Development Plan for the Causal Analysis/Diagnosis Decision Information System (CADDIS)

Development Plan for the Causal Analysis/Diagnosis Decision Information System (CADDIS)

National Center for Environmental Assessment
National Exposure Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC 20460

DISCLAIMER

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

ABSTRACT

Increasingly, the regulatory, remedial, and restoration actions taken to manage impaired environments are based on measurement and analysis of the biotic community. When an aquatic assemblage has been identified as impaired, an accurate and defensible assessment of the cause can help ensure that appropriate actions are taken. The U.S. EPA's Stressor Identification Guidance describes a methodology for identifying the most likely causes of observed impairments in aquatic systems. Stressor identification requires extensive knowledge of the mechanisms, symptoms, and stressor-response relationships for various specific stressors as well as the ability to use that knowledge in a formal method for causal analysis. This document describes a strategy for developing the Causal Analysis/Diagnosis Decision Information System (CADDIS). CADDIS is envisioned as a decision support system that will help investigators in EPA Regions, states, and tribes find, access, organize, and share information useful for causal evaluations in aquatic systems. It will include supporting case studies and analysis tools, and it will provide access to databases that contain information useful for causal evaluations. The system will be developed incrementally and iteratively, and frequent user input and feedback will be essential to the system's success.

CONTENTS

CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

PREFACE

This document presents a plan for developing the Causal Analysis/Diagnosis Decision Information System (CADDIS). CADDIS will be a decision support system to help scientists and decisionmakers determine the causes of biological impairments so that appropriate remedial, regulatory, or restoration actions can be taken.

This plan will serve as a guide for the system development team and is based, in part, on the results of the Workshop on the Causal Analysis/Diagnosis Decision Information System held in August 2002. The workshop brought together individuals from the U.S. Environmental Protection Agency's (EPA's) Office of Research and Development (ORD), Office of Water (OW), and Regions as well as representatives from the states for the purpose of conceptualizing CADDIS and identifying critical research needs for system implementation and population. From the workshop, a summary report was written (U.S. EPA, 2002a) that identifies key system functionality and research needs and makes suggestions for the system's design, platform, and architecture. The workshop report is being used as the basis of the CADDIS development plan.

The draft of this plan was put together by representatives from ORD's National Center for Environmental Assessment (NCEA), National Exposure Research Laboratory (NERL), and National Risk Management Research Laboratory (NRMRL). The draft was then reviewed by the CADDIS workgroup, which has members from each major ORD laboratory and OW. Comments from the workgroup and NCEA managers were incorporated into the draft, which was then released for peer review. Reviewers were chosen to reflect a mix of the user community (i.e., OW, Regions, and states), experienced system developers, research planners, managers, and the ORD research community. After addressing the review comments, the draft plan was finalized and made available to the public.

We hope this development strategy will cultivate a shared vision for CADDIS among managers, research coordination representatives, potential users, and the research community.

REVIEWERS

Midwest Biodiversity Institute and Center for Applied Bioassessment and Criteria
Edward Rankin
Chris Yoder

Idaho Department of Environmental Quality
303(d) / 305(b) Program
Michael Edmondson

EPA, Office of Research and Development

National Center for Environmental Assessment
Scott Freeman
Jeff Frithsen
Vic Serveiss
Chieh Wu

National Center for Environmental Research
Jonathan Smith

National Exposure Research Laboratory
Brenda Rashleigh

National Health and Environmental Effects Research Laboratory
Robert Spehar

EPA, Region 5
Water Quality Standards
Candice Bauer

LIST OF ACRONYMS

APG	Annual Performance Goal
APM	Annual Performance Measure
BASS	Bioaccumulation and Aquatic System Simulation
CADDIS	Causal Analysis/Diagnosis Decision Information System
CWA	Clean Water Act
DEP	Department of Environmental Protection
DO	Dissolved Oxygen
ECOTOX	Ecological Toxicity Database
EDAS	Ecological Data Application System
EERD	Ecological Exposure Research Division
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
ESD	Environmental Sciences Division
GAO	General Accounting Office
IT	Information Technology
LTG	Long-term Goal
MYP	Multi-year Plan
NRRSS	National Riverine Restoration Science Synthesis
NCEA	National Center for Environmental Assessment
NERL	National Exposure Research Laboratory
NRMRL	National Risk Management Research Laboratory
NRC	National Research Council
ORD	Office of Research and Development
OW	Office of Water
PC	Personal Computer
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
SAS	Statistical Analysis Software
SI	Stressor Identification
SPRC	Science Planning and Research Coordination
SSD	Species Sensitivity Distribution
STORET	STOrage and RETrieval database

LIST OF ACRONYMS (continued)

TIE   Toxicity Identification Evaluation
TMDL   Total Maximum Daily Load
UWRRC   Urban Water Resources Research Council

