implement the feature.

3.5 AN APPROACH FOR DEVELOPING INNOVATIVE SOFTWARE

Presented by Steve Fine, U.S. EPA, ORD, NERL

Typical software engineering approaches do not address the needs of innovative projects because the traditional techniques are often unsuitable for situations where requirements continue to evolve during the life of the project. U.S. EPA's Multimedia Integrated Modeling System (MIMS) is an example of such an innovative project. MIMS will include software to support the composition, configuration, application and evaluation of complex systems of multimedia models. How to do this is a complex research problem that requires experimentation and flexibility as new approaches are identified. To address these demands, MIMS is following several approaches. Principles of agile software development are being applied. This includes providing rapid feedback to software developers, using designs that are simple but leave future options open, making incremental changes, and building quality products. MIMS also applies several design techniques that simplify the process of incorporating new features, reuse good designs and software that others have developed and that support our objectives, and maintain a balanced effort with a focus on MIMS's customers' needs. The MIMS project has been successful in iteratively developing software currently used by a small group. Broader distribution of MIMS will occur this fall and provide additional feedback on the value of this

development approach.

4. BREAK-OUT GROUP REPORTS

This section discusses the results from the three break-out groups. Each of the three groups worked through three sessions. For the case study groups, the sessions corresponded with the steps of the SI process. The third group focused on different system development issues during each of the sessions. Charges to the break-out groups are also presented, although the session summaries do not necessarily follow the questions since each group approached the charges differently.

4.1 BREAK-OUT GROUPS A AND B

Groups A and B focused on case studies for the Little Scioto River, OH; and the Cabin John Watershed, MD respectively. These case studies were intended to offer food for thought to help the break-out groups identify how a diagnostic system such as CADDIS could help in the SI process. The break-out groups identified key functionalities for each of the three main steps in the SI process. Specifically, the questions Groups A and B were charged to answer in each session were:

- Of the possible functions for CADDIS, which ones are highest priority and why?
- Are there other CADDIS functions?
- What information would you like to access that is currently hard to find?
- Is there information/data that would be useful but does not exist or is one or two steps away from being directly usable?

The recommendations from the two groups had much in common; therefore, their recommendations have been combined and are synthesized by session next.

4.1.1 Session 1: Defining the Impairment and Listing Candidate Causes

The first break-out session of the workshop focused on the SI phase of defining (refining) the impairment and listing candidate causes. The four steps to this phase of SI are:

1. Mapping the situation
2. Refining the description of the impairment

3. Listing the candidate causes developed from data collated for the case
4. Developing a CM of the situation

The assistance envisioned from CADDIS for all steps falls into three categories:

1. Educating the user
2. Providing foundational information for the step
3. Providing tips, or a “wizard,” to prompt and guide the user through the process and highlight considerations that the user should make during that step

Step 1. Map the situation

Functionalities for this step are:

- Education
 - Identify the kinds of things a user would want to include on a map, e.g., National Pollutant Discharge Elimination System sites, gaging stations, industrial plants
- Foundational information
 - Provide links to resources (if they exist) that can provide those kinds of data

NOTE: The groups felt that there should not be a mapping capability in CADDIS itself; instead, the system should have the ability to upload user-generated maps developed off-line or available from external sources (e.g., EnviroMapper). However, some need was indicated for creating figures that depict the site (e.g., cartoon figures, line drawings, thumbnail sketches), even if mapping was not provided directly in CADDIS. (This is also, in part, a CM issue.)

- Tips/“wizard”
 - Provide prompts when maps might be useful in the SI process or be helpful for documenting implementation in CADDIS

Step 2. Refine the description of the impairment

Functionalities for this step are:

- Education
 - Provide a description of what an index of biotic integrity is, as well as which metrics and thresholds have been used (recognize that this might vary by State or county and the problems this would create with consistency)
 - Provide a description of stressor-response relationships
- Foundational information (e.g., taxa lists and associated categorizations, tolerant/intolerant, feeding guilds)
 - Provide the ability for the user to modify the databases given by the system to customize them for their own situation/place
 - Identify particular suites of effects that indicate particular categories of change (e.g., nutrient cycle, energy, diversity)
- Tips/“wizard”
 - Provide prompts through the thought process on how/what to refine in stating the biological effect, such as appropriate metrics that might be indicative of particular stressors, ways to disaggregate from index to indicators, or maybe, how to look at the raw data to best describe biological effect
 - Help the user define the scale of the problem (e.g., pervasive over a large area or point source issue, stream or watershed). Impairment could be caused by either local or global issues (e.g., air deposition)

Step 3. List candidate causes developed from available information from the case, from the region, and from conferring with stakeholders

Functionalities for this step are:

- Education
 - Provide a description of stressor-response relationships
- Foundational information (searchable “library”)
 - Provide a compendium of known source-stressor relationships
 - Provide a compendium of known stressor-effect relationships

- Provide a compendium of known thresholds (e.g., temperature tolerances for various fish species)
- Provide links to already existing databases that provide information above
- Provide a compendium/database of case studies that can be searched for similar situations to those that the user is evaluating
- Provide a “glossary” or two-page synthesis of knowledge about each stressor
- Tips/“wizard”
 - Provide prompts through what to consider, potential causes for particular stressors or sources
 - Provide a way for users to list their sources, then have the system list the typical stressors for each source (pick list); allow the users to then modify for their own situations
 - Identify logical ways to shorten the list to what is possible for user’s situation (e.g., is a particular kind of source not there?)
 - Provide a reminder (or maybe track through CADDIS) to document rationale for taking something off the list
 - Provide a way to extract new relationships (source-stressor or stressor-effect) from the case studies at hand

Step 4. Draw CM

Functionalities for this step are:

- Education
 - Provide a description of what a CM is and how to use it; provide example CMs
- Foundational information
 - Provide a database of “standard” CMs for typical situations (modifiable to allow customization for a user’s particular situation)
 - Include current land use in the CM
- Tips/“wizard”
 - Possibly provide a graphics tool to allow the user to pick source or stressor, tool

draws and connects

- Provide a way to click on an individual box and get underlying data if user has loaded it in
- Provide the ability to nest, or develop hierarchical models (e.g., start with a general relationship, click to more detailed model that shows more specific interactions)

Group A talked about using the term “change” instead of “impairment” to reflect the fact that what SI really measures is whether the system being investigated is undergoing change over time. This is what leads to change in stream-measured values (e.g., indices) and causes streams to become listed.

In general, flexibility should be provided for developing impairment and cause lists. It would be helpful if common associations were made available within CADDIS, but user-defined associations need to be accommodated as well. Metrics or measures that the user chooses must also be accommodated. Some discussants thought that categorization of stressor-response relationships might help the user sift through possible choices or options (e.g., biodiversity, energy, nutrients, metals, toxic chemicals, habitat alteration, point sources, non-point sources). The participants wanted long, fixed (common) lists as a starting point, but the ability to have users expand and add to lists as well, and wanted to be able to link impairments and causes.

Following this discussion was one about how to pare down the lists in response to implementation of the rest of CADDIS. The lists and relationships included in the foundations above could be used to define impairments and causes. They could be used initially to generate a broad-based list of stressor-response pairs, which could be managed (reduced or expanded) in response to progress through the SI process; an elimination step will result in change to the lists. There was recognition that the process is iterative, and so the active lists will have to be revised in response to implementation within CADDIS. In general, this revision is done by the user. For example, in the Little Scioto case, a lot of work and options were considered prior to the presentation in the papers. For another site, working through the process did not locate the cause initially, but another potential cause was added after the SI process failed to identify a cause at the outset (Dye Facility). Flexibility within the system is needed to accommodate these changes while the process is being implemented. Some concern was expressed that if lists were provided, users could be anchored to those lists and would not define their own site-specific items. However, it was felt that broad lists would be preferable.

Also, once list items are activated, the foundation is then laid for CM building, in which

case CADDIS could generate CM diagrams linked directly to stressor-response pairs included in the application. If CM diagrams are developed, then these diagrams should be updated with the change, and the record of how and why the changes were made should be tracked. Creation of CMs, graphically or otherwise, was considered very important to the process, since the CM describes the process. Tracking can be used to document changes in thinking in response to new information and evidence.

Some mention was also made of understanding the incremental cost-benefit from any action that might be taken. For example, it may be more effective to take action on a less impacted stream because recovery would be greater than to take a similar action on a more impacted stream that may not recover as well. This point was not considered beyond the initial discussion because it was felt that this came under the jurisdiction of management decisions, which is not within the scope of CADDIS at this time.

In general, there was a sense that professional judgment could not be replaced by a computer, and that CADDIS should focus on providing information, some guidance through the SI process, and capability to implement the process by capturing the professional judgments of the user, but not by making the judgments (e.g., through statistical analysis). In other words, CADDIS should support the process rather than trying to perform it. (The user must answer the site question, not CADDIS; CADDIS is simply a repository for supporting the user's judgments.)

Following this discussion, group members considered ways to pare down the lists in order to implement the rest of CADDIS.

1. The break-out groups did not envision that CADDIS would provide analytical tools and be a place to do on-line analysis of data, although Group A did believe that it might be useful to provide the ability for the user to present data graphically (spatially or temporally), perform exploratory data analysis, and calculate some basic descriptive statistics within the system (e.g., bar charts that were presented in the Little Scioto example). There was recognition that statistical analysis is used to short-list impairments based on site data, but it was felt that this capability did not have to be provided within CADDIS. Instead, CADDIS could simply be a repository for the results of those analyses (e.g., short-list of impairments). The break-out groups also agreed that CADDIS should not be envisioned as a place to go to get raw data for a particular case; the users would bring their own data to the process. Analyses of those data would best be conducted off-line, with the capability to upload the results of those analyses into the system for documentation of the process.
2. The break-out groups felt one of the strongest benefits CADDIS could bring to the SI process would be to serve as a knowledge base to add value to the user's evaluation by (1) synthesizing stressor information and providing it in one place (as opposed to

having to search the Internet), and (2) providing a repository of case studies that provide knowledge of what everyone else has had problems with and which solutions have been effective.

3. However, Groups A and B strongly felt that the final needs and requirements for the system should be defined by the intended users. This leads to the question of who is the intended audience. Once decided, those users should be contacted to see what they feel would be useful. There was some discussion in Group A of whether CADDIS should be a computer-based system or if the SI process is sufficiently well defined and should be used more as a paper system before proceeding with development of a computer system.

4.1.2 Session 2: Analyzing the Evidence

This phase of the SI process analyzes the available data from the site and elsewhere to generate evidence concerning potential causation. The two case study groups considered the support that they would like from CADDIS for this phase. The consensus of the two groups, and a few areas of differences, are described in this section. The discussions are best summarized in terms of the priorities for desired features.

Walk-through with Prompts

A fundamental feature CADDIS must have is the capability to lead the user through the process and prompt the investigator to provide information, make judgements, or determine the relevance of a step or consideration. In this phase of SI, this feature could be best performed by providing a table of types of associations that the user would fill in with relevant evidence. In general, it was felt that CADDIS also should provide an indication of where the user is in the process and that the context of the work be understood.

Information

There was strong agreement that CADDIS users would want access to outside information concerning potential causal agents. This would include: (1) stressor-response information, (2) diagnostic responses including biomarkers, (3) characteristics of the agents such as degradation rates and other physical/chemical properties, (4) reference or background levels or characteristics of the agents, and (5) results of remediation or restoration projects. In addition, some biological information is needed concerning the impaired organisms, populations, or communities, such as spawning habits and seasons or trophic characteristics. The information may be provided outside the system, and CADDIS may simply provide web links or references to published reviews.

CADDIS may provide a useful context-based search engine. It could also provide a bibliographic database that would save the exposure-response and other information that the user had accumulated and the related bibliographic sources. However, there was a consensus of both groups that it would be very useful for CADDIS to generate and provide summaries of relevant exposure-response and diagnostic information for common stressors and classes of stressors. Finally, once CADDIS is in use, it will be important to provide a means for information sharing among the users. This might be a separate listserv, the capability might be built into CADDIS, or other means may be identified.

Expert Guidance

The concepts and terminology of CADDIS may be unclear to beginning users. Therefore, CADDIS should, at minimum, provide links to definitions and explanations of individual terms and concepts and even entire steps in the process. A later addition to the system would be suggestions for how best to perform the analyses or inferences. These might include, rules of thumb (e.g., how to relate temporal patterns of exposure in the field to laboratory test durations) or built-in warnings about using the same evidence twice. They might also include examples of analyses and inferences taken from case studies or generated for this purpose. User feedback can guide the development of this feature.

Report Generation

At a minimum, CADDIS should save the results of work performed by the user in a format that can be reviewed, revised, and printed. For this step in the process, it would be valuable to generate tables of the evidence generated for each type of association. More advanced versions of this feature would generate reports that would fully document and explain the process so that it could be understood by managers and individuals other than the users.

Data Management and Analysis

This feature inspired considerable differences of opinion. Some participants believed it would be useful to incorporate simple data management and analysis capabilities, particularly for the types of analyses recommended by the expert guidance. Others felt that users already have data management and analysis capabilities that they are familiar with and that are suited to their data sources. Those users might benefit little from data management and analysis capabilities in CADDIS. Based on user feedback, more advanced analytical tools might be included. These might be built into the system, or CADDIS might provide macros or formulas for use in common

statistical platforms.

Quality Assurance

This feature also caused differences of opinion. Because CADDIS users will depend primarily on existing data, they have little control over data quality and cannot take advantage of conventional data QA approaches. Hence, the data QA is performed by the data generators, and the CADDIS user should document that by importing the relevant meta data along with the primary data. However, it may be important to evaluate and track the quality of the analyses and inferences used in the causal analysis itself. For example, one may identify a spatial trend based on 3 points or 30. Hence, several participants felt that it would be important to include a feature that would prompt the user to evaluate the uncertainty or quality of an analysis or inference at each step in the process and carry that information into the report.

Case Studies

These might include rules of thumb (e.g., how to relate temporal patterns such as exposure in the field to laboratory test durations), or built-in warnings about using the same evidence twice.

Link Forward

It may be useful to link forward from the analysis of evidence to the characterization phase. In particular, some people felt that they are likely to see opportunities for elimination of a candidate cause during this phase. They would like to jump forward to the elimination step, eliminate that cause, and then return to the analysis phase.

Link Back to the Conceptual Model

Participants thought that it might be useful to be able to easily go back and modify the CM during this phase. As the data are analyzed, the user is likely to gain an understanding of the impaired ecosystem that could result in revision of the CM. Also, the CM could be updated as candidate causes are eliminated or added, based on the analysis and subsequent characterization, so as to reflect progress in the causal analysis.

The minimal CADDIS capabilities for the analysis of evidence must include a walk-through of the process, with some ability to define or explain terms and concepts, and saving and tracking results. It was felt that users will want some information very early in the development of CADDIS, particularly related to exposure-response. Other features may be added in later versions, based on user feedback.

4.1.3 Session 3: Characterizing the Causes

Eliminating Impossible Causes

During the analysis phase of the investigation, associations are made using evidence collected at the study site. These associations can now be used to eliminate impossible causes. CADDIS can help by creating a form having all candidate causes depicted across one axis and the specific impairments along the other axis. Icons could appear that allow the user to seek a description of elimination, clarification of associations using texts, examples, a diagram or cartoon, and pitfalls to avoid.

The questions and answers developed in the analysis phase should be available for review by both the user and the investigator. The system could guide the user through each elimination step or allow a quick mechanism to highlight those causes that should be eliminated. Another option could allow the investigator to chose an automated process that eliminates all causes that have been identified as impossible causes of the impairment.

Space should be provided for caveats, such as why a candidate cause is not eliminated even though there is a “no” entered into the analysis. CADDIS could provide tips and options for deciding if and when a concentration is “really” elevated or not from one location or time to another, as well as tips on interpreting correlations to facilitate evaluation of biological gradients.

The table created for the elimination of causes should then show those causes that will continue to the diagnosis and strength-of-evidence characterizations. It would be handy for the table to provide a reference to the appropriate association developed in the analysis portion of the documentation. Another way of tracking the process would be by showing a graphic of the pathway used to eliminate the candidate cause.

The elimination step should occur prior to the collection of evidence using observations from the study area and from the literature; associations with cleanly eliminated candidate causes may not be needed, and this information would make it easier for CADDIS to create a table for the strength-of-evidence arguments.

Using Symptomology to Diagnose a Probable Cause

Observations and measurements about the specific impairment at the study location may permit the investigator to use symptomology to diagnose a cause. Symptomology collected from the published literature assembled in the analysis phase could be made available to the investigator for easy viewing. This would probably include a table of symptoms and an indication that each symptom was or was not present and weights, if appropriate.

CADDIS could provide a location for documenting the characterization and rationale for diagnosing the cause of an impairment. A graphic of the decision points could help to lead the investigator through this step; it would document the process that the investigator used to either diagnose or proceed to the strength-of-evidence analysis.

CADDIS should also provide space for composing text to support the tables. The team has not fully explored the best way to analyze and present evidence for diagnosis. Options for providing a realistic and complete example are needed.

Organizing “Strength-of-evidence” Arguments

During the analysis phase, associations were made using evidence gathered from the literature and by comparing them with measurements and observations taken at the study area. These associations could then be used to weigh various considerations within the strength-of-evidence arguments. CADDIS could facilitate this process by providing information to help the user through the process. Help should be optional so that experienced investigators can proceed to completing the characterization of causes more rapidly. One option would be to show icons that allow the user to seek a description of the strength-of-evidence process, including all considerations, examples, and diagrams or cartoons, as well as pitfalls to avoid.

To characterize each consideration, for all the impairments and candidate causes, CADDIS could return the investigator to a screen that shows a consideration, the evidence, impairment, and the candidate cause. There is a particular need to enable the investigator to become comfortable with scoring. Participants commented that the table in the SI Guidance did not provide enough assistance.

CADDIS could provide a list of options appropriate for each consideration, and the investigator could choose from among them and automatically score the consideration. The score should then be automatically recorded in a summary table and a complete strength-of-evidence table (including a reference for each piece of evidence); the investigator can then compare all the evidence for a particular candidate cause or for a consideration among multiple candidate causes. An option should be provided to use similar previous cases as templates or to pull in a column of evidence from the strength-of-evidence table for a particular candidate cause.

CADDIS can help by creating either a summary form for each impairment (that shows all the candidate causes depicted across one axis, and the considerations along the second axis), or a similar form with the specific impairment information repeated below each consideration.

The interest in creating forms and tables elicited discussions of the challenges for coding the software. The coding had to be flexible enough to accommodate variation among studies

with different numbers of candidate causes and impairments.

Generating the Final Report

Some suggestions for basic report elements included:

- Chapters for each section of the SI process, including an introduction describing the regulatory or other impetus for undertaking the causal analysis
- Various products such as tables, lists, CMs, bibliographies, and text should appear in the appropriate sections
- The template could be exported to the user's preferred platform for word processing or presentation software.

Research Needs

Two situations were identified where further consideration is needed concerning the availability of data, guidance on how to address the situation, and modifications to CADDIS to accommodate these situations. The first was the scenario wherein multiple stressors could act synergistically to cause an impairment.

The second was the need for flexibility in characterizing causes at different spatial scales. For example, the impairments and/or causes could vary for different parts of a river or components within a watershed. How CADDIS could accommodate analysis of various impairments for different waterbodies and scales was discussed.

There also was discussion about the potential for a more advanced scoring process that could include independent scoring developed by individual users. The process could accommodate user uncertainty through some form of weighting process, as well as scoring among candidate causes for each consideration.

Other Comments

There were many reminders that causal analysis is an iterative process. One situation was described to illustrate this point. If an investigator was gathering information from the literature and discovered that the impairment could be caused by a new candidate cause, then it might be added to the list of candidate causes and the CM, and require analysis and characterization.

Because of the complexity of performing both analysis and characterization, it seemed important to remind investigators to seek informal consultation with many colleagues with

diverse expertise. This was considered absolutely essential if, after completing the strength-of-evidence process, no cause had sufficient evidence.

Throughout the causal analysis, a navigator would be useful to show the investigator where he or she is in the process and which resources and work already done might be useful for that activity. This would permit the user to move to that resource or to another part of the process. Appearance could be *via* a list of completed steps or more complex processes. This is especially important during the characterization step, because it is necessary to continually refer to previous work.

Almost all of the functions listed in Appendix B-1 of the Design Options Paper were useful and often essential; however, some uncertainty existed concerning the definition and, therefore the usefulness of the modeling and weighting wizards for the strength-of-evidence analysis, as well as having the probabilistic modeling wizard for evaluating uncertainties.

There needs to be a balance between being instructional and having a place to create a document. Generally, CADDIS should prompt users for information in a step-wise fashion and record user responses to document decisions in the causal analysis. For cases where no evidence was generated in the analysis phase, CADDIS could inform the investigator that there is no reason to evaluate that consideration or could prompt the user to seek additional information.

4.2 BREAK-OUT GROUP C

Break-out Group C was tasked with addressing system development issues from the perspective of the potential users in three different sessions. The general topics discussed in each session included: (1) the issues surrounding system development, (2) QA requirements and issues and (3) interaction with and among users.

4.2.1 Session 1: Issues Surrounding System Development

In Session 1, the group addressed the following specific questions:

- Who are the potential users of the CADDIS system?
- How will the users use the system?
- Who might want to independently modify the code?
- How and when should users be involved in the development process. How do we make this happen?

- What are the pros and cons associated with different options for the system's long-term home?
- What design and implementation strategy would maximize the likelihood of CADDIS becoming a long-lived system?

The results of these discussions are summarized next.

Potential Users of CADDIS

Group C brainstormed a list of potential users of the system. It was suggested that by identifying as many potential users as possible, marketing the system would be easier and more effective. It was also suggested that the better the developers understood the users, the better the system requirements could be defined. However, one caution was noted that by trying to address the needs of all potential users, the design of the system might be overwhelmed by a large number of diverse needs. Group C identified the following potential users:

- States: Biologists
- Tribes: Biologists
- Water Quality Managers
- Local Governments: Stormwater Managers; Utilities Personnel; Biologists; Planners
- Consultants
- Citizen/Watersheds Groups
- U.S. EPA Regions
- Other Federal Agencies
- U.S. EPA, ORD
- U. S. EPA, OW
- External Stakeholders, including potential sources of stressors
- Enforcement Personnel
- Educators: Universities (students learn the process and then get hired by States/local

governments)

From this list, Group C identified that the primary users would be States, at least initially. Local governments may also be primary users, along with certain stakeholders who might be identified as the source of the stressors causing an impairment in a waterbody.

Use of the System

Major uses of the system identified by Group C included 303(d) listing of impaired waterbodies and the development of TMDLs (Figure 3), where causes needed to be confirmed and sources needed to be identified. Both States and U.S. EPA would use the system in this way. U.S. EPA Regional Offices may also use CADDIS for checking State lists of impaired waterbodies in developing TMDLs or in TMDL litigation. States might also use the system to confirm previously identified causes of biological impairment (using best professional judgement or other techniques) to make their determinations more legally defensible. CADDIS may be used to provide added value for current systems [STORET, BASINS] and to improve cost-effectiveness in SI. It was also suggested by the group that CADDIS could be a valuable training tool to enhance consistency in SI.

Modification of the System's Code

There was general agreement that the potential users listed in this section would probably not need to modify the code. However, it will be essential for the users to be able to add or change the data in the system. Given this, it was suggested by the group that data input and format be flexible for different users and sources/types/formats of information. The group also suggested that the system be modular to maintain flexibility and that adequate documentation should be available to provide confidence in the TMDL decisions when modifications to the system are made by users.

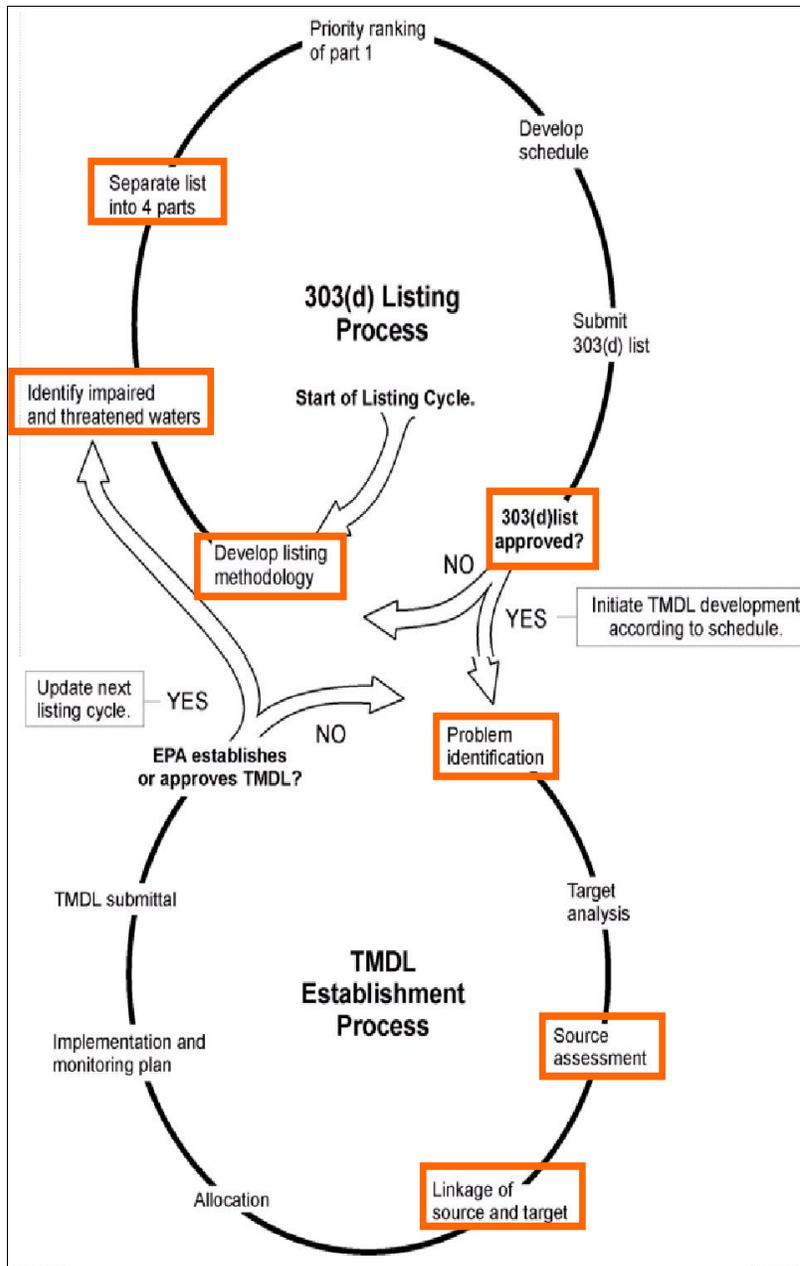


Figure 3: Where CADDIS may aid the processes of impaired waters 303(d) listing and TMDL development. Boxes around steps indicate where CADDIS might be used.

A discussion of the design of the system also ensued. However, a number of members in the group suggested that it was premature to ask design questions prior to getting enough input from the potential users. The group suggested that a series of questions could be posed as a user requirements study to help focus user input on design. The questions that were suggested are:

- What is the actual process you use when assessing an impaired stream?
- What are the top five things the computer can do for you to make your job easier?
- At what point in the process will you need to use the system?
- What would help you to get through the TMDL process?
- If you were to use the SI process, where would you need help?
- What could we provide that would make you want to use this process?
- What are the minimum features that CADDIS should have to make you want to use the process? (e.g., searchable relevant literature)
- What are the current weaknesses of SI that you think need to be addressed?
- What kind of training/educational component would be needed to use this system?

User Involvement in the Development Process

The group had a fairly long discussion of this important issue. One general agreement that emerged was that users need to be factored into the development process very early on, or right up front, and to be involved often. Some members of the group seemed to agree that as soon as a preliminary prototype was developed, it should be given to users to get their reaction and feedback. Other members of Group C suggested that users should even be involved in the development of a preliminary prototype. Some members of the group also suggested it would be good to have user(s) on the development team and not just programmers and U.S. EPA project managers. It was expressed that users need to invest the time in development to have the payoff of acceptance in the end. One member offered that CADDIS should be a system created by users for users. A few members of the group suggested that CADDIS should come out with something quickly that has value to users (core functions), and iterate system development from there.

The group also recommended that there needs to be good communication between the users and the developers – communication goes both ways. For good communication,

developers of the system should have a prioritized list of requirements collected from the user community. Developers should also have a schedule that shows that the list of user requirements will be addressed; functions on the list will not be ignored. The group also suggested that system developers need to give the users feedback on the time/cost needed for the requirements. This may change the prioritization for the system development.

Group C also recognized it was important to keep U.S. EPA senior management involved and informed about system development. It was suggested that CADDIS development team should think big, but work carefully in small steps and show the progress. Continual funding of CADDIS will be needed to maintain the system and support any enhancements. The development team could use demonstrated user involvement, such as lists of features and enhancements requested by users, to help justify a plan for incremental funding to develop and improve the system. CADDIS needs to demonstrate careful planning and flexibility for investment, one that is reasonable and affordable over time.

The group next discussed how to engage users. The group agreed that to best engage users, contact needs to be face to face, at least initially. Later contact can be informal by e-mail, web-based feedback, or discussion groups. There could also be monthly or bi-weekly meetings of an “architecture group” – developers and substantially involved users. It was pointed out that the users that are involved in development are those that want to get involved and have the time to commit to the process. Since not all users may have time to be involved, there need to be different levels of involvement by users; give users a choice of the level of his/her involvement.

The group suggested that there should be initial meetings with the user base: go out to the Regions (Regions can invite State representatives). The group suggested having a workshop, or series of workshops with users. One member suggested that U.S. EPA Regional buy-in is very important for encouraging State use.

The group next discussed the Development Team. The group suggested the team should consist of:

- Programmers
- Project Managers
- Line Management
- User Group (tiered)
 - One person intimately involved
 - Others less involved, maybe just from time to time

A Long-term Home for the System

Several options were identified for the long-term home of the CADDIS system. This was identified as a key question that would impact both the system design and the functionality that could be built into CADDIS. The first option is that CADDIS could be developed as a local application that resides on the user's computer. A local implementation of CADDIS could assume either an Internet connection or no Internet connection. Alternatively, CADDIS could be developed as a web-based application that resides wholly on a web server (e.g., the U.S. EPA National Computing Center in RTP, NC). An intermediate option, web-based with applets, was also discussed. This was considered intermediate since an applet is a Java program that may be downloaded over a network connection and run on a local machine via a web browser. The issues involved in deciding between a web-based application versus a local application are outlined in Table 1. The group generally felt that the system needed to rely on Internet access and that avoiding applets is desirable. No clear consensus on the appropriate path forward was identified by the group on whether to proceed with a web-based or local application.

The U.S. EPA and a non-government organization (NGO) were discussed as two possibilities for maintaining a web-based implementation of CADDIS. Advantages of a U.S. EPA-based home for CADDIS included U.S. EPA maintaining control over CADDIS and the credibility lent to CADDIS by the epa.gov domain name. Advantages of a professional organization hosting CADDIS included increased flexibility and functionality allowed by not having to meet U.S. EPA web specifications for security, accessibility and presentation, as well as the possibly of a built-in peer review mechanism for case studies and expert system rules. Also of concern was identifying a group within U.S. EPA or an organization outside of U.S. EPA that could ensure long-term funding of the system.

Table 1. Issues to consider when choosing between a web-based system and a local application.

Issues	Web only	Web with applets	Local application with web access	Local application without web access
Web Access	Application can be accessed by anyone, anywhere. Access could be disrupted by U.S. EPA policy, funding decisions, server problems.	Performance will be significantly impacted by access speed.	High-speed access will be needed if large remote datasets are accessed.	No access needed but removes linking to information resources, limiting functionality.
Deployment	Deployment to U.S. EPA server requires U.S. EPA National Computing Center Review that can hold up release and will require a security plan. Deployment will also require an office to take full responsibility for the website/ application.	Same as web but requires testing to include a variety of desktops to test performance of applets.	Could deploy <i>via</i> the web and/or compact disc (CD).	Have to deploy by CD. Requires maintaining mailing lists, and entails postage and mailing costs.
Updating	Automatic. Users never see updating process. Updates are universal. No problems with version control.	Same as Web-Only.	Automatic. Users likely involved with updating their programs.	Mail out CDs to users.
Maintenance	Keep up with browser changes and web specifications. Information/application easily backed-up and recovered by computer support center.	Keep up with browser changes and web specifications.	Keep up with browser changes, web specifications, and operating system (OS) changes.	Keep up with OS changes
Analytical Power	Less dependent on user's computer, more on server performance and number of concurrent users. If the application is running on shared server, it may compete for processing power with other applications.	Dependent on server performance and number of concurrent users.	Dependent on user's computer.	Dependent on user's computer.
Internet Security	U.S. EPA security requirements may limit functionality.	U.S. EPA security requirements may limit functionality	Should not be an issue.	Should not be an issue.
Capacity	Dependent on server. If limiting, could be improved with more CADDIS resources.	Dependent on server and user's system. If server is limiting, could be improved with more CADDIS resources.	User-dependent.	User-dependent.
Compatibility	Browser issues.	Browser issues.	Browser and OS issues.	OS issues.
Resources	Requires resources for computational power and data storage on a CADDIS server. The user does not store the application locally, saving disk space.	Requires resources for computational power and data storage on a CADDIS server.	Requires server resources for updating.	Resources for CD distribution.

Table 1. (Continued)

Issue	Web only	Web with applets	Local application with web access	Local application without web access
Compliance	The application must comply with the Agency standard web templates; web functions permitted in the U.S. EPA environment and with Agency security standards.	Same as Web only.	Fewer Agency requirements.	Fewer Agency requirements.
Control	User has less control over data. Version changes may limit reproducibility of results.	User has less control over data.. Version changes may limit reproducibility of results.	User has control over data.	User has control over data and software.
User Support	Browser issues.	Browser, version and OS issues.	Version and OS issues.	Version and OS issues.
Training	Incorporate into website.	Incorporate into website.	Incorporate into website.	CD based training, not easily updated.
Capabilities	More difficult to incorporate existing analytical tools. More difficult to create complex, interactive graphical user interfaces.	More difficult to incorporate existing analytical tools.	Easier to incorporate existing analytical tools.	Easier to incorporate existing analytical tools; no access to information sources on the web.
Customization	Difficult. Users would submit suggestions to CADDIS project for incorporation.	Difficult. Users would submit suggestions back to CADDIS project for incorporation.	No limitations.	No limitations.
Licensing	Varies. If the application uses a commercial database or other commercial software, the owner must purchase an “unlimited” license for web use. If simple and non-commercial software is used, there is no additional cost.	Same as web only.	Commercial software may have a distribution fee.	Commercial software may have a distribution fee.

Design and Implementation Strategy for a Long-lived System

Group C came to agreement that CADDIS would be a long-term commitment for both developers and users. As such, commitments from the developers (U.S. EPA) and from the entity operating and maintaining the system (possibly U.S. EPA or an NGO) would need to be long term; otherwise, the system would not survive. The group identified that long-term financial support would be critical to long-term survival. The group also recognized that long-term buy-in by U.S. EPA management to fund the development and operation of the system

would be necessary. The group recommended that U.S. EPA management be kept informed of the system development, products, uses, financial needs and user response. Briefings should be provided to U.S. EPA management in the short term to justify support for the development of the system, and commitments must be made to keep management informed over the long term. Clear milestone products and successes would help over the long term to maintain U.S. EPA commitment to the system.

From the user perspective, the group concluded that users need to find the system helpful, user-friendly, and available, in timely fashion, to be a long-term success. From this standpoint, the group felt that during development, the system should release the most useful features first and soon. The most useful features could be determined from the user input at the outset of system design. The group thought there would be huge implications regarding user longevity depending on the content of the early released features, whether they be literature, data or tools. If users get system components quickly and are satisfied with their usefulness and capability, this will help to ensure their long-term acceptance of the system. Some members of the group suggested that a modular design of the system would separate the user interface from the databases and from other pieces, making the design more flexible and more able to change over the long term. The group also suggested that if a web-based design is used for the system, it would be easier to implement a hierarchical design that might serve a larger user base.

4.2.2 Session 2: QA Requirements and Issues

During the second session, Group C was charged with answering four questions:

1. What models are available for dealing with QA issues in a system like this?
2. Which QA issues are most important to the various user groups?
3. How do we strike the balance between access to all relevant information versus access to high-quality information? Is the balance different for various user groups?
4. What are the experiences of the STORET, WATERS, or TMDL tracking system developers?

When considering these charge questions, members of the group noted that CADDIS QA requirements will depend on what components eventually comprise the system. Group C identified all the potential CADDIS components and then discussed their highly variable potential QA needs. These components included:

- System
 - Design
 - Programming
 - Rule base

- Literature
 - Refereed literature
 - Gray literature

- Data
 - Data within CADDIS
 - Data referred to by CADDIS
 - Data input to CADDIS by users

- Tools
 - External
 - Internal

Members of Group C were able to offer insights on QA requirements and potential issues based on experience with U.S. EPA and non-U.S. EPA information systems such as WATERS, ECOTOX, MIMS, STORET, BASINS, and American Society of Civil Engineers' Best Management Practices effectiveness database. This report section provides the first steps toward outlining a CADDIS QAPP, discusses issues affecting each CADDIS component, and concludes by offering general principles that may help guide CADDIS QA in a generic sense.

System QA Considerations

Two major categories of system QA considerations were discussed: QA of the program logic (i.e., the basic system design), and QA of the rule base. Basic system design QA is certain to dominate CADDIS QA early on. Customarily, peer review is advisable for the system design document and any related prototypes. System programming QA requirements will vary with the functionality and tools that eventually comprise CADDIS. These may be incremental, and, thus, compatibility with pre-existing CADDIS components and functions will be an important part. The group advised the development of automated unit tests and system tests that could mechanize some of the QA on an ongoing basis. Configuration management, or how components interrelate, would also need to be addressed. Paired programming was mentioned as one development technique that integrates some of the QA at the programming stage by assigning two programmers to write and continually review designs and code as they are created.

The second category of system QA involves the rule base, or decision logic, that drives the expert systems, if any, utilized by CADDIS. This may rank among the most complicated QA tasks, both in peer reviewing the causal analysis assumptions and decisions (see also “Tools” below) and in assessing program functionality. Use of existing expert system shells may help the latter.

Literature QA Considerations

One major component of CADDIS may be causal analysis literature search and retrieval. For the most part, the QA responsibilities lie with the originators of this information, and CADDIS QA role is more in characterizing than performing the QA. Literature QA considerations affect inclusion/rejection of general categories of literature as well as individual documents. Also, future decisions about how CADDIS handles its literature functions (e.g., citation retrieval only; external key words; CADDIS-assigned key words; external search engine; annotated bibliographies, etc.) will have additional QA implications that would fall among U.S. EPA’s QA responsibilities.

As a rule, peer-reviewed sources will likely be the core of CADDIS literature, but numerous gray literature sources were also considered valuable by Group C due to their often greater detail, earlier availability and difficulty of obtaining. The group noted that peer review and publication are no guarantee of relevance or quality if the causal analysis elements of the study were not specifically QA’ed. Gray literature, on the other hand, despite lacking refereed publication, may still contain well QA’ed causal information. In short, general guidelines for literature quality are advisable, but some rules for characterizing QA appear necessary as well.

Data QA Considerations

Data and literature represent more of a gradient than two distinct categories, but Group C did separately consider data QA for information ranging from unpublished case studies and assessments through raw data measurements. As a rule, raw data was considered beyond the scope of CADDIS and probably unusable, but use of case studies, assessment results, and such materials as stressor-response relationships will likely appear in CADDIS. The group addressed the general needs associated with data and studies within CADDIS, referred to by CADDIS, and newly input by users.

Data encompassed within CADDIS could include U.S. EPA-designed case studies (for demonstrating CADDIS), externally designed case studies adapted for CADDIS demonstration, and any test cases run to assess system performance. Significant QA responsibility accompanies this category. The QA metadata for these kinds of studies should be provided. External studies

might require different levels of QA effort depending on how they are used as demonstrations of the system, or in relation to the rule base; in any case, the QA procedures followed in all case studies internal to CADDIS should be reviewed and made available. Data referred to by CADDIS should meet a stated QA standard, and users should be able to access information on what the QA of a given data set is. QA considerations are similar to those discussed under the literature section, above; again the QA burden lies primarily with the originator, with CADDIS responsible for characterizing its QA. Requirements are less clear for data input by users due to the variety of possible circumstances CADDIS may encounter. For the time being, Group C assumed that locally input and locally used data QA is the full responsibility of the user but an increasing level of QA verification and characterization would become necessary where input data become accessible to other CADDIS users or may become part of the CADDIS knowledge base and decision logic.

Tools QA Considerations

As initially discussed under system QA, the decision support tools utilized by CADDIS will also present a QA challenge. External tools such as BASINS or MIMS may again follow the pattern of originator responsibility for QA, CADDIS review and acceptance of that level, and CADDIS communicating the QA level and type. Internal tools, on the other hand, connote significant QA responsibility in the soundness of their science as well as their functionality. Clearly the standards set for the supporting science in the CADDIS internal tools must be consistent with, but also will aid in the QA of, the CADDIS literature and data components.

Summary: Generic QA Principles to Consider

Group C offered the following general statements as useful guidance for the continued development of CADDIS QA elements:

- Every function designed into CADDIS should be periodically verified as performing according to plan.
- CADDIS has a responsibility to characterize and effectively communicate the type and level of QA of the information a user may encounter through CADDIS and/or use in an analysis.
- The level of effort devoted to QA should be commensurate with the intent of the analysis and the consequences of encountering error.

- Clearly articulated criteria that determine use or rejection of information sources in CADDIS can objectify the process and help avoid the appearance of Agency endorsement and favoritism or their inverse.
- Data quality objectives selected for CADDIS components can strongly influence the eventual magnitude, complexity, functional capabilities and reliability of CADDIS.

4.2.3 Session 3: Interaction With and Among Users

When considering the development of a system from the perspective of the user community, it is important to think about how the system will facilitate communication among the users, as well as between the developers and users. As such, Group C spent time in Session 3 discussing four questions:

1. How can we optimize the communication among users, CADDIS and researchers?
2. What benefits are associated with interactions within the user group? How can we minimize the risks?
3. Which website “contact us” information would attract the feedback we want?
4. How would a CD handle interactions with and among users?

The group focused on three main types of communication: outreach, technical support, and researcher interactions. Outreach to the user groups includes needs assessment, marketing, and training. Technical support was considered to include both system support and process support. Researcher interactions focused on the communication of research needs from the users to the researchers, in addition to providing direct technical assistance to the users. The group discussed not only how CADDIS could optimize these three types of communication, but also the risks to facilitating this communication.

Optimizing Communication

Outreach

Outreach to the user groups needs to happen at all phases of CADDIS development. During the early phases of development, communication with the users involves needs assessments as well as a marketing component. Later, once a system is established, while needs assessment and marketing continue, a training component should also be added. Group C recognized the importance of utilizing various forms of communication, from electronic to face-

to-face interactions, to get access to a broad cross-section of the user community. There was also an interest in capitalizing on existing methods of communication. These included:

- Speaking and holding training sessions at professional and coordinator meetings (e.g., U.S. EPA Regional meetings, Water Coordinator groups, the Council of State Governments, Society of Environmental Toxicology and Chemistry, North American Benthological Society, American Water Resources Association)
- Writing articles in newsletters and trade journals (e.g., U.S. EPA Office of Water's Watershed Events and Nonpoint Source Newsnotes, Ohio U.S. EPA Newsletter)
- Posting messages on an established listserv or e-mail distribution list (e.g., U.S. EPA techloops, STORET and NHD listservs, Water Quality Standards listserv)
- Speaking on established conference calls held by various groups
- Creating a training module on the U.S. EPA, OW Watershed Academy website
- Creating a feedback form on the current U.S. EPA SI Guidance website.

In addition to noting these existing paths of communication, members of Group C expressed interest in creating links with the industries and corporations that are likely to be impacted by the outcomes of SI. These are groups that the CADDIS developers will want to engage to make the best use of existing information databases have been created by the regulated community. Additionally, industry groups are considered potential users of CADDIS and are likely to use the system in their defense when named as the source of an impairment.

Technical Support

Once CADDIS is developed, users will need to have access to technical support. Members of Group C were adamant about the need for a long-term technical support commitment by the Agency or group that houses the system. The group considered a three-pronged approach to technical support. The first, and most resource intensive, type of support would be a dedicated technical support team that could work with users on a one-on-one basis. In addition to individual support, the use of a listserv, with a searchable archive function, was suggested as a way to allow the users to help themselves. Since there is a great deal of experience within the user community, Group C felt that facilitating user communication with a listserv would be a cost-effective method of tapping into considerable expertise. However, there would need to be some level of monitoring of the listserv by the technical support team in case

questions cannot be answered by the user community, and to ensure that the help offered by the users is correct. The third type of technical support mentioned by Group C was allowing users access to lists of bugs in the system and planned system improvements. By making these types of information readily available, the developers demonstrate their commitment to open communication with the users.

Researcher Interactions

The primary type of researcher interactions that Group C discussed was how CADDIS could be used to facilitate the planning of research in the ORD. First and foremost, members of Group C recognized the importance of ensuring that the development of CADDIS and related diagnostic research was included in the existing mechanisms of ORD planning: multi-year plans, the Research Coordination Teams (RCT) and SPRC. One way to help promote CADDIS in the planning process is to demonstrate how CADDIS can be used as a vehicle for ORD diagnostic information transfer, specifically from the Environmental Monitoring and Assessment Program (EMAP) and the Science to Achieve Results (STAR) program.

In addition to promoting research planning for CADDIS within ORD, it was stated that a mechanism for user feedback should be developed in the initial CADDIS designs. For example, as an individual uses CADDIS to do a case, he or she would have the option to fill out a form identifying research gaps. This information could be automatically placed into a database of research needs, which could then be used by the research planners. Other avenues of communication of research needs from the users to the research planners include the use of listservs, one-on-one meetings, and e-mail lists.

While discussing design options for user feedback, members of Group C also mentioned the need to design the system to accept data and information produced by researchers. In this vein, some members of the group expressed the need for synthesis and review papers, instead of single-study papers, since it is difficult and time consuming for investigators to sift through large amounts of literature on one individual subject. Also, bringing information from synthesis papers into CADDIS would be more cost-effective than adding data from single study articles. Additionally, CADDIS is another way to distill information and could build upon the information presented in review articles. Finally, as CADDIS identifies research gaps, the writing of synthesis papers may be a cost-effective method to satisfy these needs.

Specific user and researcher interactions were also discussed by the break-out group. The encouragement of direct technical assistance to the users by researchers was seen by the group as a possible role of CADDIS. CADDIS could help link technical experts in specific areas

to the users by providing them with a database of experts and their contact information, categorized by technical expertise. While this was seen as a great benefit to the users, it was also recognized that the researchers themselves may show some resistance to this contact. However, since technical assistance is being promoted by U.S. EPA management as one role of ORD scientists, there may be increasing managerial support for this function of CADDIS.

Risks

While the benefits to promoting communication lines between researchers, system developers and the user community are easy to perceive, it is important to consider the possible risks that are associated with this communication as well. Group C discussed the risks associated with the communication in outreach, technical support and researcher interactions.

Outreach

Members of Group C noted that U.S. EPA has had problems in the past with inconsistent messages being given to user groups and to the States in particular. Thus, it was felt that a communication strategy that demonstrated to the States how OW programs and ORD programs, such as EMAP and CADDIS, all fit together. In fact, CADDIS might be used as a vehicle for showing the connections between various U.S. EPA programs.

There was also a recognized risk in reaching out to the possible sources (e.g., the regulated community). Specifically, there is the possibility of misuse of the system by industries that have been identified as sources of impairments. A source could use data unavailable to the States that affects the results and points to another source. It is also feasible that industry groups may take the system and enhance it to the point that it surpasses the capabilities of the system as it is used by the States. Such enhancements might give the industries greater confidence, certainty and power in litigation to defend themselves when identified as sources.

When communicating with all user groups, Group C also identified the importance of managing expectations. Particularly when engaging in marketing activities, there is danger of raising the expectation of the user community and then not delivering the product in the end. There is also the danger of raising the U.S. EPA managers' expectations that the system, or part of it, will be available sooner than is technically feasible.

Technical Support

The risks associated with technical support are limited mostly to the difficulty in managing versions of the system. There is a danger of users having varying versions or accepting only some of the system updates. This would make technical support a difficult endeavor, and, as such, it is important to consider version control when designing the system.

Researchers Interactions

Group C identified the primary risks in this category as being those associated with losing intellectual property. From the perspectives of the researchers, there is little incentive to put their data and information up into the system prior to publication for fear of being “scooped.” Thus, there needs to be a mechanism for encouraging researchers to get information into CADDIS as soon as possible and in a way that does not compromise their intellectual property.

Summary

Group C discussed the importance of communication with and among user groups, once those groups have been identified. While there were recognized risks to fostering this communication *via* CADDIS, the benefits appeared to far outweigh perceived risks. The group noted the importance of keeping communication in mind when designing the system, as well as building effective communication pathways into the system from the beginning. As long as effective management of expectations and protection of scientists’ intellectual property are incorporated into the communication strategy, CADDIS should prove to be an effective tool for identifying research needs and for ORD research planning.

5. FINAL PLENARY DISCUSSION

At the final plenary discussion, participants provided their recommendations on the system development approach, high-priority functions, information and research needs, and timing and dates for system development and use. At the close of this discussion, each participant was asked to express a key thought about CADDIS. Upon review, several themes emerged. These are presented next along with a synthesis of the participants' remarks.

Focus on the Users

“Work toward building a community of satisfied users.”

Many of the participants emphasized the need to be in touch with potential user groups to fashion the design and implementation of CADDIS. This was expressed as a “must do” to make CADDIS a valuable system.

State and Tribal scientists working in TMDL programs were identified as a core user community. Investigators in other organizations (e.g., watershed groups) and programs (e.g., Superfund) may also use the system. While there are many potential user groups, caution should be exercised so that CADDIS does not try to do too much and be everything for everyone.

Although professional society- and U.S. EPA-sponsored meetings are good venues to reach potential users, not all the right people may be present, and incomplete distribution of materials may occur back at work. Other suggestions for identifying and engaging users included:

- Add a box for feedback and suggestions to the SI web page. The 26,000 hits on this page indicate that it is an important information source for scientists working in this area.
- Link to the Watershed Academy and have a training module to introduce SI and the development of CADDIS to the user community.
- Distribute a fact sheet that would point people to a prototype where they try out different functions and provide feedback.
- Arrange one-on-one meeting with users who have varying levels of familiarity with SI and CADDIS.
- Implement or encourage Region-sponsored forums.

All potential users and user groups do not have to be contacted at the same time during development. Rather, user needs assessment and engagement should be a continuing part of system development.

Users should help define functions of CADDIS that should receive high priority by identifying what they cannot do or find difficult to do now. Users could assist in the formation of electronic list services, give suggestions for the presentation of synthesis information on such things as stressor-response relationships, and provide feedback on specific program needs such as TMDL reviews. Such user engagement was also seen as a way to demonstrate how CADDIS might deal with the diversity within the States with respect to state-specific approaches, state-specific impairments, state-specific needs and wants, the availability of various state databases and other state-specific sources of information, and the information technology requirements across States.

Provide Functions That Guide Users Through the SI Process

“Walk this way.”

Participants viewed CADDIS as providing an electronic, step-by-step guide through the SI process. The step-by-step guide was seen as important to help users document the decision process and improve the defense of causal relationships.

System functions in the early phases of development can include tips on getting started, guidance and prompts for users as they move through the process, and advice on avoiding pitfalls. Forms and formats for tables needed for the processes can be provided. A progress tracking system can help users organize the information they use during the process and create bibliographies. More advanced functions can help users gather and organize information in one place, increase the automation of table creation, and provide easy access to particularly useful analytical tools (e.g., data visualization and correlation). Specialized features might target requirements for TMDL reporting and review.

Participants cautioned that a computer cannot substitute for professional judgment: system operations and output functionality should be designed to avoid a “black box” effect. For example, CADDIS could simply provide users with tables, or it could generate complete reports based on information supplied by users. However, the CADDIS design should prevent the possibility of automating report generation to the extent that users do not know what went into that report.

Provide Functions That Offer Helpful Information at Key Steps in the SI Process

“Knowledge is power.”

A number of participants remarked that if CADDIS was nothing more than an SI guide and tracking system it might fall short. They strongly recommended that CADDIS provide key information (i.e., a knowledge base) for the user, and that the development and synthesis of a knowledge base be a primary function of CADDIS. Providing information that is difficult to obtain could serve as incentive for getting users engaged with CADDIS.

Stressor-response relationships were foremost on the list of needs for the knowledge base. Other needs included targeting the literature and experiences regarding remediation and restoration efforts, and synthesizing information on how consistently stressors are associated with different impairments. Access to this information might involve providing one- to two-page synopses, longer and more thorough synthesis papers, as well as facilitating access to the literature. CMs can serve as an organizing framework for the information, and categorization schemes can be used to facilitate access to information by relevant spatial units (e.g., ecoregion or watershed).

Participants emphasized the need to initiate development of the knowledge base with the highest priority stressors and responses. Our current knowledge of the stressors influencing TMDLs can serve as a foundation for prioritization. Information on responses to physical stressors such as sediment and habitat alteration were seen as urgent needs. Beyond stressor-response information, participants noted that our clientele also need source-stressor relationship information. Although this stretches the current scope of CADDIS, its importance for effective management (e.g., allocation in a TMDL) strongly advocates its inclusion in the knowledge base.

Provide a Means of Sharing Information and Lessons Learned

Participants indicated that CADDIS should foster communication among users, as well as between users and the research community. This function was emphasized because there is still limited experience within the scientific community on addressing unknown causes in aquatic systems. Sharing case studies and useful sources of information would enable the community as a whole to improve its expertise in causal evaluation. In addition, our understanding of certain stressor-response relationships is incomplete. Given an effective feedback mechanism, research needs and data gaps identified by CADDIS users could be fed into existing research planning

and coordination efforts such as the OW/ORD SPRC Workgroups, the ORD multi-year research plans and the research planning committees.

Two commercial systems, S-Plus and GoldSim, were mentioned as models of possible approaches for sharing information and experiences. S-Plus maintains a listserv where users can post questions and S-Plus programs and routines. The listserv is mediated by the system developers so that they can personally reply to the more difficult questions, and it is archived to facilitate searching for relevant postings. GoldSim is a simulation engine that also encourages user interaction. The developers maintain a website where users can upload case studies. The website approach may allow more flexibility in sharing information.

Decide on System Platform and Home

“There’s no place like home.”

Early decisions on system platform and home need to be made to initiate development of CADDIS. However, final decisions should follow from a more extensive probing of user needs and technological capabilities.

Group C (see Section 4.2.1) developed a table summarizing strengths and limitations of different platform options. Different homes for the system reflect different strengths and limitations; it may prove best to pursue different homes for different functions. For example, guidance about the SI process, forms, and case studies, can all easily be done using an U.S. EPA server. In contrast, functions that involve uploading and sharing information from users external to U.S. EPA were seen as a potentially difficult issue, based on concerns for security, privacy and maintenance of server functionality. While it may be possible to provide these functions within the U.S. EPA-server constraints, a mechanism outside of U.S. EPA (e.g., *via* an NGO, academic institution, or professional society) may provide easier, earlier, and more flexible functionality.

Money, Resources and Marketing

“What’s the BUZZ?”

It was clear that the success of CADDIS as a project will require time and resources. Support and commitment of resources are needed within the Agency. More and fresher legs are needed to carry out the long-term needs related to design, development, revision, and marketing. Improving the scientific underpinning of CADDIS will also require planning and resources. This is an ambitious project that has the potential for greatly improving the States’ and Agency’s

abilities to identify problem areas and to guide environmental improvements. To accomplish this, participants emphasized that a team will be needed.

Achieving Agency recognition that CADDIS is a long-term project will require developing and articulating a clear vision. Compelling short-form descriptions for distribution within the Agency will help solidify a vision of CADDIS – its structure, function and value. Developing a strategic plan for CADDIS “the project” as well as the plan for CADDIS “the system” will help in the planning for longer term science and development needs while providing valuable functions in the shorter term.

Early successes will also help solidify commitment for the longer term vision. The planning and development process should aim for delivery in the short term, with additional functionality refined through user interaction. The scope of CADDIS should be periodically revisited with user input, but kept sufficiently narrow to insure success.

Effective marketing of CADDIS is another requirement for long-term success. Components of a marketing strategy should include engaging internal managers for budgeting support, the consulting community to stimulate their development of SI and CADDIS as an application for solving environmental problems, and scientists involved with the TMDL program.

Developing links to the TMDL program was seen as a particularly critical element to effective marketing. TMDL user groups can be engaged by provide links to CADDIS on TMDL-related sites. The familiar “figure eight” diagram of listing and TMDL development under the CWA can be marked up to show where CADDIS fits (Figure 3). The conceptual modeling component of CADDIS can help link SI with TMDL efforts. Demonstrations or prototypes of CADDIS can be provided at key meetings at which state biologists and others involved with TMDL are expected. Case studies can be used to illustrate how CADDIS can serve as the “front end” of a TMDL problem, can demonstrate how CADDIS can save time when doing SI, and can underscore the value added of CADDIS for dealing with complex situations.

Approaches

“Be happy at small steps and gains, as this is likely to be the reality.”

A number of participants emphasized the value of an iterative approach to developing CADDIS. An iterative approach can keep the focus on users’ needs, and enable the developers to learn quickly from successes and failures.

The overriding theme of these comments was to get started immediately, but to develop CADDIS in little steps guided by user input. The “agile development” user-guided approach to

development (see plenary abstract Section 3.5) was seen as a promising approach. The development team was encouraged to bite off small, doable chunks (phased scope) so that they can document progress, successes and failures. CADDIS may best be developed as a set of modules so that users can select from some or all of these modules. For example, a CM builder function might be attractive to a broad variety of Agency and State programs. A modular approach would also be consistent with a phased development of CADDIS.

Participants recommended identifying a minimum set of functions for the early development phases. These appear to include: a system to step the user through the SI process, access to information about stressor-response relationships (and perhaps other critical information) when needed, a CM support system (at least at a simple level), and a recording and reporting mechanism. Analytical capabilities within CADDIS are seen as less essential in the early phases, although user group inputs may refine this concept over time if an agile development approach is employed.

Guided by user input, CADDIS might grow in at least two ways: growth in sophistication and complexity (i.e., include analytical tools, better accounting systems, etc.) and growth in the supporting knowledge base (e.g., information about more stressors or more types of habitats). Realizing the full potential of CADDIS will require a long-term commitment.

Broader Vision

“I can see for miles and miles and miles.”

Two key points were raised with respect to looking beyond the applications of SI to date and beyond the TMDL program:

- There are lots of different aquatic and wetland habitats besides freshwater streams.
- While the TMDL writers represent a core group of users, there are many other potential users and we should be careful not to miss those opportunities. Examples include applications for Superfund and ecosystem management programs (water, wetlands, and terrestrial).

Timing and Getting Started

“Time’s a-wasting.”

Many participants emphasized that the project needs to be on a fast track, at least for the early phase. States are moving forward on how to deal with TMDL-related issues. The need is

there now. Since that need will be filled one way or another, timing is critical for success. Participants provided the following suggestions for near-term actions:

1. Get a prototype up and running. Something needs to be “out there” in a short period of time, even if only as a demonstration. At a minimum, make forms available. A second priority would be uploading case studies into the system. Shoot for early successes with low-technology options, and then repackage them for the web. ECOTOX is willing to provide downloads as samples to show how that database can provide information useful to SI.
2. Start engaging users. The upcoming Water Environmental Federation meeting in November and the U.S. EPA biocriteria clinics planned for March 2003 are two good opportunities. Provide training in both SI and the CADDIS prototypes. The number and type of case studies should be expanded to include more Regions and non-point-source issues. For example, it would be helpful to have something that is more non-point source related, and in the western part of the U.S.

6. CONCLUSIONS

There was a clear consensus among workshop participants that CADDIS should be developed, because it has the potential to greatly aid investigators in States, Tribes, and Regions to identify the causes responsible for ecological degradation. Workshop participants generated the following recommendations:

- High priority functions include a step-by-step guide through the SI process, a knowledge base that would help users find and interpret relevant information, assistance in organizing and reporting results, and a means for sharing information and lessons learned.
- The system should be developed incrementally and iteratively.
- Frequent user input and feedback will be essential to the system's success.
- Different system homes have different strengths and limitations; it may prove best to pursue different homes for different system functions.
- A clear vision and early successes will help garner the resources and commitment necessary for the long-term viability of CADDIS
- CADDIS has many potential uses beyond the TMDL program, including applications in terrestrial environments and ecosystem management. The system should be developed in a way that could support future expansion.
- Time's a-wasting. The early phases of CADDIS should be pursued immediately because States are already moving forward on many TMDL-related issues.

This workshop report will be used as the basis of a strategic plan for system development and for more specific recommendations for the system's design, platform and architecture. It will also be used to identify longer term research projects that will be useful for filling data gaps and populating the system. By helping investigators identify the stressors responsible for effects, CADDIS will provide a critical step toward identifying sources, taking action and improving the nation's environmental quality.

APPENDIX A

**AGENDA, PARTICIPANTS AND
BREAK-OUT GROUP ASSIGNMENTS**

Agenda

Causal Analysis/Diagnosis Decision Information System
 August 26 - 28 2002
 Mt. Sterling, OH

Workshop Objectives:

1. To conceptualize an information system (decision support system, expert system) that will help investigators find, access, organize and share information useful for causal evaluations in aquatic systems.
2. To identify critical research needs for system implementation and population.

Monday August 26

Morning
 Travel to Deer Creek

11:30-12:45 Registration

Afternoon

1:00	Welcome/meeting objectives	Mike Slimak
1:15	Stressor Identification History and SPRC	Bill Swietlik
1:30	Stressor identification process	Glenn Suter
2:15	The CADDIS concept, Meeting Objectives	Sue Norton
2:35	Discussion of objectives and agenda	Susan Cormier
3:00	Break	
3:30	Options papers, Prototype	Tom Stockton
4:30	Developing Innovative Software	Steve Fine
5:00	Overview of Tuesday's agenda	Leela Rao
6:00	Poster social	

Tuesday August 27

8:00	Intro to the Little Scioto Case Study	Susan Cormier
8:30	Intro to the Mid-Atlantic Case Study	Barbara Brown
9:00	Session 1	
11:30	Lunch	
12:30	Report out from Session 1	Charlie Menzie (Facilitator)
1:30	Session 2	
4:30	Break	
5:00	Report out from Session 2	Charlie Menzie (Facilitator)
6:00	Overview of Wednesday's agenda	Leela Rao

Wednesday August 28

8:00	Session 3	
10:30	Break	
11:00	Report out from Session 3	Charlie Menzie (Facilitator)
12:00	Lunch	
1:00	Final plenary	Charlie Menzie (Facilitator)
3:00	Adjourn	

Participant List

Name	Phone	Affiliation	E-mail	Address
Brad Autrey	513-569-7368	SoBran	Autrey.Brad@epa.gov	26 W. Martin Luther King Dr., Cincinnati, OH 45268
Sharon Batterman	218-529-5220	ORD/NHEERL	Batterman.Sharon@epa.gov	6201 Congdon Boulevard, Duluth, MN 55804
Paul Black		Neptune & Co.		
Barbara Brown	401-782-3088	ORD/NHEERL	Brown.Barbara@epa.gov	27 Tarzwell Dr., Narragansett, RI 02882
Cheri Butler		Menzie-Cura		
Susan Cormier	513-569-7995	ORD/NERL	Cormier.Susan@epa.gov	26 W. Martin Luther King Dr. [642], Cincinnati, OH 45268
Tod Dabolt	202-566-1186	OW/OWOW	Dabolt.Thomas@epa.gov	1200 Pennsylvania Ave., NW [4503T], Washington, DC 20460
Naomi Detenbeck	218-529-5025	ORD/NHEERL	Detenbeck.Naomi@epa.gov	6201 Congdon Boulevard, Duluth, MN 55804
Don Dorsey	240-777-7729	Montgomery Co. DEP, Watershed Mgmt. Division	don@askdep.com	255 Rockville Pike, Suite 120, Rockville, MD 20850
Chris Faulkner	202-566-1185	OW/OWOW	Faulkner.Chris@epa.gov	1200 Pennsylvania Ave., NW [4503T], Washington, DC 20460
Steve Fine	919-541-0757	ORD/NERL	Fine.Steven@epa.gov	USEPA Mailroom [E243-04], RTP, NC 27711
Jeff Frithsen	202-564-3323	ORD/NCEA	Frithsen.Jeff@epa.gov	1200 Pennsylvania Ave., NW [8601D], Washington, DC 20460
Laura Gabanski	202-566-1179	OW/OWOW	Gabanski.Laura@epa.gov	1200 Pennsylvania Ave., NW [4503T], Washington, DC 20460
Ken Galluppi	919-541-3306	ORD/NERL	Galluppi.Kenneth@epa.gov	USEPA Mailroom [D305-01], RTP, NC 27711
Jeroen Gerritsen	410-356-8993	Tetra Tech, Inc.	jeroen.gerritsen@tetratech.com	10045 Red Run Blvd., #110, Owings Mills, MD 21117
Gretchen Hayslip	206-553-1685	EPA/Region 10	Hayslip.Gretchen@epa.gov	1200 Sixth Ave. Seattle, WA 98101
Cheryl Itkin	202-564-3357	ORD/NCEA	Itkin.Cheryl@epa.gov	1200 Pennsylvania Ave., NW [8601D], Washington, DC 20460
Jan Kurtz	850-934-9212	ORD/NHEERL	Kurtz.Jan@epa.gov	Sabine Island Drive, Gulf Breeze, FL 32561-5299
Jackie Little		T.N & A.		
Charlie Menzie		Menzie-Cura		
Scott Minamyer	513-569-7175	ORD/NRMRL	Minamyer.Scott@epa.gov	26 W. Martin Luther King Dr. [G75], Cincinnati, OH 45268
Bruce Mintz	919-541-0272	ORD/NERL	Mintz.Bruce@epa.gov	USEPA Mailroom [D305-01], RTP, NC 27711
Pete Nolan	617-918-8343	EPA/Region 1	Nolan.Peter@epa.gov	11 Technology Drive [OEME], North Chelmsford, MA 01863-2431
Doug Norton	202-566-1221	OW/OWOW	Norton.Douglas@epa.gov	1200 Pennsylvania Ave., NW [4503T], Washington, DC 20460
Sue Norton	202-564-3246	ORD/NCEA	Norton.Susan@epa.gov	1200 Pennsylvania Ave., NW [8623D], Washington, DC 20460
Lynn Papa	513-569-7587	ORD/NCEA	Papa.Lynn@epa.gov	26 W. Martin Luther King Dr. [190], Cincinnati, OH 45268
Bruce Peirano	513-569-7185	ORD/NRMRL	Peirano.Bruce@epa.gov	26 W. Martin Luther King Dr. [290], Cincinnati, OH 45268
Dave Pfeifer	312-353-9024	EPA/Region 5	Pfeifer.David@epa.gov	77 West Jackson Boulevard [WT-16J], Chicago, IL 60604-3507
Ed Rankin	740-517-2275	Ohio University/ CABB	rankin@ilgard.ohiou.edu	9666 State Route 37 East, Sunbury, Ohio 43074
Leela Rao	202-564-3362	ORD/NCEA	Rao.Leela@epa.gov	1200 Pennsylvania Ave., NW [8623D], Washington, DC 20460
Chris Russom	218-529-5218	ORD/NHEERL	Russom.Chris@epa.gov	6201 Congdon Boulevard, Duluth, MN 55804
Mike Slimak	202-564-3324	ORD/NCEA	Slimak.Michael@epa.gov	1200 Pennsylvania Ave., NW [8601D], Washington, DC 20460

Name	Phone	Affiliation	E-mail	Address
Jonathan Smith	202-564-5181	ORD/NCER	Smith.Jonathanh@epa.gov	1200 Pennsylvania Ave., NW [8725R], Washington, DC 20460
Robert Spehar	218-529-5123	ORD/NHEERL	Spehar.Robert@epa.gov	6201 Congdon Boulevard, Duluth, MN 55804
Tom Stockton		Neptune & Co.		
Bhagya Subramanian	513-569-7349	ORD/NERL	Subramanian.Bhagya@epa.gov	26 W. Martin Luther King Dr. [642], Cincinnati, OH 45268
Glenn Suter	513-569-7808	ORD/NCEA	Suter.Glenn@epa.gov	26 W. Martin Luther King Dr. [117], Cincinnati, OH 45268
Bill Swietlik	202-566-1129	OW/OST	Swietlik.William@epa.gov	1200 Pennsylvania Ave., NW [4304T], Washington, DC 20460
Marge Wellman	202-566-0407	OW/OST	wellman.marjorie@epa.gov	1200 Pennsylvania Ave., NW [4305T], Washington, DC 20460
James Wickham	919-541-3077	ORD/NERL	Wickham.James@epa.gov	USEPA Mailroom [E243-05], RTP, NC 27711
Chris Yoder	740-517-2274	Ohio University/ CABB	yoder@ilgard.ohiou.edu , db8177@dragonbbs.com	P.O. Box 21561, Columbus, OH 43221-0561
Lester Yuan	202-564-3284	ORD/NCEA	Yuan.Lester@epa.gov	1200 Pennsylvania Ave., NW [8623D], Washington, DC 20460

Breakout Group Assignments

Breakout Group	Role	Participant/Affiliation
Group A: Little Scioto Case Study	Facilitator	Jackie Little T.N. & Associates
		Tech Lead
	Tech Lead	
		Notetaker
	Notetaker	
		Notetaker
	Notetaker	
		Notetaker
	Notetaker	
		Notetaker
	Notetaker	
		Notetaker
	Notetaker	
		Group B: Cabin John Case Study
Tech Lead	Barbara Brown U.S. EPA / NHEERL-Narragansett	
	Tech Lead	
Notetaker		
	Notetaker	
Notetaker		
	Notetaker	
Notetaker		
	Notetaker	
Notetaker		

Breakout Group	Role	Participant/Affiliation
		James Wickham U.S. EPA / NERL - RTP
		Chris Yoder Ohio University / CABB
		Lester Yuan U.S. EPA / NCEA - Washington
Group C: System Development/ User Perspectives	Facilitator	Scott Minamyer U.S. EPA / NRMRL - Cincinnati
	Tech Lead	Doug Norton U.S. EPA / Office of Water - OWOW
	Tech Lead	Tom Stockton Neptune & Co. (subcontractor to TN & A)
	Notetaker	Leela Rao U.S. EPA / NCEA - Washington
		Tod Dabolt U.S. EPA / Office of Water - OWOW
		Steve Fine U.S. EPA / NERL - RTP
		Cheryl Itkin U.S. EPA / NCEA - Washington
		Lynn Papa U.S. EPA / NCEA - Cincinnati
		Dave Pfeifer U.S. EPA / Region 5
		Ed Rankin Ohio University / CABB
		Chris Russom U.S. EPA / NHEERL - Duluth
		Jonathan Smith U.S. EPA / NCER
		Bill Swietlik U.S. EPA / Office of Water - OST
		Marge Wellman U.S. EPA / Office of Water - OST
	Cross-session facilitator	Sue Norton U.S. EPA / NCEA - Washington

APPENDIX B
BACKGROUND MATERIALS

**Design Options for the
Causal Analysis and Diagnosis
Decision Information System
(CADDIS)
Workshop Draft**

Prepared by

Tom Stockton and Paul Black

of

Neptune and Company, Inc.

**Prepared for T N & Associates, Inc.
under Contract Number 68-C-98-187**

**March 2002
revised July 2002**

**This document has been reviewed and approved for use as background for
the CADDIS Workshop by the EPA Work Assignment Manager.**

Introduction

This report outlines options for the Causal Analysis and Diagnosis Decision Information System (CADDIS). CADDIS is envisioned as an interactive web-based decision support tool that will help States/Tribes, Federal agencies, and other water resource managers identify causes of biological impairments or ecological stress in water bodies. CADDIS will provide investigators access to relevant information sources and decision-based tools that can be used to process information to support consistent, defensible and scientifically-based causal evaluations of impaired aquatic systems. This can lead to implementation of better water quality management controls, and more useful and effective expenditures of valuable and often limited resources. The system is based on EPA's Stressor Identification (SI) guidance (USEPA, 2000).

Three system design options for CADDIS are presented here in a hierarchical fashion from less to more complex and resource intensive to develop and implement (CADDIS¹, CADDIS², and CADDIS³). These options build on each other sequentially to define a resource efficient path forward for developing CADDIS. The move to more complex approaches is in large part a move from simply streamlining access to information in a static web-based environment, towards a more holistic framework based on decision analytic approaches aimed at adaptive decision-making in which uncertainties and decision consequences are taken into account. As CADDIS is implemented, examples can be developed and added to the system, so that CADDIS can be dynamic and iteratively update with each application.

In the description of the three options that follows, examples of the look, feel, and the technical approach for implementing the desired functionality are available at www.neptuneandco.com/caddis. The prototype CADDIS web site is only intended as a tool for exploring how a user might be guided through the SI process and for exploring the technical tools needed to implement CADDIS. This site is not intended to be a model for the aesthetics of CADDIS or to provide complete functionality. In general the web tools used in this prototype are Open Source and freely available. A table of the functionality CADDIS could provide is presented in Appendix B-1. For each function or feature, this table indicates which of the three option levels applies, and the anticipated level of effort and resources needed to implement the feature.

The descriptions of the options below focus on implementation in a web-based environment. However, most of the CADDIS functions could be served over the web but actually implemented on a CD-ROM. For example, if CADDIS¹ is implemented solely in HTML it could easily be distributed on a CD-ROM and viewed using a browser such as Internet Explorer or Netscape. CADDIS² and CADDIS³ include dynamic components that could be implemented on a CD-ROM using a "localhost" server that, in essence, creates a local web server on the user's computer, potentially providing all the functionality of the web served version. There are several disadvantages of a CD-ROM based approach. It is difficult to keep the CD-ROM version up to date, especially if each user is allowed to contribute to the building of CADDIS. The CD-ROM based user could not take advantage of other user's input, gain access to external data sources or GIS systems, or contribute their input to the continued development of the system. Also, if internet access is not

available then hyperlinked sites would not be available, limiting access to information that might be available on States, EPA, or other water resources web sites . Potential advantages are that access to the Internet will not be required, and performance will not be based on the performance of the CADDIS server.

Option 1 (CADDIS¹):

CADDIS¹ is envisioned as a guide to both general and stream specific information sources and data relevant to stressor identification. The core of CADDIS¹ would be built on hyperlinking components of the Stressor Identification Guidance Document. This could be somewhat similar to the Adobe Acrobat version of the Guidance Document but with a wizard rather than Table of Contents interface. The wizard approach could step the user through the SI process with hyperlinks to web sites with more information regarding a particular topic in the Guidance Document. This could include web sites for:

- State, Federal and local governments, local watershed groups (e.g., Ohio Environmental Protection Agency Division of Surface Water),
- mapping (e.g., EnviroMapper),
- metadata (e.g., EIMS),
- observational data (e.g., STORET), and
- decision and data analysis (e.g., StatWiz).

CADDIS¹ could present the user with hyperlinks to information and help pages describing each component of the SI process. Each description could contain further hyperlinks to examples for the particular component of the SI process. The examples section of the Guidance Document could be extensively hyperlinked to provide information at several different levels or scales of the technical analysis. Links to downloadable SI worksheets in the form of spreadsheet templates could also be provided to allow the user to complete the SI process steps off-line. In turn, this capability could allow users to submit examples and completed worksheets. The submitted information and examples would then be made available to other users in the form of static information or examples and case studies of implementing the SI process.

A secondary component of CADDIS¹ could tailor the information accessible to each user by providing a map of the United States that leads to a web page specific to each State. Then, each State's web page could provide links to relevant EPA sites, non-profit organizations sites, local watershed groups, as well as the web site of the state's water quality management agency.

Figure B-1 shows an example of the web interface that could be used in CADDIS¹. Links from the SI process components and from a map of the U.S. can be made available directly from this page. These links could access help files, information sources, databases, or State's web-pages. Additional links to allow downloading and uploading worksheets could also be made available. Databases could be downloaded for external analysis, but data analysis capabilities will not be available in CADDIS¹.



Figure B-1. CADDIS page for accessing the SI process or choosing a State.

In summary, CADDIS¹ will facilitate access to information directly through access to a web-based environment, facilitate access to other relevant information sources and data bases either through the SI links or through a link to geographic locations (States), and can be built with downloadable worksheets with the capability to upload the completed worksheets to a database of case studies and examples. This system will be primarily static, providing access to information without also providing dynamic analysis capabilities, although worksheets can be uploaded to continue to supplement the database of case studies and examples.

Technical Tools

The features of CADDIS¹ are limited to allow CADDIS to be implemented solely in static **HTML**. This minimize the resources required to built CADDIS and limits the resources required to serve CADDIS on the internet. This also limits the dynamic interactivity of CADDIS and doesn't allow for dynamic updating or saving user information. This system can help inform the user, but will not directly assist with analysis and implementation of the SI process.

Resources Required

The steps envisioned for building CADDIS¹ include:

- write an HTML wizard to the Stressor Identification Guidance Document, and
- build and populate each State specific web page.

Writing the HTML wizard would be relatively straightforward and could probably be accomplished in a few months. Building and populating each State specific web page would be more resource intensive, although this information may already be compiled and available. Alternatively, CADDIS **Option 2** (CADDIS²) provides an approach that allows the State specific pages to be built gradually by the CADDIS developer within the CADDIS system, while augmented by the CADDIS users. In this way CADDIS² could be more resource efficient alternative than CADDIS¹.

Option 2 (CADDIS²):

The move from CADDIS¹ to CADDIS² is largely a move from a static to a dynamic interactive tool that could have a user updating component. The interactive tool will allow the user to implement the SI process within CADDIS and will provide direct access to some analysis capabilities to facilitate the implementation.

The core component of CADDIS² is a stream specific wizard that allows the user to enter, store, and retrieve information for their stream application. The wizard could step the user through each phase of the SI process, building a record for a stream in a database of all streams for which CADDIS has been applied. Each application can be stored and retrieved at will, allowing a database of applications to be developed on-line. This database will effectively provide a resource tool for other potential users of the system, who will be able to review case studies and examples, or even copy them as a template for their own stream specific or water management problem.

In addition to allowing the user to complete the SI process electronically within CADDIS², the user will also be provided access to databases and other information sources. This option of CADDIS will also include some data analysis functionality, so that data that are retrieved from external sources can be analyzed on-line, and the results can be used to populate that SI process worksheets and spreadsheets, which will be stored in an internal database for future review and editing.

CADDIS² will provide access to the information sources specified in CADDIS¹, and will also provide on-line implementation with analysis functionality. Analysis capabilities will be limited to individual components, so that retrieved external data, for example, can be analyzed directly for the user to input the results in the SI process pages, but will not also include a capability for taking a holistic view of the SI process in which all of the inputs are integrated into a single model. CADDIS² will provide access to information as specified in CADDIS¹, as well as data bases, GIS overlays, data analysis functionality and a presentation tool.

In application the wizard is a series of HTML forms presented to the user in web pages. The following prototype web tool presents one approach to providing this capability and provides a means for exploring and comparing the potential of the various web tools available. The first step or page allows the user to either enter a new stream name in a text form or to chose a previously entered stream from a select form (Figure B-2). The streams made available in the select form could be dynamically generated from the database of previously entered streams. Continually updating a stream database and providing access to the database through the stream name would allow the user to stop and return to the wizard at the most recently completed step without starting over. This page could be preceded by a user login screen. This could serve to monitor the use of CADDIS².

Once a stream name is entered the next step could be to provide the stream a geographical context. If the stream does not already exist in the database the user could be presented with a map of the United States with clickable States, Watersheds, and/or EPA Regions. Figure B-3 provides an example based on a State map. Clicking on a State adds a "State" field to the stream database and allows CADDIS to guide the user to State specific information and data sources. If the stream already exists in the stream database the geographical context could be provided by the State field in the database. In this manner each CADDIS user always has a geographical context that CADDIS can use to refine and tailor itself to the user. A State database would be access by CADDIS that provides links to State specific, as well as, regional and nationwide information sources. If the user discovers information sources not originally provided by CADDIS, the user could add this hyperlink to the State or US database via a text form. Subsequent users of CADDIS then have the advantage of these updated links databases. This allows CADDIS to be continually updated with each application.

Once the geographical context of the CADDIS user has been identified the user moves to the *Define the Impairment* page (Figure B-4). This page could provide a text field form into which the stream impairment information can be entered. If impairments for the stream were previously entered, the text field could be dynamically filled from the stream database and edited as necessary. The new or edited impairment information could then be stored in the stream database. On the *Define the Impairment* page with the impairment form would be links to aid the user in defining the impairments if they are have not already been identified. One of these links could be a mapping and data retrieval tool. The user could be presented with a map of the relevant State with streams, watersheds, towns, roads, gaging stations, water quality monitoring stations, NPDES point sources, land use, and other relevant GIS data layers. Ideally in such a tool, each GIS data layer would provide clickable access to raw data that could be imported into a data analysis wizard.

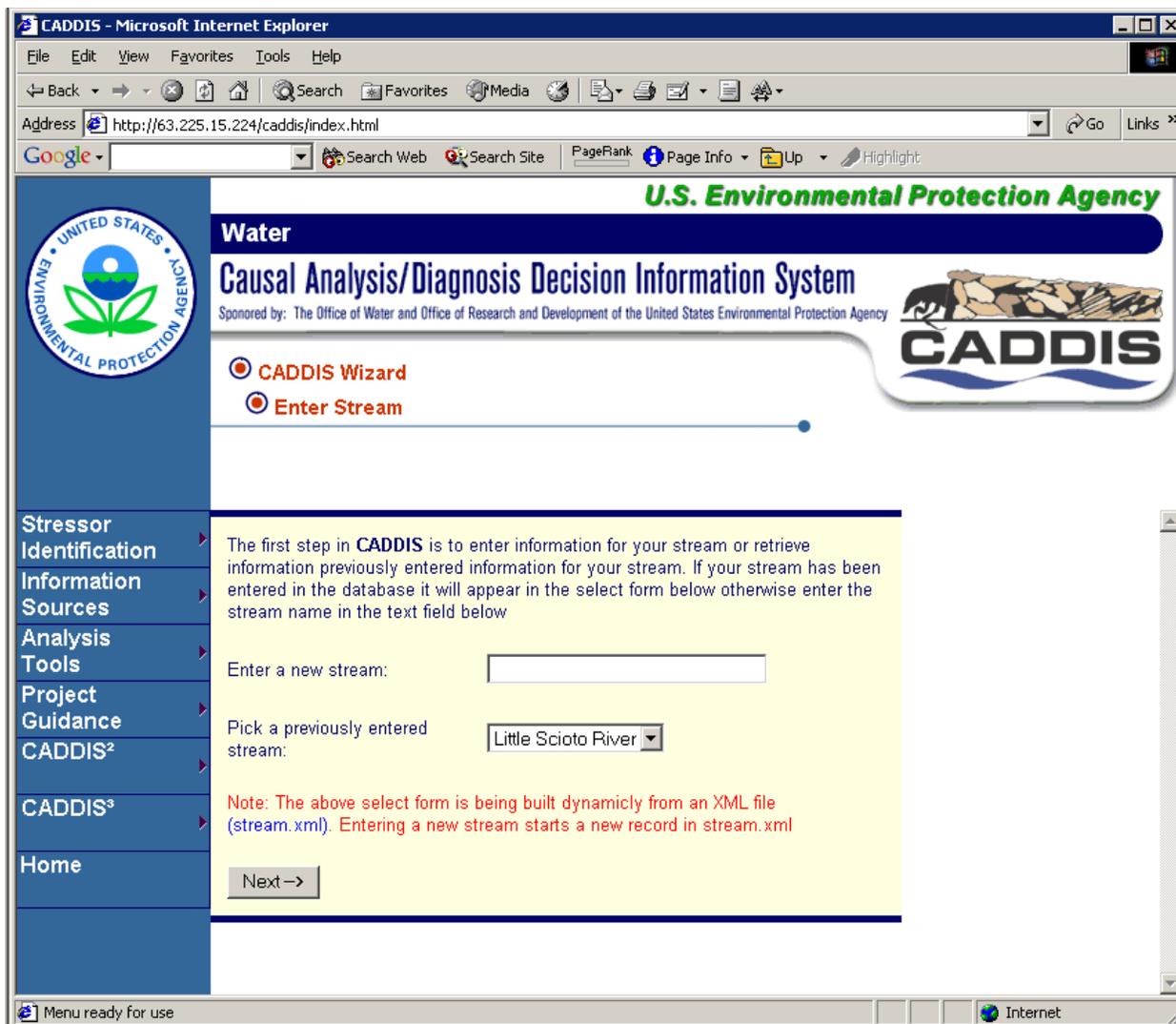


Figure B-2. Initial CADDIS wizard page for entering the stream name. The stream name select form is populated from a database of previously entered streams.

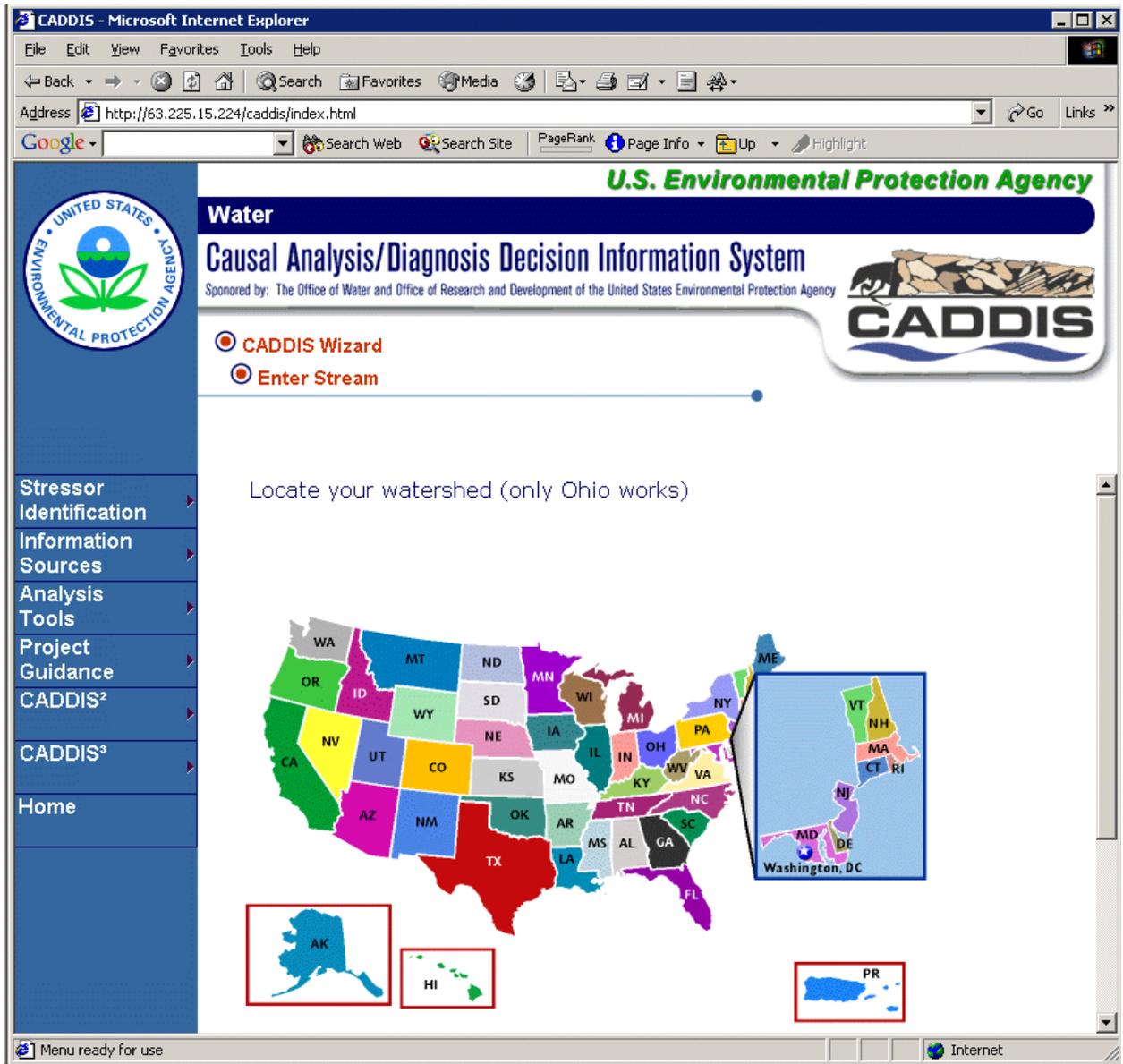


Figure B-3. CADDIS page for assigning and retrieving relevant stream information by State.

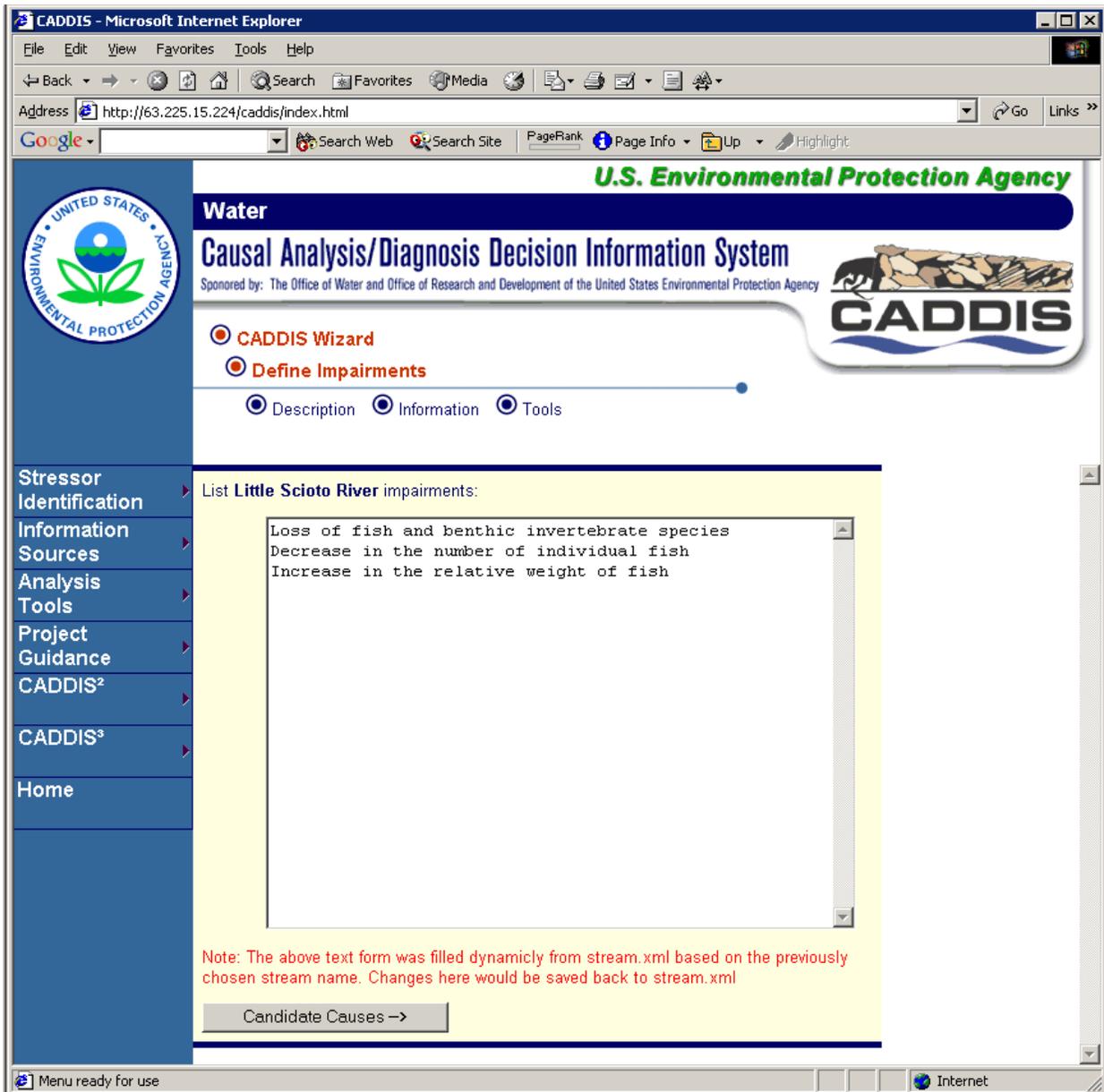


Figure B-4. *Define the Impairment* page. This page includes a link to a *Define the Impairment* Help page and a text field for entering and editing impairments. The impairments are then stored in a database referenced by stream name.

Currently EPA's EnviroMapper (maps.epa.gov/enviromapper) provides some of the desirable functionality. The data layers EnviroMapper provides that would be most relevant to CADDIS include streams (apparently Reach File 1), towns, watersheds, counties, and point sources. Data layers not apparently provided by EnviroMapper include detailed streams (Reach File 3), Federal and State water quality monitoring stations, stream gaging streams, local government or watershed group monitoring stations. EnviroMapper also does not provide a direct data retrieval mechanism. EPA's EnviroFacts web page provides access to some of the data relevant to the CADDIS user that would be identified through EnviroMapper, such as some NPDES point source information, however, direct query of EnviroFacts databases by the public is currently restricted. EPA's Environmental Information Management System (EIMS) provides metadata for data sets, databases, documents, models, projects, and spatial data. However, in general EIMS does not appear to provide access to the underlying data. EIMS could be useful to CADDIS as a guide to data sources other than data sources that already have been identified. An alternative to these existing tools would be development of a mapping and data retrieval tool specifically designed for CADDIS (discussed further in Options 2 and 3 with example available on the CADDIS web site).

STORET could be a very useful data source but it may be difficult to provide a wizard to access the data. STORET is split between the STORET Legacy Data Center and Modernized STORET. The STORET Legacy Data Center includes data prior to January 1, 1999 while Modernized STORET includes post January 1, 1999 data. Neither of these data sources appear to allow data query other than through forms on the STORET web page. Data retrievals are run overnight and requesters are notified via email of a web page where the data can be "picked up".

After the impairments are entered and saved to the Stream database the wizard sends the user to *Candidate Causes* page with a similar look and feel to the *Impairment* page (Figure B-5). Candidate Causes can be entered into a text field form that is saved to the Streams database. Alternatively, help on developing a list of candidate causes can be accessed. This could provide links to examples, maps, local, state, regional, nationwide information data sources, as well as a link to a data analysis wizard. Ideally data and information would be accessed visually with a map interface. Of particular importance would be links to facilitate conceptual model building. Conceptual models are often built in graphical form. Although Open Source free software is available (e.g., dia) there doesn't appear to be a clear path forward for developing a web-based conceptual model drawing tool. One possibility may lie in scalable vector graphics (SVG).

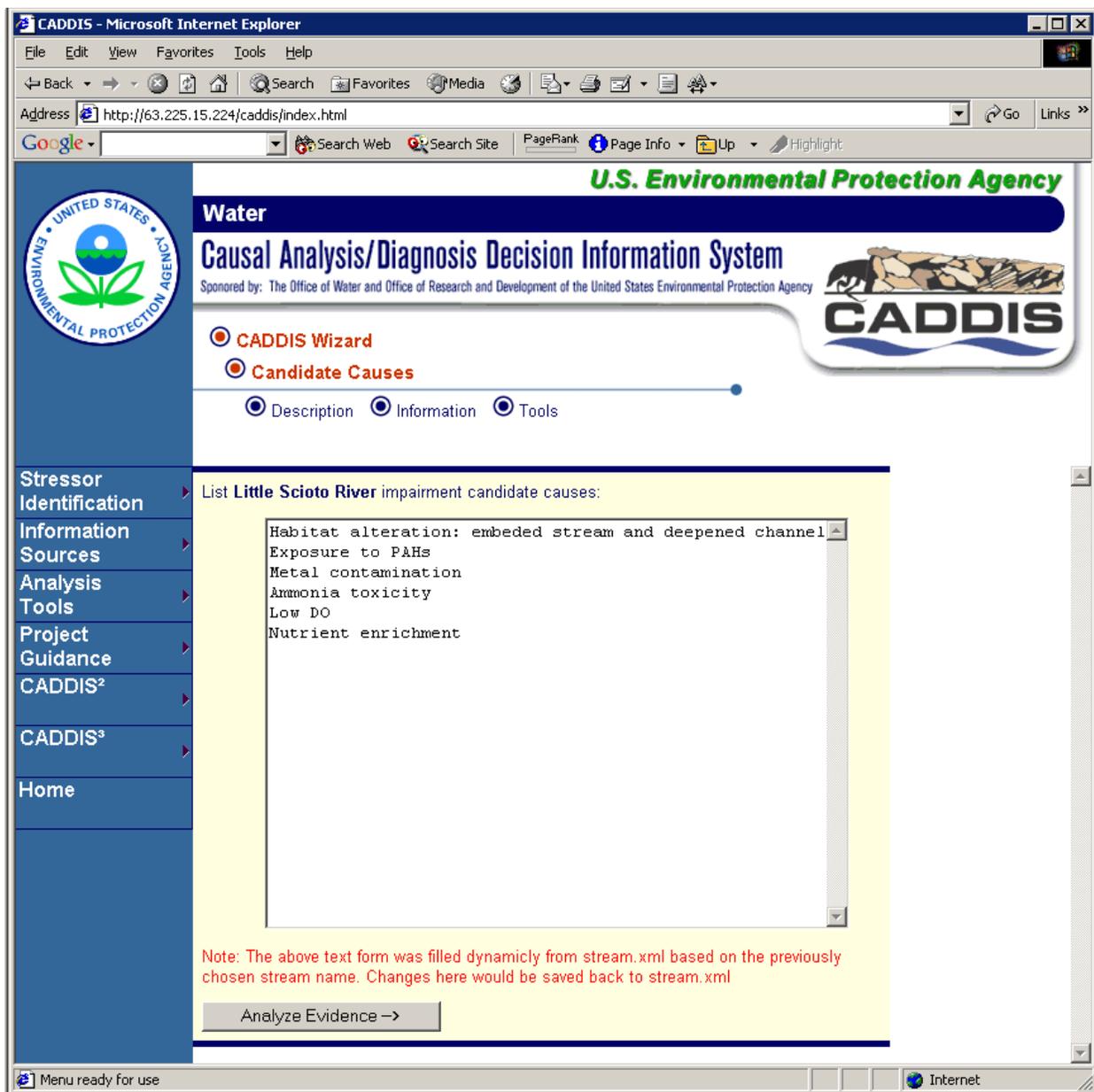


Figure B-5. List Candidate Causes page.

The *Analyze Evidence* page then follows the *List Candidate Causes* page (FigureB-6). This page could provide a link to a data analysis wizard, as well as, guidance documents for interpreting the data analysis results. These guidance documents could include:

1. general statistics guidance,
2. the Analyze Evidence Chapter in the Stressor Identification Guidance Documents, and
3. State and EPA guidance on interpreting data and water quality criteria.

A web-based water quality data analysis wizard could be built with modules designed, for example, to facilitate the comparison of up and down stream biological data, analyze data for trends, or compare data to water quality criteria or thresholds. A data analysis wizard could also provide for visualization and mapping of the stream and data. An important piece of a data analysis wizard would be a data import wizard to facilitate moving data from a water quality database (e.g., STORET) or from a standard software package (e.g., Excel). The data analysis wizard should also provide summary results that facilitate completing each step of the SI process. An example of such as data analysis wizard that implements some of these features is available at www.neptuneandco.com/StatWiz. Analysis and interpretation of data could be a component in each step of the SI process. Therefore the links available on the *Analyze Evidence* page could be viewed as a compilation of all the links available on the help pages for the previous pages.

After the *Impairments* and *Candidate Causes* are defined and the *Evidence Analyzed* the CADDIS wizard provides a *Characterization of Causes* page that allows all *Candidate Cause* and *Impairment* combinations to be scored in several causal considerations categories (Figure B-7, the causal considerations presented in this figure are not meant to be exhaustive but only serve as an example). Scores are input in a matrix of text forms and saved to the Stream database. These scores are synthesized and presented in the *Identify Cause* page. This page could provide a summary visualization of the Cause characterization table to aid in Identifying causes.

Overall, CADDIS² will provide access to information sources similar to CADDIS¹, but will also provide access to databases with retrieval capabilities for standard database structures (Oracle, ACCESS, Excel, text files), and access to GIS information and coverages. If data are pulled or are provided from external sources, a data analysis wizard can be made available that is tailored to the needs of the SI process steps. The user will be able to use and manipulate information that is gathered, and will be provided a capability to implement the SI process directly on-line. Examples will be stored in an internal database format for future retrieval and editing (with some password protection probably), and these examples will add to the list of case studies and examples that can be made available to all potential users. In this way, CADDIS will be continually updated by new applications. Once the user-supplied information is input into the CADDIS forms and worksheets, the stressors can be scored following the SI guidance and the most likely stressors using this mechanism can be identified. This stops short of providing an integrated modeling capability that is deferred to Option 3.

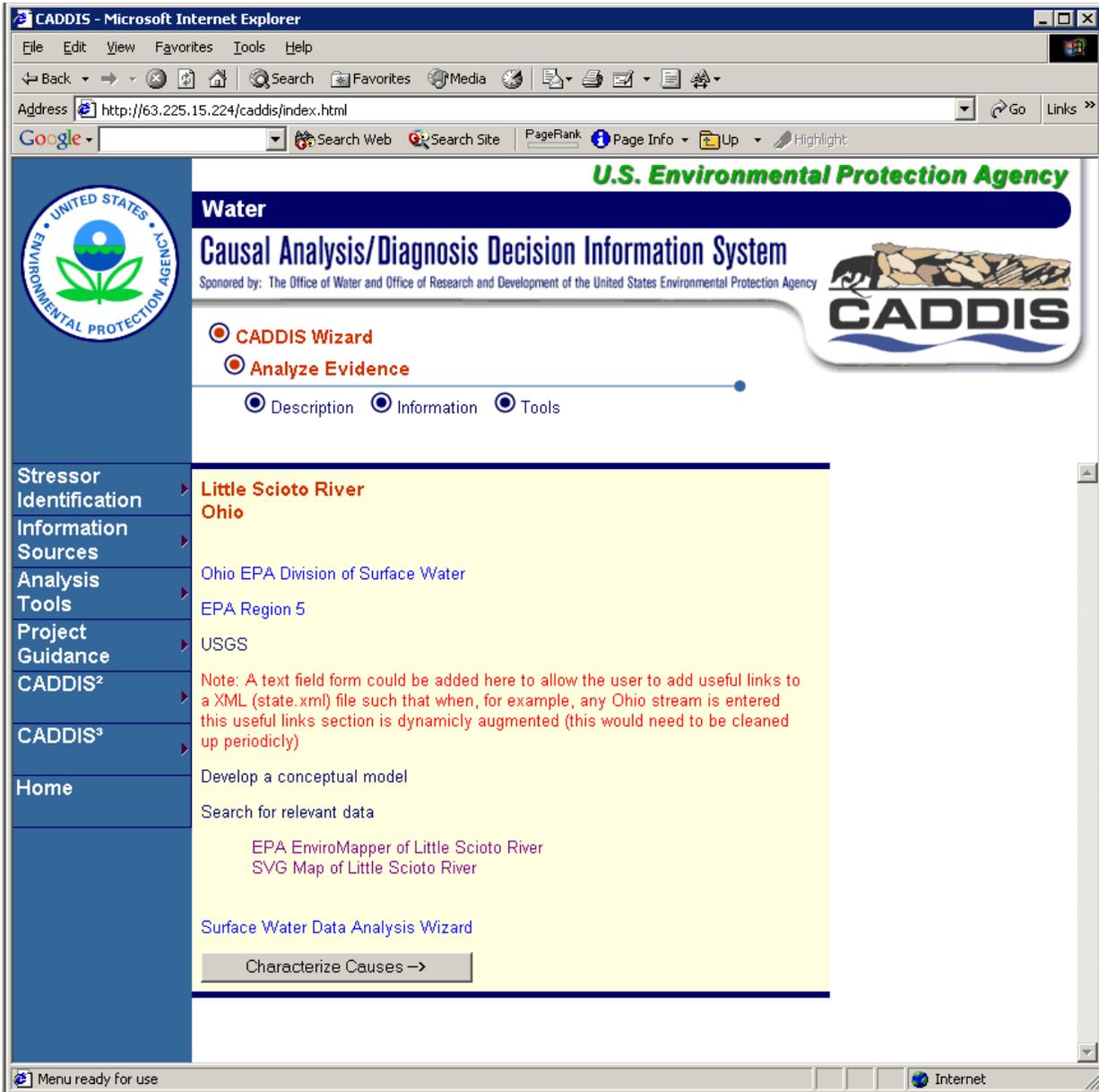


Figure B-6. *Analyze Evidence* page.

U.S. Environmental Protection Agency
Water
Causal Analysis/Diagnosis Decision Information System
 Sponsored by: The Office of Water and Office of Research and Development of the United States Environmental Protection Agency

CADDIS

CADDIS Wizard
 Characterize Causes
 Description Information Tools

CHARACTERIZATION OF CAUSES
Little Scioto River

	Habitat alteration: embed stream and deepened channel	Exposure to PAHs	Metal contamination	Ammonia toxicity	Low DO	Nutrient enrichment
Causal Consideration						

Loss of fish and benthic invertebrate species

cooccurrence	NE	NE	NE	NE	NE	NE
temporality	NE	NE	NE	NE	NE	NE

Decrease in the number of individual fish

cooccurrence	NE	NE	NE	NE	NE	NE
temporality	NE	NE	NE	NE	NE	NE

Increase in the relative weight of fish

cooccurrence	NE	NE	NE	NE	NE	NE
temporality	NE	NE	NE	NE	NE	NE

Next Steps →

Scoring
 NE: no evidence
 NA: not applicable
 ...

Figure B-7. *Characterization of Causes* page. Web based worksheet for organizing impairment and causal relationships. The causal considerations list here are not meant to be exhaustive but are provide for illustrative purposes.

Technical Tools

CADDIS² requires web tools that enable CADDIS to be dynamic and interactive. The dynamic components of the prototype CADDIS were built using HTML, Common Gateway Interface (CGI) Perl scripting, and Perl XML (Extensible Markup Language) parsing. Alternative approaches to CGI include Java Server Pages (JSP), Extensible Server Pages (XSP), and Active Server Pages (ASP). The example data analysis wizard, StatWiz (www.neptuneandco.com/statwiz), uses CGI Perl as a web interface to R, an Open Source statistical programming language. R provides a powerful statistical package that can be tailored to the needs of CADDIS including probabilistic simulation.

The prototype CADDIS uses XML for data storage and retrieval. Alternatively a Relational Database Management System (RDMS) such as Oracle could be used. A RDMS would likely complicate CADDIS and limit portability without providing much value added, since XML is well-suited for storing and managing data and information including features such as text, figures and drawings, and numerical information.

The missing functionality from CADDIS² that would be highly desirable includes data retrieval wizards for some of the standard water quality databases including STORET and USGS water quality and stream flow databases. Providing CADDIS with this functionality does not appear to be achievable in the near future. Links to these data sources could be included in CADDIS and data import wizards could be developed that facilitate import of the data into a data analysis wizard. The EIMS system is designed to allow remote queries and could provide a valuable information resource for CADDIS. The information in EIMS is largely metadata thus does not appear to provide a source for data. In general, current sources of the type of information that CADDIS would use are not organized in a way that allows easy or suitable retrieval. Access to standard databases (e.g., Oracle or ACCESS) can be provided through the data analysis wizard or more directly.

Resources Required

The resources required to build CADDIS² would include the resources required to build the hyperlinked Stressor Identification Guidance Document and the State specific pages outlined in CADDIS¹. Much of the dynamic capabilities of CADDIS² are exemplified in the prototype of CADDIS. Fully implementing these features, as well as, fully developing a surface water data analysis wizard would require additional resources.

The main functions that are needed to make the system functional (e.g., linking to information sources, obtaining data from external data bases, GIS functionality, data analysis functionality, interactive implementation of the SI process) can probably be implemented as version zero in approximately 1 year after initial design with sufficient resources. The initial system will require peer review, beta testing and iteration before it can be released for more general use. The system will also be updated from examples that are implemented and recorded. System management will be necessary, although the system can be built such that it is self-managed to some extent. When the design steps are undertaken, some consideration will need to be given to system management,

hosting, centralization and distribution across networks, and hardware requirements (storage capacity and operational functionality) for the server.

A CADDIS tailored GIS approach would provide a much more functional system than can be provided by EnviroMapper, however significant resources may be required to create the State specific maps and data organization. This would be somewhat similar to the approach used by BASINS and could require computer server resources to store, manage and serve.

Another resources consideration is that CADDIS² places a larger burden on the server since the dynamics of CADDIS are based on some form of server based scripting (CGI, JSP, etc.). CADDIS² also includes a data storage and retrieval component, which further complicates the web server configuration. The data analysis wizard component could provide a substantial computational burden on the server as well, depending on the level of use CADDIS sees. These issues can be managed by increasing hardware capabilities, but might need to be considered early in the planning and design process. Given that CADDIS² will not be ready for about 2 years, and hardware capabilities continue to improve at an alarming rate, hardware specifications might become less of an issue. Similar considerations apply for web-based software tools that might be applicable. The capabilities of these tools has increased dramatically in the last year or so, allowing different software to be linked to produce a fully operational integrated system. For example, web-based tools such as Perl/CGI and JSP now can work hand-in-hand with HTML and link with data bases, GIS programs, statistical software and graphical software. It will be important to stay on top of hardware and software issues, so that the most up to date and necessary computer technology is used when CADDIS² is built, but it should be expected that the appropriate computer resources will exist.

Option 3 (CADDIS³)

In addition to the functionality provided by Options 1 and 2, CADDIS³ would encompass analytical tools that would facilitate modeling of aquatic systems in the context of the SI guidance. That is, fully coupled systems could be developed in which relationships among stressors can be identified and characterized simultaneously. This would be a more holistic approach to stressor identification than provided in CADDIS². Stressor models could be developed using deterministic or probabilistic approaches, either of which could be extended to accommodate decision analysis if desired. Probabilistic modeling will allow uncertainty to be propagated through each step of the SI process. Using this approach uncertainty can be managed effectively and sensitivity analysis can be performed to identify the major sources of uncertainty in the system. Steps that would need to be included in CADDIS³ to facilitate this type of model-based approach (beyond those already specified in CADDIS²) would include building of conceptual models, translating conceptual models into numerical models, specifying the numerical models, simulation, uncertainty and sensitivity analysis, and model updating as more information is collected. CADDIS³ would hence encompass analytical tools that allow uncertainty to be propagated through each step of the SI process. Ideally this would be conducted within a Bayesian paradigm incorporating prior probabilities and data. Prior probabilities

could be generated from expert judgment, prior data collection and analysis efforts as various scales ranging from EMAP at a national scale to studies performed in nearby watersheds.

In any probabilistic modeling system such as this, propagating uncertainty, whether in a Bayesian approach or in a classical probabilistic simulation approach, generally requires some form of Monte Carlo sampling. Influence diagrams are a visually and logically appealing way to implement a probabilistic simulation of a conceptual model. Figure B-8 presents the example conceptual model from the Stressor Identification Guidance Document (EPA 2000). Conceptual models are typically represented in this fashion as flow diagrams. The relationships defined by the flow diagram can be interpreted as defining a joint probability distribution for all components of the conceptual model. Thus if the conceptual model is developed and simulated as an influence diagram the results are probability distributions for all the components of the conceptual model. Using an appropriate simulation algorithm (e.g., Markov Chain Monte Carlo, MCMC) allows the joint probability distribution to be integrated to define the relationship between the uncertainty in stressors and the associated uncertainty in the responses. The real appeal to this approach is that all available information, regardless of its source (primary data, secondary data, expert opinion, non-expert opinion), can be appropriately incorporated into the analysis by expressing the level of confidence in the data source through the variance parameter of the information probability distribution. If the input distributions are objectively parameterized then the analysis will represent the state of knowledge qualitatively described by the conceptual model .

A secondary appeal of this approach is it provides a rigorous mechanism and guide for collecting more information and iteratively updating the conceptual modeling. An ideal approach is to initially develop a conceptual model influence diagram that places minimal resources on collecting new data and information collection. Input probability distributions are specified that reflect the current state of knowledge and the conceptual model influence diagram is simulated. The sensitivities of conceptual model responses to inputs provides a rigorous means of evaluating the cost and value of collecting more information and iteratively updating the conceptual model.

Ideally CADDIS³ would provide a mechanism for drawing, storing, specifying, and solving conceptual model influence diagrams. There are three basic issues to be resolved in implementing such a web-based tool

- drawing and storing the conceptual model influence diagrams,
- providing guidance and data analysis tools to the user for specifying input probability distributions, and
- solving a conceptual model influence diagram.

There exists standalone software that allow influence diagrams to be drawn, specified, and solved, including Analytica and WinBUGs. Analytica is a commercial Windows package that facilitates model building through formal decision-based influence diagrams that can then be solved using a classical Monte Carlo simulation approach. WinBUGs (www.mrc-bsu.cam.ac.uk/bugs/) is a free (currently free but may not be free in the future) Windows package that implements a Bayesian

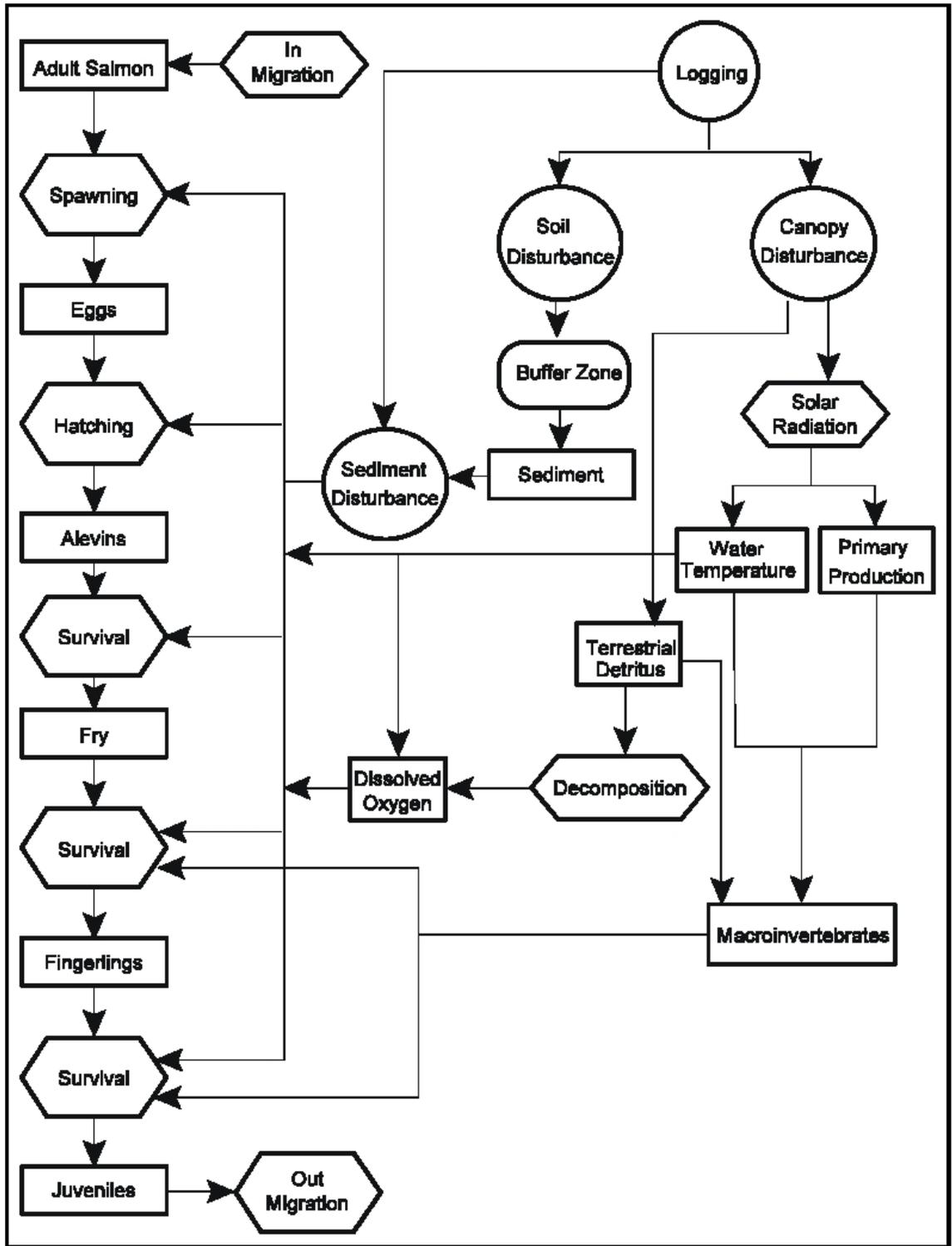


Figure B-8. Example conceptual model diagram from the Stressor Identification Guidance Document (EPA 2000).

MCMC approach to solving complex probabilistic models. One advantage of WinBUGs is that it has built in functionality for Bayesian probability theory, which would facilitate updating of a model once new information is collected. Other software such as GoldSim (Golder Associates) is also available for building environmental models and performing Monte Carlo simulations. Some of the same functionality could be developed using R, although some interaction between platforms is also possible (for example, R can call WinBUGs programs). These software products are continually being updated to add functionality, and between them can be expected to provide the necessary capabilities for CADDIS³ in the time frame in which this system might be developed.

The path forward is not clear on how to implement a similar web-based tool. An approach that would seem to have a lot of potential is to use Extensible Markup Language (XML), Extensible Stylesheet Language (XSL), and scalable vector graphics (SVG) for the web-based drawing and storing of conceptual model influence diagrams. Since SVG is a text-based graphic description language implemented in XML both the graphical description and the probability distributions of the influence diagram can be stored and manipulated as an integrated fashion. Using XML/XSL to draw, store, and present the conceptual model also allows the conceptual model to be integrated with the stream information and data analysis described under CADDIS² in a text based non-proprietary standardized form that is transportable between operating systems. An SVG approach could also have potential for visually eliciting and specifying input probability distributions for a conceptual model. Solving the conceptual model influence diagram could in one of a number of programming languages including R, Java, or C++. Alternatively a SVG based conceptual model influence diagram could generate input files for BUGS (the command line version of WinBUGS) to provide the influence diagram solution engine. Also, a new version of Analytica is being written using XML and Java.

The XML and SVG approach could be mixed with drawing tools built in Java or simulation tools such as Hydra (<http://software.biostat.washington.edu/statsoft/MCMC/Hydra>). Hydra is an open-source, platform-neutral library programmed in Java for performing MCMC sampling.

Technical Tools

CADDIS³ would require development of web-based interactive conceptual model influence diagram drawing and solution software. How this might be achieved is not entirely clear at this time, but it is clear that such an approach will be feasible in the time frame of interest because of the rapid advancements that are being made in web-based software and linking to analytical software. Potential solutions that are available now include XML, XSL, SVG, and Java for developing and managing the web-based influence diagram drawing software, as well as, for a user interface to MCMC software such as Hydra and BUGS. MCMC sampling could also require significant computational resources depending on the level of use of CADDIS. Other potential solutions include using a web-based version of Analytica. Given the time frame of interest, what will be important is to keep track of software developments as they occur so that once CADDIS³ is designed, the appropriate software is identified.

Resources Required

Given the uncertainty in the development of CADDIS³, determining the resources required at this time is very difficult. The web-based tools that would be relied upon are developing rapidly and may provide a clearer path forward in the near future. One might expect that a system such as CADDIS³ could be built within 3 years after completion of Option 2 (give or take a couple of years), depending on the features required, and depending on the nature of Option 2 when it is implemented. Also, it would be possible to add functionality of this type in a piecemeal fashion, highlighting each step of the SI process in turn towards development of a complete system. This way, each component could be tested for feasibility before the next component is attempted.

Summary

Three Options for a path forward of the design of CADDIS have been presented. CADDIS¹ provides a simple interface to information sources relevant to the SI guidance along with a capability to download SI worksheets, and to upload completed worksheets into a database that is continually updated. CADDIS¹ will be essentially a static system. CADDIS² will provide dynamic capabilities in terms of on-line implementation of CADDIS, interaction with databases, GIS coverages, and data analysis functionality. Scoring rules can also be built into CADDIS², so that stressor identification will have a numerical basis, but a complete holistic system is reserved for CADDIS³. This final option for CADDIS will also include modeling capabilities so that SI problems can be solved in a fully integrated and coupled system. Option 2 requires that the user solve their SI problem outside of CADDIS, whereas, Option 3 will provide the capability for CADDIS to numerically solve the problem, or will provide a numerical tool that will help managers solve SI applications on-line.

CADDIS¹ and CADDIS² provide valuable SI functionality that could be rapidly developed while laying a foundation for continually improving CADDIS in a resource efficient manner. The development steps for CADDIS³ are unclear at this time but as web-based tools continue to rapidly develop the vision for the path forward should clarify in the near future, or in the time frame in which CADDIS² would be built.

References

U.S. Environmental Protection Agency (USEPA). 2000. Stressor Identification Guidance Document. EPA-8220B-00-025.

Appendix B-1. *Table of potential CADDIS functionality.*

Function	CADDIS Option	Complexity	Level of Effort	Comments
STRESSOR IDENTIFICATION PROCESS				
Stressor Identification Guidance Document online	CADDIS ¹	low	low	Could be easily done.
CADDIS Wizard	CADDIS ²	low	medium	Partially implemented in CADDIS prototype.
DEFINE THE IMPAIRMENT				
Links to guidance on stakeholder involvement	CADDIS ¹	low	high	Requires locating appropriate documents and web sites.
Provide regulatory/programmatic context of impairment Links with regulations and water management programs <ul style="list-style-type: none"> • CWA: 305(b), 303(d) TMDLs links, 319, NPDES, ... • State and Local watershed management links 	CADDIS ¹	low	high	This type of functionality is demonstrated in the CADDIS prototype.
Biological data visualization <ul style="list-style-type: none"> • This will require the user to upload their data implying some pre-specified form • Plotting the data will require a statistical graphics package (R) • Underlay GIS coverages <ul style="list-style-type: none"> • Provide a GIS base map search (EIMS, EnviroMapper, EnviroFacts) • Merge and load GIS data with the impairment data into a statistical graphics package (R) 	CADDIS ²	low	low	Demonstrated in StatWiz and the CADDIS prototype.
	CADDIS ²	high	medium	
	CADDIS ²	high	high	
	CADDIS ²	high	medium	

Function	CADDIS Option	Complexity	Level of Effort	Comments
Gather other biological data <ul style="list-style-type: none"> Local geographic search for other relevant biological data National geographic search for other relevant biological data <ul style="list-style-type: none"> Identify Merge and load biological data into a statistical graphics package 	CADDIS ¹ CADDIS ¹ CADDIS ²	low low high	high high medium	Demonstrated in CADDIS prototype. Demonstrated in StatWiz.
Gather other biological information <ul style="list-style-type: none"> Internet search for relevant biological information 	CADDIS ¹	low	high	There are likely to be many sources for this type of information.
LIST CANDIDATE CAUSES				
Map data on sources and stresses <ul style="list-style-type: none"> Provide a GIS base map search (links to EIMS, EnviroMapper, EnviroFacts) Merge and load maps, sources, and stresses with the impairment data into a statistical graphics package 	CADDIS ¹ CADDIS ²	low high	high high	Demonstrated in EnviroMapper and CADDIS prototype. Demonstrated in StatWiz.
Identify common sources and stressors in impairment area <ul style="list-style-type: none"> Link to State/local water management programs Query State/local water management databases/information 	CADDIS ¹ CADDIS ²	low high	high high	Demonstrated in CADDIS prototype.
Develop Conceptual Model <ul style="list-style-type: none"> Links to guidance on developing conceptual models Links to library of existing conceptual models Wizard for web-based Conceptual model drawing tool 	CADDIS ¹ CADDIS ¹ CADDIS ³	low low high	low medium high	Requires locating sources. Requires building library Difficult but a SVG base approach looks interesting.

Function	CADDIS Option	Complexity	Level of Effort	Comments
Gather other biological information <ul style="list-style-type: none"> Internet search for relevant stressor information 	CADDIS ¹	low	high	Requires access to State/Tribe, Federal, citizen watershed groups and other web sites.
ANALYZE EVIDENCE				
Links to EPA and other guidance on statistics and modeling. This could include links to EPA's Quality program, such as DQOs, DQA, etc.	CADDIS ¹	low	low	Providing links to guidance is simple.
Case studies and Examples links	CADDIS ¹	low	medium	Building the capability is straightforward. But, effort will be required to maintain it as it grows.
Analyze Data <ul style="list-style-type: none"> Implement conceptual model Associate candidate causes with observed effects Associate observed effects with mitigation or manipulation of causes Combine data from site with other information 	CADDIS ²	medium	medium	Possibilities demonstrated in StatWiz.
	CADDIS ³	high	high	Implies a Bayesian approach, difficult
Causal Mechanisms <ul style="list-style-type: none"> Decision Logic depicting candidate causes by dimming eliminated causes, and highlighting remaining causes based on mitigation/manipulation analysis 	CADDIS ²	high	high	Possibilities demonstrated in the StatWiz decision logic approach

Function	CADDIS Option	Complexity	Level of Effort	Comments
Search/link and implement TIE methods & results	CADDIS ¹	low	high	High level of maintenance as more examples are made available.
Search/link to remediation and restoration projects	CADDIS ¹	low	high	High level of maintenance as more examples are made available.
CHARACTERIZE CAUSE				
Eliminate <ul style="list-style-type: none"> • Links to philosophy of approach • Examples links • Upstream/downstream • Time precedence • Site/reference comparison 	CADDIS ¹ CADDIS ¹ CADDIS ¹ CADDIS ¹ CADDIS ¹	low low high high high	low high medium medium medium	Examples will prove to require a high level of effort in the long term. Statwiz functions.
Diagnose <ul style="list-style-type: none"> • Links to philosophy of approach • Examples links • Links to specific diagnostic tools and keys 	CADDIS ¹ CADDIS ¹ CADDIS ¹	low low low	low high low	Much of CADDIS ¹ is simple to implement. The high LOE is because of maintenance.
Strength of evidence analysis <ul style="list-style-type: none"> • Links to philosophy of approach and examples • Link to Causal Considerations • Modeling Wizard • Weighting Wizard 	CADDIS ¹ CADDIS ¹ CADDIS ³ CADDIS ²	low low high high	low low high medium	Links are simple once the sources are identified. Modeling is difficult. Weighting schemes are not difficult but are prone to decision biases.

Function	CADDIS Option	Complexity	Level of Effort	Comments
Identify cause <ul style="list-style-type: none"> • Link to suggested report outline and examples • Generate organized material for report 	CADDIS ¹ CADDIS ²	low medium	medium medium	Requires developing material XML will facilitate implementation Requires defining reporting format and requirements
Evaluate uncertainties and sensitivities <ul style="list-style-type: none"> • Guidance links • Examples links • Probabilistic modeling wizard 	CADDIS ¹ CADDIS ¹ CADDIS ³	low low high	low high high	Sources of information are well known to us. High maintenance. Bayesian approach with web-based influence diagram drawing and simulating tool.
NEXT STEPS				
Iteration Options <ul style="list-style-type: none"> • Links on combining sensitivity analysis and data collection • Link: "Reconsider the impairment: False positives" • Value of information links 	CADDIS ¹ CADDIS ¹ CADDIS ¹	low low low	low low low	Simple links to sources of information. Adding functionality is more complex and will require a greater LOE.

U.S. EPA Information Sources

Prepared by

**MENZIE • CURA & ASSOCIATES, INC.
Environmental Consultants**

**One Courthouse Lane, Suite Two • Chelmsford, Massachusetts 01824
Phone (978) 453-4300 • Fax (978) 453-7260**

**Prepared for T N & Associates, Inc.
under Contract Number 68-C-98-187**

**This document has been reviewed and approved for use as background for
the CADDIS Workshop by the EPA Work Assignment Manager.**

This memo summarizes some of the key features of several information sources developed and/or maintained by Environmental Protection Agency (EPA) that may be useful to users of the Stressor Identification (SI) Guidance Document (USEPA, 2001). For each source we provide:

The information contained in each source that might be useful in the SI process;

The platforms/structure of the database/system; and

Whether the information has a locational component and if so, whether it is georeferenced.

These characteristics are summarized in Table B-1 and are described in more detail in the text below. The information sources we reviewed are listed in alphabetical order, not by their usefulness or application to particular steps in the Stressor Identification Process.

AQUATOX

<http://www.epa.gov/ost/models/aquatox/>

The U.S. EPA has developed a freshwater ecosystem simulation model called AQUATOX that predicts the fate of various pollutants, such as nutrients and organic chemicals, and their effects on the ecosystem, including fish, invertebrates, and aquatic plants. The AQUATOX model can help someone using the SI process to establish the cause and effect relationships between chemical water quality, the physical environment, and aquatic life. It can test sensitivity of biological responses to individual stressors. It may also help to determine the most important of several environmental stressors, e.g. where there are both nutrients and toxic pollutants.

AQUATOX is a process-based, or mechanistic, ecosystem model that simulates the transfer of biomass and chemicals from one compartment of the ecosystem to another. It does this by simultaneously computing important chemical and biological processes over time. AQUATOX can predict not only the fate of chemicals in aquatic ecosystems, but also their direct and indirect effects on the resident organisms. It can represent a variety of aquatic ecosystems, including lakes, reservoirs, ponds, rivers and streams. It also simulates the fate and effects of multiple environmental stressors, such as nutrients, organic toxicants, temperature and turbidity, and changing depths and flow regimes. Understanding these cause and effect relationships can help determine effects that have been shown to be diagnostic of particular stressors (as part of Characterizing Causes in the Stressor Identification Process).

Table B-1. Summary of Information Sources for Use in the Stressor Identification Process

Information sources	Use in SI process	Web Address	Database Platform	Mode of Access	Georeferenced?
AQUATOX	Model cause and effects in water body of concern	http://www.epa.gov/ost/models/aquatox/	Windows-based software	Order from the National Service Center for Environmental Publications or download from web site	No
ASTER	Identify chemical stressors associated with particular effects	http://www.epa.gov/med/databases/aster.html	NA	Request search via e-mail	No
BASINS	Identify stressors, quantify impact of multiple stressors on watershed	http://www.epa.gov/ost/basins/index.html	NA	Free software	Yes, may be used with ArcView GIS software.
ECOTOX	Identify chemical stressors associated with particular effects	http://www.epa.gov/ecotox/	NA	Internet	No
Ecotox Thresholds	Identify chemical stressors	http://www.epa.gov/superfund/resources/ecotox/index.htm	DOS-based software	Download from website	No
Envirofacts	Identify candidate causes (point sources)	http://www.epa.gov/enviro/index_java.html	ORACLE	Internet	Yes
EIMS	Identify candidate causes and links between cause and effect based on case studies	http://www.epa.gov/eims/eims.html	ORACLE	Internet	Yes
EMAP	Identify candidate causes and links between cause and effect based on case studies	http://www.epa.gov/emap/	ASCII text files and ARC/INFO export files	Internet	Yes
EVISTRA	Identify effects associated with particular stressors	http://www.epa.gov/med/databases/evistra.html	NA	NA	NA

Table B-1. (Continued.) Summary of Information Sources for Use in the Stressor Identification Process

Information sources	Use in SI process	Web Address	Database Platform	Mode of Access	Georeferenced?
FIELDS	Delineate spatial distribution of stressor (includes watershed mapping feature)	http://www.epa.gov/region5fields/text/pages/index.html	EQuIS by Earthsoft & database developed by FIELDS Team	Software	Yes
Inventory of Ecological Restoration Projects Within the Mid-Atlantic Integrated Assessment (MAIA) Region	Identify wetland stressors (and possibly links to effects)	http://www.epa.gov/owow/wetlands/restore/ http://yosemite1.epa.gov/water/restorat.nsf/PYPMapPg6?OpenPage	NA	Internet	No
National Water Quality Standards Database (WQSDB)	Identify levels of stressors associated with potential effects	http://www.epa.gov/wqsdatabase/	NA	Internet	Yes. Linked to on-line mapping tool, EnviroMapper.
River Corridor and Wetland Restoration Project Directory	Identify wetland stressors (and possibly links to effects)	http://www.epa.gov/owow/wetlands/restore/ http://yosemite1.epa.gov/water/restorat.nsf/rpd-2a.htm?OpenPage	NA	Internet	No
STORET	Identify stressors in water body of concern (chemical, nutrient, other)	http://www.epa.gov/storet/	UNIX/ ORACLE database server	Internet or free software	Yes
Toxicity/Residue Database	Identify chemical concentration in tissue associated with effects	http://www.epa.gov/med/databases/tox_residue.html	NA	NA	NA
WATERS	Determine status of waterbodies of concern	http://www.epa.gov/waters/	Microsoft Access	Internet	Yes. Link to ArcView or use EnviroMapper

Notes:

The designation “NA” indicates that we were unable to find information on the World Wide Web to fully describe the information source. The desired information could most likely be obtained by talking with the persons who developed or maintain these information sources.

AQUATOX is designed to provide a realistic representation of aquatic ecosystems with a minimal amount of detailed site specific information or site calibration. The following are required input data:

- Loadings to the waterbody (environmental loadings can be from multiple sources)
- General site characteristics
- Chemical characteristics of any organic toxicant
- Biological characteristics of the plants and animals

AQUATOX comes bundled with data libraries that provide default data. This is of particular importance for the biological data, which are probably the most difficult for a user to obtain. AQUATOX also allows the user to evaluate model uncertainty by varying the value or distribution of input parameters, or to test the sensitivity of a given endpoint to specific inputs.

AQUATOX has many potential applications to water management issues and programs, including water quality criteria and standards, total maximum daily loads (TMDLs), and ecological risk assessment. AQUATOX can be used to predict ecological responses to proposed management alternatives. AQUATOX can be helpful where the user needs to understand the processes relating the chemical and physical environment with the biological community and the following conditions exist:

- Where ecological and biological processes are complex
- Where indirect effects are important but difficult to monitor
- When one needs to articulate linkages between nutrients and biotic community
- Where the environmental conditions may change appreciably

AQUATOX is a Windows-based software program that was developed for EPA by Richard A. Park and Jonathan S. Clough of Eco Modeling. The model, users guide, technical documentation and validation reports may be ordered from the National Service Center for Environmental Publications, or they can be downloaded from the web site. The results of several validation studies that have been conducted using AQUATOX are also available on the website.

ASTER (Assessment Tools for the Evaluation of Risk)

<http://www.epa.gov/med/databases/aster.html>

ASTER is a Unix-based computer program that is useful when looking for toxicity information for pollutants that are less commonly found in the environment or whose toxicity is not well studied. This database integrates the toxic effects information from Aquatic Toxicity Information Retrieval (AQUIRE) and the Quantitative Structure Activity Relationships (QSAR) system to develop toxicity profiles for compounds for which no empirical toxicity information is available.

“When empirical data are not available, mechanistically-based predictive models are used to estimate ecotoxicology endpoints, chemical properties, biodegradation, and environmental partitioning. The QSAR system includes a database of measured physicochemical properties such as melting point, boiling point, vapor pressure, and water solubility as well as more than 56,000 molecular structures stored as Simplified Molecular Input Line Entry System (SMILES) strings for specific chemicals.” The results found in the Fathead Minnow database compiled by MED-Duluth are incorporated into the ASTER database.

These toxicity data in this database represent concentrations at which specific biological effects have been observed. The information in the ECOTOX database can help someone using the SI process to identify effects that are associated with exposure to particular chemical stressors.

ASTER is currently not available to the public, but affiliates of the government (local, state, federal, international) or people who are cooperating with or contracted with a government agency can ask technical support staff to conduct a search for them. To request a search, you can e-mail technical support at ecotox.support@epa.gov. Be sure to include the CAS # of the chemical you are interested in. You will receive a toxicity profile that will include the following types of information:

- Chemical Identification - chemical name, CAS number, SMILES and chemical formula
- Environmental Exposure Assessment—molecular weight, melting point, boiling point, vapor pressure, heat vaporization, solubility in water, log P, pKa, adsorption coefficient, Henry’s Law constant, hydrolysis half-life, BOD half life and fugacity
- Ecotoxicological Hazard Assessment—species-specific acute toxicity data, chronic toxicity data, bioconcentration data
- Ecological Risk Characterization—some interpretation of values given in exposure assessment and ecotoxicological hazard assessment sections.
- References

EPA-Duluth hopes to make the ASTER database public sometime in 2002. For more information on ASTER contact the Scientific Outreach Program at (218) 529-5225 or e-mail to ecotox.support@epa.gov.

BASINS (Better Assessment Science Integrating Point and Nonpoint Sources)
<http://www.epa.gov/ost/basins/index.html>

BASINS is a system developed (1) to facilitate examination of environmental information, (2) to provide an integrated watershed and modeling framework, and (3) to support analysis of point and nonpoint source management alternatives. It supports the development of TMDLs, which require a watershed-based approach that integrates both point and nonpoint sources. BASINS can support the analysis of a variety of pollutants at multiple scales, using tools that range from simple to sophisticated. It has been useful in identifying impaired surface waters from point and

nonpoint pollution, wet weather combined sewer overflows (CSO), storm water management issues, and drinking water source protection. It has also been used in urban/rural land use evaluations, animal feeding operations, habitat management practices and as a free resource of GIS and surface water data for the U.S. for schools and educational institutions.

BASINS can be used in the stressor identification process to georeference stressor data (e.g., point source and water quality information) and to quantify the impact of multiple stressors on a watershed.

BASINS is a software package that can be installed on a personal computer. BASINS allows the user to assess water quality at selected stream sites or throughout an entire watershed. In BASINS 3.0, the stream network within the watershed may be determined from Digital Elevation Modeling or preexisting digitized stream networks (e.g., Reach File version 1 (RF1), RF3, NHD, or the Census Tiger Files) may be overlain and burned in. Its databases and assessment tools are directly integrated within an ArcView GIS environment. By using GIS, a user can fully visualize, explore, and query to bring a watershed to life. The simulation models run in a Windows environment, using data input files generated in ArcView. Training courses and on-line tutorials are available to help you learn how to use the BASINS software. The BASINS version 3 system, documentation, and GIS data of watersheds in each of the ten EPA Regions are available through National Service Center for Environmental Publications (NSCEP).

ECOTOX

<http://www.epa.gov/ecotox/>

The ECOTOXicology database is a source for locating single chemical toxicity data for aquatic life, terrestrial plants and wildlife. ECOTOX integrates three toxicology effects databases: AQUIRE (aquatic life), PHYTOTOX (terrestrial plants), and TERRETOX created by the U.S. EPA, Office of Research and Development (ORD), and the National Health and Environmental Effects Research Laboratory (NHEERL), Mid-Continent Ecology Division, Duluth, MN (MED-Duluth).

The U.S. EPA Office of Pesticide Program's Environmental Effects Database (EEDB) of toxic effects data for registered pesticides is also included within ECOTOX via periodic updates from the Office of Pesticide Programs. The EEDB includes toxicity data for aquatic and terrestrial life. These data have been reviewed and categorized as acceptable for fulfillment of pesticide registration and re-registration guideline requirements as explained under FIFRA Subdivision E, Parts 158.145 and 158.150.

These toxicity data represent concentrations at which specific biological effects have been observed. The information in the ECOTOX database can help someone using the SI process to identify effects that are associated with exposure to particular chemical stressors. Users of ECOTOX should consult the original scientific paper cited in the database to ensure an understanding of the context of the data retrieved.

The AQUIRE (AQUatic toxicity Information REtrieval) database was established in 1981. Scientific papers published both nationally and internationally on the toxic effects of

chemicals to aquatic organisms and plants are collected and reviewed for AQUIRE. Independently compiled laboratory data files that include AQUIRE parameters and meet the quality assurance criteria are also included. The majority of aquatic literature that has been reviewed and entered into the system describes studies conducted between 1970 and the present. Toxicity test results and related testing information for any individual chemical from laboratory and field aquatic toxicity tests are extracted and added to AQUIRE. Lethal, sublethal and bioconcentration effects are recorded for freshwater and marine organisms. All AQUIRE data entries have been subjected to established quality assurance procedures. (See description of EVISTRA database and its accompanying guidance document.)

The PHYTOTOX database is a computerized information resource that permits the rapid retrieval and comparison of data pertaining to lethal and sublethal responses, excluding residue effects, of terrestrial plants to the application of chemicals. Both natural and synthetic organic compounds administered to native, crop, or weed species have been considered. Independently compiled laboratory data files that include PHYTOTOX parameters and meet the quality assurance criteria are also included. The PHYTOTOX database was developed as a PC-based program through a joint effort by the University of Oklahoma and the U.S. EPA, Western Ecology Division (formerly Environmental Research Laboratory-Corvallis). PHYTOTOX is currently maintained at MED-Duluth.

TERRETOX is a terrestrial wildlife toxicity database established to provide data linking quantified chemical exposures with observed toxic effects. TERRETOX includes results for lethal, sublethal and bioaccumulation effects. TERRETOX identifies sources of alternative data (domestic or laboratory animal toxicity and bioaccumulation information) when there is a paucity of information on wildlife species. Animals associated with the aquatic environment that do not breathe using gills (e.g., ducks, whales) are included in the TERRETOX database. Results from exposures of terrestrial life stages of amphibians are included in the TERRETOX database. Exposures to the aquatic life stages of amphibians are included in the AQUIRE database component of ECOTOX. Independently compiled laboratory data files that include TERRETOX parameters and meet the quality assurance criteria are also included. The TERRETOX database was developed as a PC-based system by the U.S. EPA, Western Ecology Division and is currently maintained by MED-Duluth.

AQUIRE, PHYTOTOX and TERRETOX may be accessed through the ECOTOX Search Homepage. Access to these databases is open to the public without restriction. Reports can be retrieved as a browser viewable report or as an ASCII delimited file. For more information on any of these databases, contact the Scientific Outreach Program at (218) 529-5225 or E-mail: ecotox.support@epa.gov.

The ECOTOX database has several useful features:

- It is accessible via the World Wide Web
- It is easy to use and includes a detailed Help section
- It is updated quarterly and you can select “view recent update information” at the ECOTOX Search Homepage. Therefore, you have access to up-to-date information.

- Each study is reviewed to see if it meets data quality requirements before it is entered into the database.
- It states the limitations of the data in the database as well as the limitations of the web version of the database itself.

Ecotox Thresholds

<http://www.epa.gov/superfund/resources/ecotox/index.htm>

EPA has developed software that calculates Ecotox Thresholds (ETs) for selected chemicals and can print out a table of ETs and their sources. Ecotox Thresholds (ETs), are defined as “media-specific contaminant concentrations above which there is sufficient concern regarding adverse ecological effects to warrant further site investigation.” ETs are designed to provide Superfund site managers with a tool to efficiently identify contaminants that may pose a threat to ecological receptors and focus further site activities on those contaminants and the media in which they are found. ETs are meant to be used for screening purposes only; they are not regulatory criteria, site-specific cleanup standards, or remediation goals. For those chemicals with the potential to bioaccumulate to toxic levels in upper trophic wildlife (e.g., methyl mercury, PCBs, DDT, dioxins, and lead), these benchmarks may not be low enough at some sites.

The ET software calculates site-specific ETs by adjusting for pH and hardness in surface water and total organic carbon in sediment. The software can also compare the site-specific ETs to the concentrations detected at the site. The January 1996 ECO Update on ETs (EPA 540/F-95/038) describes, in detail, how to calculate and use ETs. This issue of ECO Update is available as a Portable Document Format (PDF) file that may be downloaded from the Ecotox Thresholds website. Someone using the SI process can use these Ecotox Thresholds to develop a list of chemical stressors of concern in the water body he or she is evaluating.

Envirofacts

http://www.epa.gov/enviro/index_java.html

The Envirofacts Warehouse allows you to retrieve environmental information from EPA databases on Air, Chemicals, Facility Information, Grants/Funding, Hazardous Waste, Risk Management Plans, Superfund, Toxic Releases, and Water Permits, Drinking Water, Drinking Water Contaminant Occurrence, and Drinking Water Microbial and Disinfection Byproduct Information [Information Collection Rule (ICR)]. You may retrieve information from several databases at once, or from one database at a time. Online queries allow you to retrieve data from these sources and create reports, or you may generate maps of environmental information by selecting from several Mapping Applications. The locational information in Envirofacts contains all of the latitude and longitude coordinate data for EPA related facilities.

The Envirofacts database is helpful in creating a list of candidate causes of impairment at your site because it allows you to identify nearby pollution sources that may influence the biological conditions at your site.

The Envirofacts database is only available for online querying via the Internet. To access selected data tables from Envirofacts, you can use the online Query forms. To access to all the available data, you can connect to the database directly over the Internet by following the instructions on the "Accessing the Envirofacts Database" webpage. Access to the Envirofacts from the Internet is restricted to registered users. You can complete the Envirofacts Registration Form and e-mail it to enviromail@epa.gov to obtain an EPA UserID and password. Registration is free. Envirofacts is an ORACLE database and can be accessed by any software package that can connect to an ORACLE database.

Environmental Information Management System (EIMS)

<http://www.epa.gov/eims/eims.html>

EPA's Office of Research and Development (ORD) has developed EIMS to organize descriptive information (metadata) for data sets, databases, documents, models, projects, and spatial data. The EIMS design also provides a repository (giant card catalog) for scientific documentation that can be easily accessed with standard Web browsers. EIMS is revised periodically and information about the latest updates is available at the website. Users can search within EIMS to find information sources of interest based upon topic or defined criteria related to types of environmental resources, geographical extent, date, or content origin.

Each metadata entry in EIMS describes the actual data object, and it includes a contact name, e-mail address, and phone number for someone who is knowledgeable about the information resource. Contacts can provide details on how to get an information resource, and may even be able to provide it themselves.

EIMS establishes intra-agency partnerships with projects and Program and Regional offices to build the content of its database. These partnerships promote data and information sharing across EPA. The following partners are storing their metadata in the growing EIMS collection:

ORD Organization Partners

- National Center for Environmental Assessment (NCEA)
- National Center for Environmental Research (NCER)
- National Exposure Research Laboratory (NERL)
- National Health and Environmental Effects Research Laboratory (NHEERL)
- National Risk Management Research Laboratory (NRMRL)
- Office of Resources Management and Administration (ORMA)

ORD Project Partners

- Environmental Monitoring and Assessment Program (EMAP)
- Human Exposure Database System (HEDS)
- Regional Environmental Vulnerability Assessment (ReVA)
- Science Information Management Coordination Board (SIMCorB)

EPA Program Offices

- Office of Environmental Information
- Office of Water Surf Your Watershed Program

EPA Regional Office Partners

EPA Region 2
EPA Region 5
EPA Region 8
EPA Region 9
EPA Region 10

Intra-Agency Partnerships

Global Change Research Program
Great Lakes National Program Office (GLNPO)

Inter-Agency Partnerships

Enviro Science e-Print

EIMS also became EPA's node of the National Spatial Data Infrastructure (NSDI) (<http://www.fgdc.gov/>) as of August 24, 2000. EIMS is a registered Geospatial Clearinghouse Node of NSDI, and as such allows the national network of NSDI Gateway servers to search EPA-EIMS for geospatial metadata at EPA. The EIMS database supports the full Geospatial Metadata Content Standard developed by the Federal Geographic Data Committee (FGDC). NSDI is an excellent opportunity for scientists to publish, share, and use digital geographic information over the Internet.

This database may be used in several steps in the Stressor Identification Process. It can be used to identify common stressors found in estuaries systems by accessing metadata from EPA's Environmental Monitoring and Assessment Program (EMAP). EIMS may also help you analyze the evidence you've collected for each candidate cause to determine which one is likely causing the observed impairment at your site. For example, EIMS can be used to locate other studies that found certain associations between candidate causes and effects or suggest ways that you can analyze your site-specific data to establish associations. Different EPA regions, programs or offices may also post results from studies of the effect of deliberately eliminating or reducing a candidate cause (e.g., a field experiment, lab experiment or regulatory or remedial action) to establish cause and effect relationships. Finally, EIMS can be used to analyze the strength of evidence for each candidate cause at your site, for example, in terms of whether the hypothesized cause and effect relationship at your site is similar to any well-established cases.

The EIMS application was developed to operate with the web browser Netscape. Netscape and ORACLE, the database platform used for EIMS, follow the open standard for use of Java. The proprietary Java standard employed by Microsoft Internet Explorer is not compatible for all operations of EIMS.

If you want to add new metadata to EIMS, you can register with EIMS by providing your name, phone number, and email address. A simple on-line form steps you through the registration process and on-line forms guide you through the data input process.

Environmental Monitoring and Assessment Program (EMAP)

<http://www.epa.gov/emap/>

“The Environmental Monitoring and Assessment Program (EMAP) is a research program to develop the tools necessary to monitor and assess the status and trends of national ecological resources (see EMAP Research Strategy). EMAP's goal is to develop the scientific understanding for translating environmental monitoring data from multiple spatial and temporal scales into assessments of ecological condition and forecasts of the future risks to the sustainability of our natural resources. EMAP's research supports the National Environmental Monitoring Initiative of the Committee on Environment and Natural Resources (CENR). ” [A database of Federal Environmental Research and Monitoring Programs in the Mid-Atlantic Region can be queried at the national Environmental Monitoring Initiative website at:

<http://www.epa.gov/monitor/owsquery/>]

One of the goals of EMAP is to develop and demonstrate indicators to monitor the condition of ecological resources. EMAP funds academic research into ecological indicators and other topics through STAR grants (see <http://es.epa.gov/ncercqa/grants/>). Therefore, the information on this website may be used identify common stressors found in aquatic systems that are affecting the condition of ecological resources (Step?/Chapter 2 of the Stressor ID process). It may also help you characterize causes of an observed impairment at your site (Step?/Chapter 4 of the Stressor ID process) by considering the results from EMAP projects.

You can access EMAP metadata from the EIMS website described above, but the EMAP website contains additional information such as program information, data and metadata files, and publications. Data are formatted in two EPA standards: ASCII text files and ARC/INFO export files. Many files available for download from the EMAP website are compressed or ZIP'ed to save disk space. The EMAP website was created for best viewing with Netscape 4+ and Microsoft Internet Explorer 4+ web browsers; however, it should still be reasonably functional and readable with other browsers and earlier versions of Netscape and Internet Explorer.

EVISTRA (EValuation and Interpretation of Suitable Test Results in AQUIRE)

<http://www.epa.gov/med/databases/evistra.html>

The purpose of this database is to present results that were obtained from aquatic toxicity tests on selected chemicals and were evaluated for suitability and quality and interpreted when necessary.

“Supplementary information (e.g., results of quality evaluations, taxonomic and geographic information concerning test species, etc.) and software can be accessed to facilitate use of test results in EVISTRA.” Also as part of the EVISTRA effort, a draft guidance document titled "Guidance for Evaluating Results of Aquatic Toxicity Tests" has been written at MED-Duluth.

The database and guidance document can help the user identify effects associated with particular stressors as well as learn how to evaluate the quality of aquatic toxicity test data obtained from other sources (e.g., journal articles, technical reports).

FIELDS (Fully Integrated Environmental Location Decision Support) system
<http://www.epa.gov/region5fields/text/pages/index.html>

The FIELDS system is a cooperative project between the USEPA Water Division and Superfund Division. The Fully Integrated Environmental Location Decision Support (FIELDS) Team's mission is to identify, assess, communicate and help solve priority environmental problems in specific geographic areas. To achieve their mission, the FIELDS team has developed a process to assist environmental organizations who are seeking better and more efficient ways to investigate, communicate and help solve environmental problems.

Software has also been developed by the FIELDS team to aid site investigators in the following areas:

Base Mapping Help—Obtaining a good base map is an important first step to be able to display spatial data and provide a geographical reference for the area under investigation. This is also listed as a key first step in the Stressor Identification Guidance Document.

Sampling Design Module—This part of the software allows the user to create sampling plans in GIS and then export the latitude and longitude information to a GPS unit in order to conduct the field sampling for the project.

Data Entry and Storage—FIELDS uses two relational databases for environmental data. One is a database developed by the FIELDS team for sediment and fish. The other is the EquiS database developed by Earthsoft which is used for groundwater well, surface water and soil data.

Modeling—“The FIELDS modeling tools allow discrete sampling data points to be interpolated into a surface. Important uses of these interpolated surfaces include delineating hot spots, calculating average concentrations, estimating contamination mass and volumes, and developing post-remediation scenarios.”

Analysis—The Analysis Tools FIELDS extension currently consists of three tools: Average of Grid, Estimation Error, and Mass and Volume Calculation. These tools are used in conjunction with the modeling results to allow decision-makers to isolate any particular contaminant, level of contamination, or any other characteristic of the site they may wish to analyze. The tools for quantifying these analyses include 2D and 3D imaging as well as traditional text-based data reports.

The FIELDS system could be particularly helpful in visualizing available stressor data at very large sites or at sites with multiple stressors. The FIELDS system has been implemented at more than 20 sites in the U.S. to-date. A Beta version of the FIELDS system software is now available to Federal, State and Tribal agencies.

National Water Quality Standards Database (WQSDB)

<http://www.epa.gov/wqsdatabase/>

This website provides access to several WQS reports that provide information about designated uses, waterbody names, State numeric water quality criteria, and EPA recommended numeric water quality criteria. The WQSDB allows users the ability to compare WQS information across the Nation using standard reports. The website contains EPA's compilation of the State, Territory and authorized Tribal water quality standards effective as of May 30, 2000. The site is updated periodically to reflect EPA approved revisions and additions to State and Tribal water quality standards.

Currently, the database includes designated use information for 16 states and numeric water quality criteria for one pilot state (Missouri). Work is progressing to include this information for the remaining States, Tribes, and Territories in the U.S.

This database can be used when analyzing the available evidence to establish cause and effect relationships in the SI process (Step?/Chapter 3). "Water quality criteria developed under section 304(a) are based on data and scientific judgments on the relationship between pollutant concentrations and environmental and human health effects"

<http://www.epa.gov/waterscience/standards/critsum.html>).

River Corridor and Wetland Restoration

<http://www.epa.gov/owow/wetlands/restore/>

There are two directories/databases of restoration projects in the United States that are accessible from this website:

Inventory of Ecological Restoration Projects Within the Mid-Atlantic Integrated Assessment (MAIA) Region

<http://yosemite1.epa.gov/water/restorat.nsf/PYPMMapPg6?OpenPage>

River Corridor and Wetland Restoration Project Directory

<http://yosemite1.epa.gov/water/restorat.nsf/rpd-2a.htm?OpenPage>

The USEPA's Office of Research and Development, working jointly with the Office of Water, has developed an Internet-accessible inventory of ecological restoration projects within the Mid-Atlantic Integrated Assessment (MAIA) Region. The inventory encompasses individual restoration projects conducted in both aquatic and terrestrial environments by federal, state, and local government entities and non-government organizations. Each project listing includes such information as a description of the restoration site, the problem being addressed, goals of the project, critical ecosystem/environmental factors, technical approaches, costs, monitoring activities, and a contact for more detailed information. The purpose of the inventory is to provide a central repository of restoration project information specific to the MAIA Region. It is intended to be useful to researchers, restoration practitioners, environmental resource managers, policy makers, consultants, and communities within the Mid-Atlantic area. All project information is provided by the organizations conducting the restoration activities.

This inventory has been integrated with the Office of Water's River Corridor and Wetland Restoration database and website. The River Corridor and Wetland Restoration Project Directory lists all ecological restoration projects by State (including projects in the MAIA Region) and is also accessible via the Internet. The following types of information are provided for each project:

Location (City/Town and country)	Project Number
Abstract	Adjacent Land Use
Date Provided	Greater Surrounding Area Land Use
Is it a MAIA project?	Descriptive Characteristics of Project Area
Project Identification	Restoration Goals
Lead Restoration Contact	Restoration Technique/Approach
Project Status	Project Outlook
Project Characterization	Restoration Site Description
Principal Funding Organizations	Restoration Projects/Results, Reports and Articles
Type of Funding Organization	Lessons Learned and Assistance Desired
Main Monitoring Focus	General Comments on Lessons Learned
Invasive Species	

Someone using the stressor identification process could use these databases to identify some typical wetland stressors and how these stressors were mitigated to restore wetland function.

STORET (STorage and RETrieval)

<http://www.epa.gov/storet/>

STORET (data STORage and RETrieval system) is a repository for water quality, biological, and physical data. This database uses reach numbers to help users query, analyze, and map water quality related issues. The U.S. Environmental Protection Agency (EPA) maintains two data management systems containing water quality information for the nation's waters: the Legacy Data Center, and STORET.

The Legacy Data Center, or LDC, contains historical water quality data dating back to the early part of the 20th century and collected up to the end of 1998. STORET contains data collected beginning in 1999, along with older data that has been properly documented and migrated from the LDC.

Both systems contain site-specific raw biological, chemical, and physical data on surface and ground water collected by federal, state and local agencies, Indian Tribes, volunteer groups, academics, and others. Data are submitted by many different organizations and data posting is voluntary. All 50 States, territories, and jurisdictions of the U.S., along with portions of Canada and Mexico, are represented in these systems. The data in LDC and STORET is not reviewed by EPA, so the agency makes no warranty regarding the accuracy of the environmental data stored in those databases. There are no data quality requirements for submittal, however all data kept on the LDS/STORET web site is fully documented with respect to field and laboratory methods employed, monitoring technologies used, project goals and plans, and QA/QC practices of the submitting organizations. Questions concerning the specific data of any organization are best directed to the organization from whom the data were obtained.

Each sampling result in the LDC and in STORET is accompanied by information on where the sample was taken (latitude, longitude, state, county, Hydrologic Unit Code and a brief site identification), when the sample was gathered, the medium sampled (e.g., water, sediment, fish tissue), and the name of the organization that sponsored the monitoring. In addition, STORET contains information on why the data were gathered; sampling and analytical methods used; the laboratory used to analyze the samples; the quality control checks used when sampling, handling the samples, and analyzing the data; and the personnel responsible for the data.

These databases can be used in the Stressor Identification Process to help identify any existing stressors in the waterbody of concern. Users can, for instance, input a reach number and "ask" STORET to return all point source dischargers, water quality monitoring stations, drinking water intakes, etc. upstream of that Reach. This provides users with a powerful tool for identifying what environmental data is available and also the location of that data relative to their Reach of concern.

LDC or STORET data in the EPA Data Warehouse are located on a UNIX/Oracle database server that may be browsed or downloaded by anyone using a web browser such as Netscape® or Microsoft IE® (Internet Explorer) by clicking on "Obtaining Water Quality Data" on the main STORET homepage. No user registration is required for this type of access, and with the exception of data marked by its owner as "preliminary", all data in STORET will be freely available to the public. You can query the LDC or STORET databases on the Internet and generate data reports for download in html, pdf or txt format. The data from a STORET query can also be downloaded in "Tilde Delimited" flat ASCII text format.

You may also order a free copy of STORET software to create your own data base of water quality data. (Call 1-800-424-9067 or send an email to STORET@epa.gov for your free copy of STORET software). The STORET software along with ORACLE database engine (which must be obtained separately) will allow you to create and update your own STORET database and, if you choose, export your data to the EPA Data Warehouse, where it will be made available to the general public.

Toxicity/Residue Database

http://www.epa.gov/med/databases/tox_residue.html

This database contains concentrations of inorganic and organic chemicals in tissue of aquatic organisms that are associated with effects to the organism. This database can be used to investigate hypotheses related to effect/residue relationships (e.g., body burdens of certain chemicals shown to be associated with particular effects). "The database contains more than 3,000 effect and no-effect endpoints for survival, growth and reproductive parameters for invertebrates, fish and aquatic life-stage of amphibians. Data were abstracted from approximately 500 literature references on approximately 200 chemicals and 190 freshwater and marine test species. Survival endpoints account for about 74% of the total amount of data, with growth and reproduction accounting for 19 and 7% respectively."

U.S. EPA Mid-Continent Ecology Division

The U.S. EPA's Mid-Continent Ecology Division, part of the Office of Research and Development and located in Duluth, MN (MED-Duluth), has developed the following databases: AQUIRE, ASTER, ECOTOX, EVISTRA, Fathead Minnow, PHYTOTOX, TERRETOX, Toxicity/Residue. All of these databases provide information to link exposure concentrations with biological effects. Each database is described separately in this memo.

WATERS (Watershed Assessment, Tracking & Environmental Results)

<http://www.epa.gov/waters/>

WATERS links several state agency and EPA databases in one place to provide information on:

- The uses that waterbodies have been designated for, such as drinking water supply, recreation, or fish protection. These designated uses are part of a state's water quality standards, provide a regulatory goal for the waterbody and define the level of protection assigned to it. This information is available in the National Water Quality Standards Database (see description above).
- Those waterbodies listed by the state as impaired under Section 303(d) of the Clean Water Act, and for those waterbodies, the status of control actions known as TMDLs (calculations of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and an allocation of that amount to the pollutant's sources). This information is also available in the TMDL Tracking System.

The next release of WATERS (Fall 2001) will include:

- Drinking water intakes
- Recreational beaches
- No discharge zones
- Water quality monitoring stations (STORET LDC)
- Water quality assessments from Section 305(b) of the Clean Water Act

Other water information that may eventually be included in WATERS include:

- 319 Non-point source projects (GRTS)
- Facility outfall locations
- Source water protection boundaries
- Fish Consumption Advisories

- Water related National Priority List (NPL) locations
- Total Maximum Daily Load delineation boundaries
- Combined sewer overflows

State and federal water quality managers, as well as interested citizens, can use WATERS to quickly identify the status of individual waterbodies of interest to them. It can also be used to generate summary reports on all waters of a state. WATERS uses EnviroMapper, the EPA standard mapping application, to display this information. EnviroMapper for Water is a web-based Geographic Information System (GIS) application that dynamically displays information about bodies of water in the U.S. This interactive tool allows you to create customized maps portraying the nation's surface waters along with a collection of environmental data using your web browser. The application can be used to view environmental information from the national level down to community level (within one mile), and provides the ability to pan, zoom, label and print maps. You also can link to text reports after identifying a specific waterbody of interest. A remaining question is whether the user can access any of the raw data displayed by EnviroMapper (e.g., georeferencing data or water quality data) directly from WATERS database.

The National Hydrography Dataset, or NHD, is the nationally consistent waterbody network that serves as the foundation and common “language” for WATERS. In the 305(b) database that was created before the broad use of GIS, waterbodies were often defined as small watersheds. This ended up being problematic because water quality assessment data stored in the database often only applied to portions of these watershed waterbodies. An advantage of the WATERS database is that it uses the National Hydrography Dataset Reach Indexing Tool (NHD-RIT), which is an interactive GIS application that allows users to georeference surface water data using the National Hydrography Dataset (NHD). This allows all water quality data to be georeferenced based on a consistent standard (the NHD). The NHD-RIT can be used by any organization that has a mandate to monitor and track surface water information. Users may use the NHD-RIT with ArcView to visualize the data or use WATERS as a web-based alternative for displaying the data in the NHD.

Currently there are three databases that are used in conjunction with the Reach Indexing Tool (RIT) and the National Hydrography Dataset (NHD):

TMDL Tracking System: this database is used by EPA to maintain information about the States TMDL lists.

Contact: Chris Laabs

Download: <http://www.epa.gov/owow/tmdl/trcksys.html>

Water Quality Standards Database (WQSDB): contains designated use information as it was assigned by the States.

Contact: Bill Kramer

WQS Reach Indexing Website: WQS Review Site

305(b) Assessment Database (ADB): is a database used by the States to maintain their surface water under the 305(b) Clean Water Act.

Contact: Dwane Young

Download: <http://water305b.rti.org>

Currently only the TMDL Tracking System and the Water Quality Standards Database are linked to WATERS, but the Assessment Database (ADB) will be integrated into the next release of WATERS in the Fall of 2001. All states assess their individual waterbodies for degree of designated use support (e.g., “fully supporting aquatic life; not supporting primary contact recreation”). If a waterbody’s uses are impaired, the stressors and sources of impairment are also determined (e.g., “causes/stressors are nutrients and sediment; sources are urban runoff and row crop agriculture”). The ADB is designed to make tracking and reporting this data accurate, straightforward and user- friendly for participating states, territories, tribes and basin commissions. It also enables users to meet the requirements of Section 305(b) of the Clean Water Act. All three databases can help develop a list of candidate causes in the SI Process. The National Water Quality Standards Database can also be used when analyzing available evidence to determine the cause of observed impairment using the SI process.

Summary:

Table B-2 summarizes how these information sources can be used in each step of the stressor identification process. The information sources we have reviewed are most useful in developing a list of stressors, visualizing available data, obtaining toxicological effects data for chemical stressors, and locating case studies to which the user can compare his or her site. Additional resources are needed for identification of physical stressors (e.g., hydrological changes) and the effect of nutrients on aquatic systems. It would also be useful to have a software program that may be used in conjunction with the Stressor Identification Guidance to construct a site conceptual model. We have provided a list of additional, non-EPA websites that may prove useful in the stressor identification process (Table B-3). However, it was beyond the scope of this task to evaluate each of them as thoroughly as we have for the EPA information sources in this memo.

Table B-2. Uses of EPA Information Sources in the Stressor Identification Process

Information sources	List candidate causes	Analyze the evidence	Characterize Causes
AQUATOX			X
ASTER		X	
BASINS	X		
ECOTOX		X	
Ecotox Thresholds	X		
Envirofacts	X		
EIMS	X	X	X
EMAP	X	X	
EVISTRA		X	
FIELDS		X	
Inventory of Ecological Restoration Projects Within the Mid-Atlantic Integrated Assessment (MAIA) Region	X		
National Water Quality Standards Database (WQSDB)		X	
River Corridor and Wetland Restoration Project Directory	X		
STORET	X		
Toxicity/Residue Database		X	
WATERS	X	X	

Table B-3. Internet Sources of Stressor Identification Information

URL	Subject
http://www.nal.usda.gov/wqic/	Agriculture Dept Water Quality Info Center
http://www.wes.army.mil/	Army Corps of Engineers' Waterways Experimental Station (WES)
http://www.wes.army.mil/el/arams/arams.html	Army Risk Assessment Modeling System
http://atsdr1.atsdr.cdc.gov:8080/atsdrhome.html	ASTDR Home Page
http://www.atsdr.cdc.gov/toxfaqs.html	ASTDR ToxFAQs
http://www.atsdr.cdc.gov/mrls.html	ATSDR's Minimal Risk Levels (MRLs) for Hazardous Substances
http://www.epa.gov/ost/basins/index.html	BASINS (Better Assessment Science Integrating Point and Nonpoint Sources)
http://www.chemfinder.com/	Chemfinder
http://info.cas.org/	Chemical Abstract Service
http://www.wes.army.mil/el/t2dbase.html	Environmental Residue-Effects Database (ERED), Biota Sediment Accumulation Factor (BSAF) Database, Ocean Disposal Database
http://www.chesapeakebay.net/bayprogram/	EPA Chesapeake Bay Program
http://www.epa.gov/superfund/resources/ecotox/index.htm	EPA Ecotox Thresholds
http://www.epa.gov/emap/	EPA Environmental Monitoring and Assessment
http://www.epa.gov/wqsdatabase/	EPA National Water Quality Standards Database (WQSDB)
http://www.epa.gov/ow/soft.html	EPA Office of Water Databases and Software
http://www.epa.gov/region5fields/text/pages/index.html	EPA Region 5 - FIELDS (Fully Integrated Environmental Location Decision Support) system
http://www.epa.gov/med/databases/tox_residue.html	EPA Toxicity/Residue Database
http://www.epa.gov/ost/models/aquatox/	EPA's Aquatox Release 1 - A Simulation Model for Aquatic Ecosystems
http://www.epa.gov/med/databases/aster.html	EPA's ASTER Database for Ecotoxicological Profiles
http://www.epa.gov/medecotox/	EPA's ECOTOX Database
http://www.epa.gov/enviro/index_java.html	EPA's Envirofacts Data Warehouse and Applications
http://www.epa.gov/eims/eims.html	EPA's Environmental Information Management System (EIMS)
http://www.epa.gov/med/databases/evistra.html	EPA's EVISTRA (EVALuation and Interpretation of Suitable Test Results in AQUIRE) Database
http://yosemite1.epa.gov/water/restorat.nsf/PYPMapPg6?OpenPage	EPA's Inventory of Ecological Restoration Projects Within the Mid-Atlantic Integrated Assessment (MAIA) Region
http://yosemite1.epa.gov/water/restorat.nsf/rpd-2a.htm?OpenPage	EPA's River Corridor and Wetland Restoration Project Directory
http://www.tera.org/	Toxicity Excellence for Risk Assessment
Table 3 continued.	
URL	Subject
http://es.epa.gov/ncerqa/grants/	EPA's Science to Achieve Results (STAR) Program

Table B-3 cont.

URL	Subject
http://www.epa.gov/storet/	EPA's STORET (STORage and RETrieval) Database
http://www.epa.gov/waters/	EPA's WATERS (Watershed Assessment, Tracking & Environmental Results)
http://www.epa.gov/watertrain/	EPA's Watershed Academy Web - Distance Learning Modules on Key Watershed Management Topics
http://irptc.unep.ch/irptc/invent/igo.html	Inventory of Information Sources on Chemicals
http://www1.nature.nps.gov/datasci/	National Park Service - Data on Park Resources
http://www1.nature.nps.gov/toxic/	National Park Service Environmental Contaminants Encyclopedia
http://response.restoration.noaa.gov/cpr/sediment/squirt/squirt.html	NOAA Screening Quick Reference Tables for inorganic/organic contaminants in environmental media
http://www.opsd.nos.noaa.gov/tp4days.html	NOAA Tide Predictions
http://www.esd.ornl.gov/programs/ecorisk/ecorisk.html	Oak Ridge National Laboratory (ORNL) Ecological Risk Analysis: Tools and Applications
http://risk.lsd.ornl.gov/homepage/eco_tool.shtml	ORNL- Risk Assessment Information System- Ecological Benchmarks
http://www.riskworld.com/NEWS/99q1/nw9aa081.htm	Risk World Index of Endocrine Disruptor Resources
http://www.toxnet.nlm.nih.gov/	TOXNET
http://www.usgs.gov/	USGS
http://www.best.usgs.gov/	USGS Biomonitoring of Environmental Status and Trends (BEST) Program Databases
http://www.nbii.gov/	USGS National Biological Information Infrastructure
http://www.nwrc.gov/index.html	USGS National Wetlands Research Center
http://www.nwrc.gov/publications/specindex.html	USGS National Wetlands Research Center - Species Profiles
http://www.pwrc.usgs.gov/	USGS-Patuxent Wildlife Research Center
http://fwie.fw.vt.edu/WWW/macsis/index.htm	VA Tech-Marine and coastal species information system-fish and wildlife information exchange
http://www.vims.edu/env/research/risk/software/other_software.htm	VIMS Environmental Toxicology and Risk Assessment - Software and Documentation
http://www.ecy.wa.gov/services/as/iip/eim/index.html	Washington State Department of Ecology's Environmental Information Management (EIM) System
http://www.terrene.org/wirsdata.htm	Watershed Information Resource System (WIRS) Database (in partnership with EPA's Clean Lakes Program)
Key:	
Shading indicates that we have described this resource in our memo.	

DRAFT

**Additional U.S. EPA Information Sources:
Mid-Atlantic Focus**

Prepared by

MENZIE • CURA & ASSOCIATES, INC.

Environmental Consultants

One Courthouse Lane, Suite Two • Chelmsford, Massachusetts 01824

Phone (978) 453-4300 • Fax (978) 453-7260

Prepared for T N & Associates, Inc.

under Contract Number 68-C-98-187

**This document has been reviewed and approved for use as background for
the CADDIS Workshop by the EPA Work Assignment Manager.**

DRAFT

This memo summarizes some of the key features of several information sources for the mid-Atlantic region that may be useful to users of the Stressor Identification Guidance Document (USEPA, 2001). For each source we provide:

- The information contained in each source that might be useful in the Stressor Identification (SI) Process
- A website or other means to access the information
- The type of information and format in which it is available
- Whether the information is georeferenced.

These characteristics are summarized in Table B-4 and are described in more detail in the text below. The information sources we reviewed are listed in alphabetical order, not by their usefulness or application to particular steps in the SI Process.

Alliance Citizen Monitoring Database

<http://www.acb-online.org/monitoring/site.cfm>

The Alliance Citizen Monitoring program is a regional network of more than 145 trained volunteers who perform weekly water quality tests that help track the condition of waters flowing toward the Chesapeake Bay. The data collected by these volunteers has been compiled in a database containing monitoring site information, monitoring event data and water quality data collected from 8/17/1985, to the present for just under 200 waterbodies, across the Chesapeake region in Pennsylvania, Maryland and Virginia. The database also contains jellyfish data from 1/14/2001, to the present. The database is updated continuously and includes data for waterbodies ranging from small creeks to large rivers. This data can be queried on the website via a series of prompts. The user can then download data as text, tab-delimited files or view graphical representations of the data for a particular monitoring location.

The information in this database can be used to find common sources and stressors in the mid-Atlantic region as part of the “List Candidate Causes” step in the Stressor Identification Process. More information about the Alliance Citizen Monitoring program is available at <http://www.acb-online.org/citmon.htm>.

DRAFT

Table B-4. Summary of Information Sources for the Mid-Atlantic Region For Use in the Stressor Identification Process

Information source	Use in SI Process	Web Address	Data Type and Format	Georeferenced?
Alliance Citizen Monitoring Database	List Candidate Causes	http://www.acb-online.org/monitoring/site.cfm	Can select data to download in text, tab-delimited format or graph data from the webpage.	Yes
Chesapeake Information Management System (CIMS)	List Candidate Causes	http://www.chesapeakebay.net/cims/index.htm	Can view metadata and watershed profiles on-line.	Yes
Delaware Department of Natural Resources and Environmental Control Environmental Navigator	List Candidate Causes	http://www.dnrec.state.de.us/enweb/	Access to base map of DE and web-based interactive mapping application linked to a database of information on contaminant source sites and ambient environmental monitoring sites.	Yes
Delaware Stream Watch	List Candidate Causes	http://www.delawarenaturesociety.org/swtechmon.htm	Excel spreadsheets & hard copies available upon request.	Yes
Delaware Surface Water Quality Management	List Candidate Causes & Analyze Evidence	http://www.dnrec.state.de.us/dnrec2000/Library/Water/305breport.htm and http://www.dnrec.state.de.us/dnrec2000/Divisions/Water/WaterQuality/Standards.htm	Can download State of Delaware's 305(b) reports and Surface Water Quality Standards as PDF files.	No
Environmental Monitoring and Assessment Program (EMAP) Data	Define the Impairment & List Candidate Causes	http://www.epa.gov/emap/html/datal/surfwatr/data/	Can download data (ASCII) and metadata (PDF) from the web	Yes
Maryland Biological Stream Survey	Define the Impairment & List Candidate Causes	http://mddnr.chesapeakebay.net/mbss/search.cfm	Can view data on the website or download 1995-1997 data as text, tab-delimited files	Yes
Maryland Department of Natural Resources –Streams	List Candidate Causes & Analyze Evidence	http://www.dnr.state.md.us/streams/pubs/pub_list.html	Some technical reports are available for download in PDF format.	No
Maryland Water Monitoring Council	Define the Impairment	http://www.mgs.md.gov/mwmc/	Can download MS Access database of water monitoring projects.	Yes (some)
Mid-Atlantic Highlands Coordinating Council; Potomac River Basin Pilot Project	Define the Impairment & List Candidate Causes	http://ael.er.usgs.gov/groups/gis/potomac/mahaw.htm	Can view metadata online & save as HTM or TXT file	No

DRAFT

Table B-4. Summary of Information Sources for the Mid-Atlantic Region For Use in the Stressor Identification Process

Information source	Use in SI Process	Web Address	Data Type and Format	Georeferenced?
Mid-Atlantic Integrated Assessment (MAIA) Project	List Candidate Causes	http://md.usgs.gov//maia/ and http://www.epa.gov/maia/assets/pdf/MAHASStreams.pdf	Can download data in ASCII, PDF and ArcInfo format from the web.	Yes
National Environmental Monitoring Initiative – Mid-Atlantic Inventory	Define the Impairment & List Candidate Causes	http://www.epa.gov/cludygxb/site-mid.html	List of available monitoring data and contacts for monitoring programs	No
PA Department of Environmental Protection – Water Quality Assessment and Standards	List Candidate Causes & Analyze Evidence	http://www.dep.state.pa.us/dep/deputate/watermgt/wqp/wqstandards/wqstandards.htm and http://www.dep.state.pa.us/dep/subject/eqb/2000/June20/fch93Annex.pdf	Can download State of Pennsylvania’s 305(b) reports in PDF format	No
Pennsylvania Senior Environmental Corps (PaSEC)	List Candidate Causes	http://www.environmentaleducation.org/default.lasso	Searchable on-line by “registering”. Can view data tables in your web-browser	Yes
Pennsylvania Spatial Data Access (PASDA)	Define the Impairment & List Candidate Causes	http://www.pasda.psu.edu/access/download.shtml	Free, downloadable data from web or all data on PASDA can be obtained on a CD-ROM for a fee (multiple formats).	Yes (most)
SRBC Nutrient Assessment Program	List Candidate Causes	http://www.srbc.net/nutrientdata.htm http://www.srbc.net/trends.htm http://www.srbc.net/nutrientloads.htm	Can view summarized data on web or download data in comma delimited text file, PDF or Excel format	Yes
USGS National Water Quality Assessment (NAWQA) Program	Define the Impairment & List Candidate Causes	http://orxddwimdn.er.usgs.gov/servlet/page?_pageid=543&_dad=portal30&_schema=PORTAL30	Can view or download data in several formats (XLS, WKS, TXT, CSV, HTM)	Yes
Virginia Department of Environmental Quality	List Candidate Causes	https://www.deq.state.va.us/webapp/wqm_homepage	Can query data by county/city or stream and view it on web	Yes
West Virginia DEP, Division of Water Resources, Watershed Assessment and Strategic Planning Section	Define the Impairment & List Candidate Causes	http://www.dep.state.wv.us/wr/index.cfm?page=OWR_Website/C_and_D/WAP.htm	Can download data summaries in PDF format	?

DRAFT

Chesapeake Information Management System (CIMS)
<http://www.chesapeakebay.net/cims/index.htm>

“The Chesapeake Information Management System (CIMS) is an organized, distributed library of information and software tools designed to increase basin-wide public access to Chesapeake Bay information. The Internet sites in CIMS are maintained by CIMS Partners who have enacted Memoranda of Agreement (MOA) to follow certain guidelines for assembling and publishing Chesapeake Bay related information.” Internet applications that can be accessed from the CIMS website and may be useful in the Stressor Identification Process include:

Watershed Profiles

“Watershed Profiles assembles maps, charts and information that portray the environmental condition of Chesapeake Bay watersheds. It operates at a variety of scales from the entire 64,000 square mile Chesapeake Bay watershed to small tributary watersheds. Information on landscape changes, Bay Program activities, other organization activities, and places to visit are some of the information displayed in easy to read charts, maps and tables.” This application would be very useful to gather information about common sources and stressors (landscape, nutrients and toxins) in large and small watersheds (e.g., watersheds for individual creeks and rivers). This information could help “List Candidate Causes” in the Stressor Identification Process.

CIMS Search

The website listed above also provides a link to a Search Engine for searching the CIMS metadata database and COMET (CIMS Online Metadata Entry Tool) which is used to maintain the database. Metadata is available for water bodies of all sizes and data types include databases, data sets, GIS coverage, maps, spatial data bases, spatial datasets, and reach files to name a few. This web page can be used to locate data sources that may be useful to map your data or list stressors and sources in the “List Candidate Causes” step of the Stressor Identification Process.

Delaware Department of Natural Resources and Environmental Control (DNREC)
Environmental Navigator
<http://www.dnrec.state.de.us/enweb/>

“The Environmental Navigator is a web-based interactive mapping application linked to a database of information on potential, possible, and known contaminant source sites and on ambient environmental monitoring sites in the state of Delaware”. It is very similar to EPA’s Enviromapper application, but is specific to the state of Delaware. It will eventually include additional information such as aerial photos, wetlands, land use, biodiversity indices, soil types, air and water ambient monitoring results, and facility violations. A base map of Delaware is available from this web page. (Note: This application does not work in Netscape (something to do with Java.)

DRAFT

“The Environmental Navigator includes summary information on over 8000 sites of more than twenty types. Information on each site includes site name, ID number, status, location (XY, county, basin, watershed), and a hyperlink to more information about that site if it exists. Potential contaminant source sites also include a Contaminant Source Potential Rating and monitoring status information by media and contaminant class.”

The information in the environmental navigator can be used to find and map common sources and stressors in Delaware as part of the “List Candidate Causes” step in the Stressor Identification Process. Maps showing user-selected data can be saved as JPG files.

Delaware Stream Watch

<http://www.delawarenaturesociety.org/swtechmon.htm>

The Technical Monitoring program of Delaware Stream Watch is a collaboration between the Delaware Nature Society and the Delaware Division of Natural Resources and Environmental Control. Volunteers monitor assigned sites on a monthly basis, testing for dissolved oxygen, pH, alkalinity, nitrate, conductivity, salinity (in tidal reaches) and temperature. Additional procedures ensure quality control in sampling techniques. The data collected by Stream Watch volunteers is put into the state's database and copies are maintained (in Excel spreadsheets) at the Delaware Nature Society. Data collected by Stream Watch in the past 6 years is for the Christina Basin, which includes the watersheds of the Brandywine Creek, Red Clay Creek, White Clay Creek and the Christina River. To access this data, individuals can contact Julie Smith, the Technical Monitoring Coordinator for Stream Watch at 302-239-2334 ext. 52 or send email to stream@dnsashland.org. This information may be useful in to “List Candidate Causes” in the Stressor Identification Process.

Delaware Surface Water Quality Management

The Watershed Assessment Section of the Delaware Department of Natural Resources and Environmental Control (DNREC) has developed a website to provide information about Surface Water Quality Management in the State. The State of Delaware’s 1998 305(b) report can be [downloaded from: http://www.dnrec.state.de.us/dnrec2000/Library/Water/305breport.htm](http://www.dnrec.state.de.us/dnrec2000/Library/Water/305breport.htm). The 305(b) report identifies water quality problems and sources for watersheds in Delaware, which is useful in the “List Candidate Causes” step in the Stressor Identification Process.

The most recent copy of the State of Delaware’s Surface Water Quality Standards can be downloaded in a PDF file from the following web page: <http://www.dnrec.state.de.us/dnrec2000/Divisions/Water/WaterQuality/Standards.htm>. These criteria are based on scientific judgments about stressor levels that, if met, will protect designated uses of water bodies in the State. Thus, a user can compare the stressor levels (e.g., chemical concentrations) at his or her site to these criteria in the “Analyze Evidence” step of the Stressor Identification Process.

DRAFT

More information about Surface Water Quality Management in Delaware is accessible at:
<http://www.dnrec.state.de.us/dnrec2000/Divisions/Water/WaterQuality/WQM.htm>

Environmental Monitoring and Assessment Program (EMAP)
<http://www.epa.gov/emap/html/data/surfwatr/data/>

Surface waters data and metadata for Mid-Atlantic streams can be accessed from the EMAP web page listed above. Benthic, water chemistry, sediment, fish and periphyton data collected from thirteen Mid-Atlantic streams between 1993 and 1998 can be downloaded in ASCII format and metadata can be downloaded in PDF format. This biological and chemical data may be useful to “Define the Impairment” & “List Candidate Causes”, respectively, in the Stressor Identification Process at your site.

Maryland Biological Stream Survey (MBSS)
<http://mddnr.chesapeakebay.net/mbss/search.cfm>

The MBSS collects data to look at the physical, chemical and biological status of a stream, the geography (landscape) surrounding the stream and the relationships between the living resources (fish and bugs) and their physical surroundings to try to define the problem of acid deposition in Maryland. The MBSS is a program led by the Monitoring and Nontidal Assessment Division of the Maryland Department of Natural Resources.

In cooperation with U.S. EPA's Mid-Atlantic Integrated Assessment Program the Maryland Department of Natural Resources (DNR) recently completed the first-ever, comprehensive survey of the almost 9,000 miles of small and medium-sized freshwater streams. Sample sites were randomly selected 75 meter segments of first, second, and third order streams (Strahler; 1:250,000 scale USGS maps) in each of 18 major river basins in Maryland, sampled in both spring and summer where possible. Parameters measured included physical habitat characteristics, water chemistry, fish, and benthic macroinvertebrates. Herps, submerged aquatic vegetation, bivalves and crustaceans are also measured when present. Data collected between 1995 and 1997 is available for download as text, tab-delimited files and data sets from 1995-1997 and 2000 may be searched on the MBSS web page by county, stream name, river basin or watershed.

This biological and chemical data may be useful to “Define the Impairment” & “List Candidate Causes”, respectively, in the Stressor Identification Process at your site. Also, the report summarizing the survey findings, "From the Mountains to the Sea - The State of Maryland's Freshwater Streams," may be a useful case study for identifying associations between candidate causes and effects in the “Analyze Evidence” step of the Process for sites in Maryland. This report can be downloaded from EPA 's Mid-Atlantic Integrated Assessment (MAIA) web site at: <http://www.epa.gov/maia/html/reports.html>. Other MBSS publications are available at: http://www.dnr.state.md.us/streams/mbss/mbss_pubs.html.

DRAFT

Maryland Streams – part of the DNR website
http://www.dnr.state.md.us/streams/pubs/pub_list.html

Monitoring and Non-Tidal Assessment (MANTA) does much work in Maryland streams. The Monitoring and Non-Tidal Assessment Division (MANTA) of the Maryland Department of Natural Resources (DNR) is organized into three interactive programs - Atmospheric Deposition (AD), Ecological Assessments (EA) and Monitoring (M). The Division is responsible for assessment of status and trends of biological communities in the non-tidal portions of tributaries in Maryland. A bibliography of reports and technical documents written by MANTA, the Stream Corridor and Management Division of MD DNR and the Maryland Water Monitoring Council can be found at: http://www.dnr.state.md.us/streams/pubs/pub_list.html. Examples of publications that may be useful in the Stressor Identification Process are:

Relationship between effects on fish stocks with hydrographic conditions and pollution loadings;

Relating Nutrient Loading to Production of Selected Fish Populations;

Assessment of Chesapeake Bay Benthic Macroinvertebrate Resource Condition in Relation to Water Quality and Watershed Stressors;

Status and Temporal Trends in Benthic Macroinvertebrate Communities as Indicator of Water Quality at Maryland's Core Monitoring Stations, 1976-1992.

Basin Fact Sheets (describing the current status of wadeable streams)

Some reports are available on-line in PDF format and others are available upon request from the MD Department of Natural Resources (see web page for contact information). The Maryland 2000 305(b) Report is available for download in PDF format from:
http://dnrweb.dnr.state.md.us/download/bays/MD2000_305b.pdf

These or other publications on the Maryland DNR web site may provide useful information in order to “Identify Candidate Causes” and “Analyze Evidence” (e.g., search for associations between causes and effects from other sites).

Maryland Water Monitoring Council
<http://www.mgs.md.gov/mwmc/data.html>

The Maryland Water Monitoring Council is an umbrella organization that includes representatives from professional and volunteer monitoring organizations. “The MWMC is an organization created in 1995 to foster cooperation among groups involved in all types of water monitoring Activities. The Council is interested in physical, chemical, and biological monitoring,

DRAFT

as well as the evaluation of those land use factors that affect changes in aquatic habitat quality and quantity.”

From this website, you can download a database of non-tidal tributary water quality monitoring programs in the Chesapeake Bay Watershed compiled by The Chesapeake Information Management System (CIMS). The database is available for download as a zipped Microsoft Access file. The database provides information about over 330 monitoring programs conducted by over 160 organizations throughout the Chesapeake Bay watershed. The database indicates whether data sets gathered by different monitoring programs are available to the public and, if so, how to obtain them. Thus, this database can potentially be used to locate monitoring data from locations in your project area to “Define the Impairment” or “List Candidate Causes” in the Stressor Identification Process.

Mid-Atlantic Highlands Coordinating Council; Potomac River Basin Pilot Project

<http://ael.er.usgs.gov/groups/gis/potomac/mahaw.htm>

At the request of the Mid-Atlantic Highlands Coordinating Council (MAHCC), a coalition of federal and state natural resource agencies, personnel at USGS compiled a preliminary inventory of biological, chemical and physical data for the upper Potomac River drainage basin (in MD, PA, VA and WV). Detailed information was identified for the Cacapon River, Evitt's Creek, the North Branch of the Potomac River, Opequon Creek and Sideling Hill Creek. Inventory tables accessible on this web page list several data sets that were collected for these waterbodies between 1950 and 1997. The following information is provided for each data set: the data format, the agency that collected the data and a contact person at that agency. Obtaining the types of data listed on this web page could be helpful to “Define the Impairment” or “List Candidate Causes” in the Stressor Identification Process. [Note: Upon contacting USGS, I was told that no further work has been done on the Potomac River Basin Pilot Project and the MAHCC no longer exists.

Mid-Atlantic Integrated Assessment (MAIA) Project

<http://md.usgs.gov/maia/>

<http://www.epa.gov/maia/assets/pdf/MAHAStreams.pdf>

The Mid-Atlantic Integrated Assessment (MAIA) is a research, monitoring, and assessment initiative. From the MAIA website (<http://www.epa.gov/maia/>), there are links to estuaries, surface waters, landscape ecology, forest and agro-ecosystems data. The U.S. Geological Survey (USGS) has recently completed analyses of the occurrence of nitrate and pesticides in waters of the Mid-Atlantic Region as part of the Mid-Atlantic Integrated Assessment (MAIA) project of the U.S. Environmental Protection Agency (USEPA). The MAIA project encompasses USEPA Region III and adjacent parts of major river basins in New Jersey, New York, and North Carolina. Information about this USGS project for MAIA can be found at:

<http://de.usgs.gov/publications/ofr-98-158/>.

DRAFT

Data collected from 463 sites between October 1973 and March 1997 were used to characterize the occurrence of pesticides in surface water of the Mid-Atlantic Region in the report: Pesticides in Surface Water of the Mid-Atlantic Region, by M.J. Ferrari, S.W. Ator, J.D. Blomquist, and J.E. Dysart (1998). Compiled nitrate and pesticide data for the Mid-Atlantic Region and supporting information are contained in four tab-delimited ASCII files and may be obtained along with the report listed above from <http://md.usgs.gov//maia/>. This data could be helpful to “List Candidate Causes” in the Stressor Identification Process.

The report Mid-Atlantic Highlands Streams Assessment can be downloaded from the MAIA website at <http://www.epa.gov/maia/assets/pdf/MAHASStreams.pdf>. One of the purposes of the report is to identify and rank the relative importance of stressors affecting the conditions of streams in the Mid-Atlantic Highlands region. Therefore, it could be helpful to “List Candidate Causes” in the Stressor Identification Process as well.

National Environmental Monitoring Initiative Mid-Atlantic Inventory
<http://www.epa.gov/cludygxb/site-mid.html>

“This inventory presents 128 federal research and monitoring programs that operate in the mid-Atlantic region. Programs are organized by sampling design into three tiers recognized by the National Environmental Monitoring Initiative: spatially continuous monitoring over a large region 910,000 km² or more), spatially sub-sampled surveys and monitoring designed to evaluate the status of a large region by sampling a subset of the total area, intensive research sites that are selected due to their known ecological condition or suitability for experimental manipulation.” Metadata (e.g., name of monitoring program, contact information, data available) for the Mid-Atlantic region can be accessed using user-defined queries. Data sets identified from this source may be useful to “Define the Impairment” or “List Candidate Causes” in the Stressor Identification Process.

Pennsylvania Department of Environmental Protection - Water Quality Assessment and Standards

<http://www.dep.state.pa.us/dep/deputate/watermgt/wqp/wqstandards/wqstandards.htm>
<http://www.dep.state.pa.us/dep/subject/eqb/2000/June20/fch93Annex.pdf>

The State’s 2000 Section 305(b) - Water Quality Assessment Report and 2001 Section 305(b) Report – Update as well as older copies of the 305(b) reports are accessible through the water quality assessment and standards web page (<http://www.dep.state.pa.us/dep/deputate/watermgt/wqp/wqstandards/wqstandards.htm>). These reports discuss causes of observed impairment in the state’s waters and thus may be useful in the “List Candidate Causes” step in the Stressor Identification Process.

The latest water quality amendments to Chapter 93 of DEP's Regulations (Title 25, PA Code) include modifications to Pennsylvania water quality standards. The amendments can be

DRAFT

downloaded at:

<http://www.dep.state.pa.us/dep/subject/eqb/2000/June20/fch93Annex.pdf>

These standards can be used in the “Analyze Evidence” step of the Stressor Identification Process for sites in Pennsylvania.

Pennsylvania Senior Environmental Corps (PaSEC)

<http://www.environmentaleducation.org/default.lasso>

The PaSEC program began in 1997 and is the result of a partnership between Pennsylvania’s DEP, Department of Aging and the Environmental Alliance for Senior Involvement (EASI). The EASI Water Monitoring Database has two major components:

- One component is for the members of the Pennsylvania Senior Environmental Corps (PaSEC) and any other group using the standardized protocols. This component has fields for seven physical and chemical indicators of water quality including pH, conductivity, water temperature, dissolved oxygen, nitrates, total phosphates and sulfates. A habitat assessment and a water quality rating based on benthic macroinvertebrate sampling are also part of this component.
- The second component is an "open" component available for use by any volunteer monitors who are using their own protocols and quality control plans. This part of the database has fields for the seven physical and chemical indicators listed above as well as for hardness, turbidity, chlorides, chlorophyll a, total dissolved solids, total suspended solids, aluminum, manganese, iron, fecal coliform and Secchi depth.

Both of these databases are searchable on-line and users can view the results of their data queries in table format in their web browser. This monitoring data is another source of information that could be helpful to “List Candidate Causes” in the Stressor Identification Process for sites in Pennsylvania.

Pennsylvania Spatial Data Access (PASDA)

<http://www.pasda.psu.edu/access/download.shtml>

“PASDA is Pennsylvania’s official geospatial information clearinghouse and the Commonwealth’s node on the National Spatial Infrastructure (NDSI). The PASDA clearinghouse provides for the widespread sharing of geospatial data, eliminates the creation of redundant data sets, and serves as a resource for locating data throughout the Commonwealth through its data storage, interactive mapping/webgis applications, and metadata/documentation efforts. The data available on PASDA is the result of efforts by state, local, regional, and federal governments, non-profit organizations, and academic institutions who wish to make their data or metadata available through the clearinghouse.”

Data can be downloaded for free from the PASDA Data download web page

(<http://www.pasda.psu.edu/access/download.shtml>) or all data on PASDA can be obtained on CD-ROM through a cooperative arrangement with the Pennsylvania State Data Center for a

DRAFT

nominal charge. The following types of data can be downloaded from the PASDA website for the municipal, county, watershed or state level:

- Water quality sampling records for 622 sites throughout Pennsylvania (Alliance for Aquatic Resource Monitoring (ALLARM) data).
- Drainages, streams, road, and population coverages from the Chesapeake Bay drainage are (compiled by Chesapeake Bay Program)
- Maps with locations of natural heritage sites classified according to biological diversity and ecological integrity (not georeferenced).
- Physiographic provinces and glacial extents.
- Aerial photographs
- Various coverages from the EPA concerning Acid Mine Drainage, forest cover types, sub-ecoregions, toxic resource inventory, CERCLIS, and reach files.
- Historical fish species database from the Pennsylvania Fish and Boat Commission
- State-wide and county-wide floodplain coverages.
- Urban, county and state-wide watersheds.
- Map of landform subdivisions of Pennsylvania (preliminary).
- Natural Resources Conservation Service soil survey maps
- Census block data, ecoregion boundaries, Pennsylvania hydrology, and zip codes.
- State-wide boundary and county-wide roads coverages from the Pennsylvania Dept. of Transportation.
- Multiple use coverages from the Pennsylvania Department of Environmental Protection
- Satellite imagery
- County-wide streams coverages from Penn State's Environmental Resources Research Institute (Streams Networks).
- State- and county-wide surface geology coverages.
- Digital elevation models for Pennsylvania available from USGS National Elevation Database (NED).

One of the data sets that can be downloaded from the PASDA web site is the ALLARM database. ALLARM is a project of the Environmental Studies Department at Dickinson College, who partners with Pennsylvania communities and individuals who are working to protect and restore watersheds. ALLARM's partner watershed groups include:

- Conodoguinet Creek Watershed Association (CCWA), Cumberland County
- Codorus Creek Monitoring Network, York County
- Pine Creek Headwaters Protection Group, Tioga County
- Kiski-Conemaugh Alliance, Armstrong, Indiana, Westmoreland, Cambria, and Somerset Counties [Three groups: Roaring Run Watershed Association, Conemaugh Valley Conservancy, Stony Creek - Conemaugh Rivers Improvement Project]
- Coplay Creek Monitoring Project - Lehigh Valley Chapter of the Sierra Club, Lehigh County
- Ridge & Valley Streamkeepers, Bedford County

DRAFT

- Shermans Creek Conservation Association, Perry County
- Schuylkill Office, Delaware Riverkeeper Network
- Powells and Armstrong Creek Conservation Association, Dauphin County
- Fishing Creek Watershed Association, Columbia County

ALLARM's database contains more than 33,466 records on water quality from 1986 to the present from 622 testing sites throughout Pennsylvania. Information in records includes at least alkalinity and pH and includes nitrates and phosphates for some sites since 1996. A ZIP file with ALLARM's full database in Microsoft® Access© format with records linked through SQL to ArcView© shapefiles can currently be downloaded from the PASDA data download web page. The PASDA database is updated every six months. Data summaries are also available from the ALLARM website (<http://www.dickinson.edu/storg/allarm/mainpages/database.html>).

Also, on PASDA's "PA Explorer" web page (<http://www.pasda.psu.edu/explore/index.shtml>), there are several links to web sites that allow users to create their own maps of PASDA data over the Internet. This section is ideal for users who are unfamiliar with GIS or lack GIS software to map their data.

The data sources accessible from the PASDA website can help "Define the Impairment" or "List Candidate Causes" in the Stressor Identification Process for sites in Pennsylvania.

Susquehanna River Basin Commission (SRBC) Nutrient Assessment Program

<http://www.srbc.net/nutrientdata.htm>

<http://www.srbc.net/trends.htm>

<http://www.srbc.net/nutrientloads.htm>

"The Susquehanna River Basin Commission implemented a five-year nutrient-monitoring program in October 1984 to establish a database for estimating nutrient and suspended sediment loads in the Susquehanna River Basin. The initial five-year program was concluded at the end of December 1989, and five of the twelve original sites were selected for continued long-term monitoring." Thus, the data set available at this web site includes nutrient and sediment data collected between 1984 and 2000 at various waterbodies in the Susquehanna River Basin. The raw data as well as data summaries (trends and loads) can be viewed on the web or downloaded.

The nutrient and sediment data web page (

<http://www.srbc.net/nutrientdata.htm>) summarizes the sample collection methods, parameters analyzed and types of data that can be downloaded. The current data files are formatted as comma delimited text files. Nutrient and sediment trends and loads statistics and maps can be downloaded in PDF or Excel format from the following web pages:

<http://www.srbc.net/trends.htm> and <http://www.srbc.net/nutrientloads.htm>. This monitoring data could be helpful to "List Candidate Causes" in the Stressor Identification Process for sites in the Susquehanna River Basin.

DRAFT

USGS National Water Quality Assessment (NAWQA) Data Warehouse

http://orxddwimdn.er.usgs.gov/servlet/page?_pageid=543&_dad=portal30&_schema=PORTAL30

The U.S. Geological Survey (USGS) began its NAWQA (National Water Quality Assessment) program in 1991, systematically collecting chemical, biological, and physical water quality data from study units (basins) across the nation. This data warehouse contains data from the first 36 study units. Data from the 15 study units started in 1997 will be incorporated later. The data warehouse currently contains and links the following data:

- Chemical concentrations in water, bed sediment, and aquatic organism tissues for about 500 chemical constituents;
- Site, basin, well and network characteristics with many descriptive variables;
- Daily stream flow information for fixed sampling sites;
- Ground water levels for sampled wells;
- 2,800 stream sites and 5,000 wells;
- 26,000 nutrient samples and 15,000 pesticide samples as well as 5,000 VOC samples;
- 1200 samples of bed sediment and aquatic organism tissues.

There are 6 study units in the mid-Atlantic region: Delaware River Basin, Lower Susquehanna River Basin, Delmarva Peninsula, Potomac River Basin, Allegheny and Monongahela Basins and Kanawha-New River Basin. Data collected between 1991 through 1997 is available for download from the data warehouse for a few of these study units. More information about each study unit (e.g., sampling efforts, summary reports) can be accessed at the web pages listed below.

Allegheny-Monogahela River Basin NAWQA data
http://pa.water.usgs.gov/almn/almn_main_data.html

Delaware River Basin

<http://nj.usgs.gov/delr/sampling.html>

Kanawha-New River Basin

<http://www-wv.er.usgs.gov/nawqa/>

This study unit was one of the second set of study units which began investigations in 1994. The period of the study has ended and the data are no longer available online. For information on the Kanawha-New River Basin study unit or to request available reports contact:

District Chief <dc_wv@usgs.gov>
U.S. Geological Survey Water Resources
11 Dunbar Street Charleston, WV 25301
(304) 347-5130

DRAFT

Lower Susquehanna River Basin

<http://www.pah20.er.usgs.gov/projects/lsus/>

Further information on the survey and on the specific data for each sampling site can be obtained from Kevin J. Breen or Michael D. Bilger (New Cumberland, PA) at (717) 730-6900.

Potomac NAWQA Datasets

<http://md.water.usgs.gov/pnawqa/datarpt/>

Water quality and spatial data collected as part of the Water-Quality Assessment of the Potomac River Basin from 1992 to 1996 is available for download from the web page listed above. The data set includes chemical, physical, streamflow, bed-sediment contaminants, aquatic-tissue contaminants, fish community, and selected stream habitat data. These data were collected at a variety of water bodies in the Potomac River Basin.

The data from the data warehouse can be viewed on-line, printed, or exported in several formats (e.g., Excel, Lotus 1-2-3, HTM, tab delimited, comma delimited). Data on the web pages specific to a study unit can be viewed on-line or downloaded in HTM, PDF or TXT format from the web page (there is no export function). This monitoring data may be useful to “Define the Impairment” & “List Candidate Causes” as part of the Stressor Identification Process for sites located within the study units mentioned above.

Virginia Department of Environmental Quality –Water Quality Program

<https://www.deq.state.va.us/webapp/wqm.homepage>

Virginia DEQ monitors water quality in rivers, lakes and estuaries at some 1,100 locations to detect pollutants and to assess pollution prevention efforts. DEQ staff in each of the regional offices collects water samples on a routine schedule at more than 1,000 locations across the Commonwealth. The tens of thousands of samples and chemical test results generated each year are kept in a computer database that is updated daily. In addition to recent data, the database contains historical STORET data collected by the VA DEQ as well as other historical data going back as far back as far as 1941. This water quality data can be queried by County/City or Stream at the Water Quality Monitoring Homepage (

<https://www.deq.state.va.us/webapp/wqm.homepage>). Sample location information (e.g., stream name, type of waterbody, watershed, latitude & longitude), sample date and time, and chemical results are given as a result of the query. Biological data (e.g., macroinvertebrate data) is not included in this database. Rather, this type of data is compiled in databases at 7 regional DEQ offices. DEQ also supports a cooperative arrangement with hundreds of citizen volunteers who collect water quality data throughout the Commonwealth, but the data collected by these volunteers are not compiled in one central database. This monitoring data may be useful to “List Candidate Causes” as part of the Stressor Identification Process for sites located in Virginia.

Virginia’s 2000 305(b) Water Quality Assessment Report can be downloaded from

<http://www.deq.state.va.us/water/305b.html>. This report may also be helpful to “List Candidate

DRAFT

Causes” at sites in Virginia. Also, Virginia’s water quality criteria and standards can be downloaded in PDF format from <http://www.deq.state.va.us/wqs/#vawqs>. These criteria can be used to “Analyze Evidence” at sites in Virginia.

West Virginia Department of Environmental Protection, Division of Water Resources,
Watershed Assessment and Strategic Planning Section
http://www.dep.state.wv.us/wr/index.cfm?page=OWR_Website/C_and_D/WAP.htm

“The Watershed Assessment Program collects and compiles water quality data on waters statewide. The Strategic Planning group uses the water quality data to compile the state's list of impaired streams and write and implement water quality improvement plans known as total maximum daily loads, or TMDLs. “ Activities undertaken by these programs that may provide useful information for the “Define the Impairment” and “List Candidate Causes” steps of the Stressor Identification Process include:

Watershed Assessments

The results of chemical, habitat, and biological monitoring at many of the states 9000+ streams are summarized in Watershed Assessment Reports that can be downloaded from this website.

Citizens Volunteer Monitoring

WV Save Our Streams is a program empowering citizens to monitor the water quality of their local streams via chemical and/or biologic methods. Volunteers collect information about: water quality (pH, DO, temperature, conductivity), water conditions (clarity, color, odor), macroinvertebrate counts, characteristics of the stream bed, characteristics of surrounding riparian zone and surrounding land uses. A few groups also collect water and sediment samples for chemical analysis. There are links to the survey forms used by the volunteers at the WV Save Our Streams website (http://www.dep.state.wv.us/wr/index.cfm?page=OWR_Website/wvsos/wvsos.htm). Most of the stream survey data (collected since the early 1990’s) are in paper format. However, in the past few years, an access style database has been set-up to log data collected by volunteers. Each year a report is produced from this database. These annual reports are available on the WV DEP website. To obtain data for a particular waterbody or watershed, you can contact:

Timothy Craddock, Citizens Monitoring Coordinator
WVDEP, Office of Water Resources
1201 Greenbrier Street
Charleston, WV 25311
Phone: 304-558-2108
Fax: 304-558-2780
tcraddock@mail.dep.state.wv.us

Large Stream Ambient Monitoring Network

Quarterly monitoring for chemical constituents at 26 locations on West Virginia’s larger rivers and streams. A map of and information about ambient stations is available from this website.

DRAFT

Special Biological Studies

Analysis of point source effluents on biological stream quality at selected sites statewide.

305(b) Reporting

Assessment of the condition of the state's waters, released every two years in a report. The 1998 305(b) Report and 2000 305(b) Report are available on-line. The 2002 305(b) report is expected to be released in April of 2002. The following information is provided in the 305(b) report for each watershed in the surface water assessment section: Stream sampling locations, overall designated use report, use support matrix summary, relative assessment of causes, relative assessment of sources, and 303(d) listed streams. There are also sections on Lake Water Quality Assessment, Groundwater Quality and Wetlands.

303(d) Listing

Listing of streams which are not meeting water quality standards, often referred to as the impaired streams list. Work is underway to compile the 2002 303(d) list and is expected to be finalized in April of 2002. The 1996 303d List, 1998 303d List and 2002 303d List Call for Data are available on-line.

The Environmental Quality Board of the WV DEP is responsible for issuing rules that set the water quality standards for West Virginia's surface and ground waters. West Virginia's surface water quality standards can be accessed by following the links at:

<http://www.dep.state.wv.us/eqboard/index.cfm?page=waterquality.htm>

APPENDIX C

CASE STUDY MATERIALS AND REFERENCES

CADDIS Workshop
Mid Atlantic Case Study
Identify Impairment and List Candidate Causes

The Premise of the Case Study: Stressor identification sufficient to support effective water quality management can be performed at the scale of small watersheds in addition to reach/site specific scales.

Biological impairment is often the result of multiple processes occurring over the drainage area. There may be multiple “pollutants” (i.e., proximal stressors) affecting the biota. However, these stressors in many cases may derive from a single source or human activity. As an example, biota in the receiving waters of a small watershed may be impaired due to the multiple insults of excess nutrients and suspended solids, increased temperature regimes, and increased sedimentation. However, all these may be the result of conversion of riparian forest in the watershed to agriculture. The common source of these stressors is therefore the agricultural fields along the streams.

The current framework for water quality standards focuses on identifying the causative stressors/pollutants, identifying the contributing sources, then allocating maximum loading for the stressors among those sources. However, the larger-scale activity (e.g., the areal pattern of agriculture in the drainage basin relative to its distance to the receiving waters) may be the reason for a particular pattern of multiple site-specific instances of biological impairment. It may therefore be more efficient and cost-effective to recognize and deal with the principle cause of the impairments (the pattern of sources) as opposed to the site-specific, stressor-specific cause(s) of any particular biological impairment.

At this scale of attention, the focus on causality is one step back in the pathway. The presence, co-occurrence, etc. of a particular suite of stressors is supporting evidence for a particular source or land use pattern. Conversely, if evidence for a particular source or land use pattern is easily identified, and the cumulative effects of such a pattern are already scientifically supported, resources need not be devoted to identifying specific stressor-response relationships for each individual stressor at each specific site, but can instead be devoted to determining the most effective management action to take to reduce the overall biological impairment.

The Challenge: The Office of Water is increasing emphasis on the states’ Continuing Planning Process, supporting increased use of watershed management practices. To deal with the huge volume of TMDLs, OW is also strongly recommending “batching” or grouping TMDLs in the same watershed, with the recognition that many of the TMDLs are interrelated. Solutions for those impairments will require effective holistic management action to get at the root source or land use pattern. States need methods and case studies to implement OW requirements.

Obviously, there are significant gaps in our knowledge. As one example, quantification of the correlations between particular land use patterns and the resultant environmental effects (particularly biology) is in its infancy. However, so is quantification of stressor-response relationships, particularly in terms of understanding multiple, cumulative and synergistic stressor-response. The Stressor Identification Guidance articulates a flexible structure to evaluate evidence, using well-known principles for strength of evidence analysis. It recognizes gaps in both current knowledge and in data availability, and provides an approach for drawing

conclusions anyway.

Using two kinds of case studies at the CADDIS workshop (one focused on a reach-specific analysis, one focused on a small watershed approach) will help us to design a more useful decision support information system, and, as added benefits, may help refine the method for this type of case as well as provide an example for states struggling to implement the new OW directions. The Mid Atlantic case study walk through the structured SI approach for the Cabin John watershed with the intent of conceptualizing an information system that will help investigators find, access, organize and share information useful for evaluation of the cause(s) of impairment at a small watershed scale.

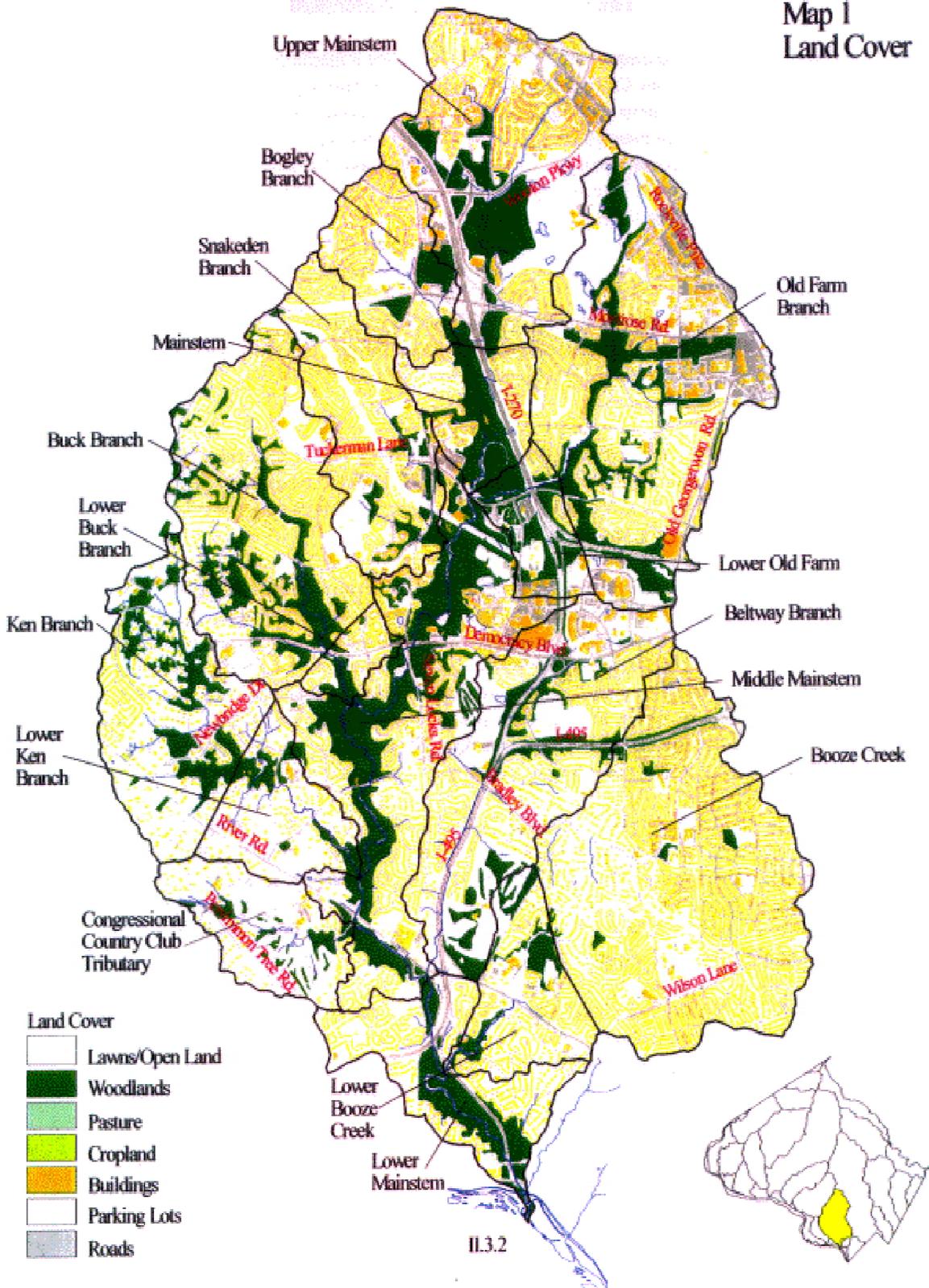
The Situation: (Van Ness and Haddway, 1999)

The Cabin John Creek watershed is within the Potomac River drainage in southern Montgomery County (Map 1). The watershed, located in the Piedmont ecoregion, occupies approximately 16,022 acres. Cabin John Creek originates within the city of Rockville and flows in a southerly direction to a confluence with the Potomac River between the Little Falls dam and Great Falls. The Watershed is bounded by Rockville Pike (Rt. 355) and Old Georgetown Pike (Rt. 187) to the east and Falls Rd.(Rt. 189) to the west.

The watershed has been significantly impacted by suburban development patterns centered around the County's main transportation corridors. The Interstate 495/270 corridor passes through the central part of the watershed, and commercial and high density residential development are common along this corridor, particularly affecting the eastern tributaries. Rockville Pike and the City of Rockville occupy the headwaters of Cabin John. In contrast, the western tributaries transition to lower density residential communities with far less commercial development. On-site stormwater runoff controls are uncommon in Cabin John. Like many of the watersheds in the lower part of the county, this area developed before environmental regulations for stream buffers and stormwater management went into effect.

Cabin John Creek Watershed

Map 1
Land Cover



Define the Scope of the Investigation:

Montgomery County uses biomonitoring to assess the cumulative environmental impacts of economic development on the health of its local streams and ecosystems. This data has allowed them to develop a countywide stream protection strategy to use in their planning process. The local-level biomonitoring data feed into the state’s NPDES point source permitting, and the state’s CWA 305(b) and 303(d) reporting processes, as well as identifying county priorities for stream restoration. The county adopted the state’s MBSS protocols in 2001, thus ensuring full compatibility with state-level assessment.

The county collects information on fish, benthic macroinvertebrates, and habitat, and has developed IBIs based on the MBSS Eastern Piedmont indices. Site-specific IBI results are used to estimate the extent of streams in excellent, good, fair and poor condition with respect to the biotic integrity of the fish and benthic communities. Table C-1 contains detailed descriptions for each of the IBI categories. The highest scores were designated as good recognizing that reference sites may not represent the highest attainable condition

Table C-1: Example of Biological Integrity Classification Method Using Total Index of Biological Integrity Scores, Narrative Biological Integrity Class Descriptions, and Characteristics of Each Class

IBI Score	Narrative Integrity Class	Characteristics
46 to 50	Excellent	Comparable to the biological community found in reference streams. Exceptional assemblage of species with a balanced community composition
34 to 45	Good	Decreased number of sensitive species, decreased number of specialized feeding groups, some intolerant species present
22 to 33	Fair	Intolerant and sensitive species are largely absent; unbalanced feeding group structure
10 to 21	Poor	Top carnivores and many expected species absent or rare; general feeders and tolerant species dominant.

Describe the Impairment:

Analysis of the 1996 Stream Monitoring and Habitat Data

Ten stations were monitored in Cabin John Creek in 1996 (Map 3). The data is shown in Table C-2. The overall resource condition of Cabin John Creek was determined by assessing the cumulative impacts that occurred in the watershed as indicated by the use of an interim Index of Biological Integrity (IBI) for freshwater fish and benthic macroinvertebrates (Map 3). A yearly assessment of “excellent”, “good”, “fair”, or “poor” resource condition was made by examining the trends expressed by the two IBI’s. This is not the same as averaging the two scores. Seasonal trends were examined and a yearly stream condition was established for the subwatersheds. Resource conditions were evaluated for 13 subwatersheds. Buck Branch and Ken Branch received an overall resource condition of good. Snakedon Branch, Thomas Branch, and Booze Creek received an overall resource condition rating of poor. The entire mainstem received an overall resource condition assessment of fair (Map 3).

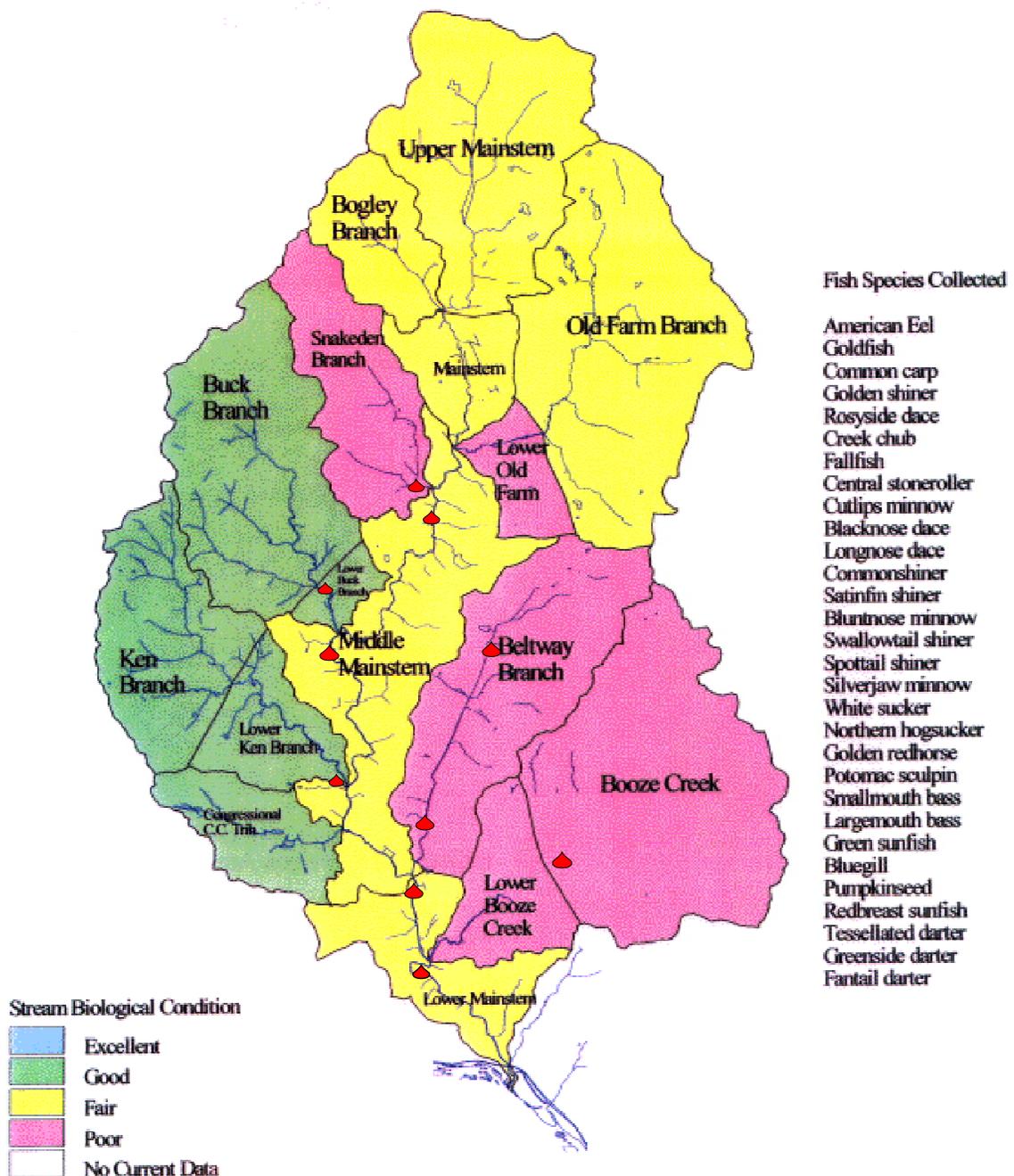
**Table C-2
Monitoring Data, Cabin John Creek Watershed, 1996**

STATION NUMBER	IBI (benthic)	IBI (fish)
CJBB201 (w.br.Buck Br.)	22 - FAIR	4.3 - GOOD
CJBC202 (Booze Cr.)	18- FAIR	1.2 - POOR
CJCJ302 (mainstem)	16 - POOR	2.8 -FAIR
CJCJ303 (mainstem)	14 - POOR	3.2 -FAIR
CJCJ305 (mainstem)	24 - FAIR	3.7 - GOOD
CJCJ307 (mainstem)	14 - POOR	3 - FAIR
CJKB201 (w.br. Ken Br.)	24 - FAIR	3.7 - GOOD
CJSB101 (Snakedon Br.)	24 - FAIR	1- POOR
CJTB101a (Thomas Br.)	14 - POOR	1.7 - POOR
CJTB101b (Thomas Br. trib)	14 - POOR	1.2 - POOR

Cabin John Creek Stream Condition

Based on biological indicators.
See Chapter 2 for details.

Map 3



II.3.4

REFERENCES

Van Ness, K; Hadway, M. (1999) Cabin John Creek Watershed Study. Montgomery County Department of Environmental Protection, Watershed Management Division. Rockville, MD. Available for download at <http://www.co.mo.md.us/services/dep/SPA/pdf%20files/cabinReport99.pdf> (accessed August 20, 2002).

LITTLE SCIOTO CASE STUDY

REFERENCES

- Cormier, SM; Norton, SB; Suter, GW; et al. (2002) Determining the Causes of Impairments in the Little Scioto River, Ohio, USA: Part 2. Characterization of Causes. *Environmental Toxicology and Chemistry*, 21(6): 1125-1137.
- Norton, SB; Cormier, SM; Suter, GW; et al. (2002) Determining Probable Causes of Ecological Impairment in the Little Scioto River, Ohio, USA: Part 1. Listing Candidate Causes and Analyzing Evidence. *Environmental Toxicology and Chemistry*, 21(6): 1112-1124.
- Suter, GW; Norton, SB; Cormier, SM. (2002) A Methodology for Inferring the Causes of Observed Impairments in Aquatic Ecosystems. *Environmental Toxicology and Chemistry*, 21(6): 1101-1111.



United States
Environmental Protection Agency/ORD
National Center for
Environmental Assessment
Washington, DC 20460

Official Business
Penalty for Private Use
\$300

EPA/600/R-02/078
October 2002

Please make all necessary changes on the below label,
detach or copy, and return to the address in the upper left-
hand corner.

If you do not wish to receive these reports CHECK HERE ;
detach, or copy this cover, and return to the address in the
upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35