1. EXECUTIVE SUMMARY

This document presents a plan for developing the Causal Analysis/Diagnosis Decision Information System (CADDIS). The primary goal of the CADDIS project is to support investigators in EPA Regions, states, and tribes in determining the most likely causes of aquatic impairments. CADDIS will make remediation and restoration strategies more effective by ensuring that management actions are directed at the true causes of aquatic impairments. CADDIS will be a decision support system based on the *Stressor Identification Guidance Document* (U.S. EPA, 2000); however, it will make that causal assessment methodology easier to use and will provide access to the required information. The stressor identification (SI) methodology was developed primarily to support the development of total maximum daily loads (TMDLs) under the Clean Water Act (CWA). However, both the SI methodology and CADDIS will be useful in effluent permitting, watershed assessments, contaminated site assessments, and other situations in which a biological impairment is observed.

The CADDIS development plan will serve as a guide for the system development team by describing the process for building CADDIS and the key functions that will make up the system. It also describes the development of the knowledge base that is needed to conduct defensible causal evaluations, including databases that make relevant research more accessible and case studies that illustrate the application of the SI process. The development strategy laid out in the plan will also promote a shared vision for CADDIS among managers, research coordination representatives, potential users, and the research community.

The CADDIS development process consists of three components that will be developed in roughly parallel time frames (Figure 1). The first component is the development of the CADDIS platform, which provides the user interface; the logical structure for analyzing, organizing, and synthesizing causal evidence; and the mechanisms for accessing relevant information. The second component is the development of the knowledge base that, when combined with site-specific information, provides the substantive basis for causal inference. The third component is the development of case studies that provide concrete illustrations of the process and a mechanism for sharing experiences and increasing investigator expertise. As a whole, the system is intended to evolve from a tool that provides streamlined access to information in a static environment to an adaptive framework applied by the user to support and document the decision-making process for individual sites. The plan is based on the assumption that sufficient resources will be available and maintained.

	2003	2004	2005	2006	2007	2008	2009
CADDIS Platform	Development Plan		CADDIS 1		CADDIS 2		CADDIS 3
Data base and info. needs		Sediments	Toxics	Conceptual Models			
Case Studies	Willimantic			Long Creek		Compilation	
Training & Guidance			Watershed Academy Module				Guidance
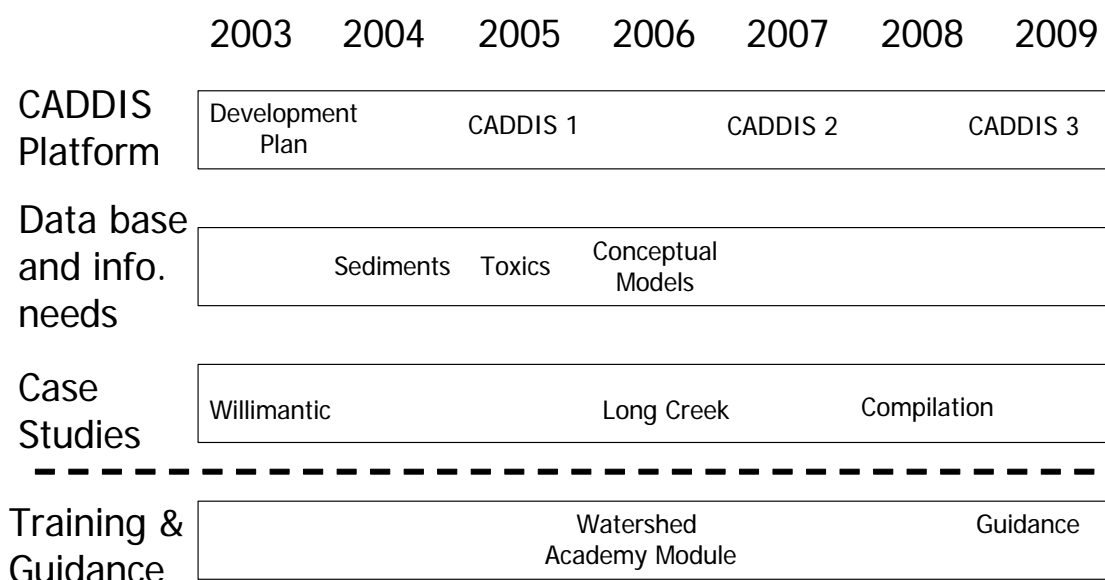
Figure 1. An overview of the CADDIS development timeline. The products shown will be developed directly by the CADDIS development team. Databases and case studies developed by others also will be incorporated into CADDIS, as discussed in Chapters 4 and 5 of this plan. Concurrently with the three components, training will be developed and the Stressor Identification Guidance revised.

An essential characteristic of our proposed approach to development is the solicitation and incorporation of user input throughout the process. To this end, the development team will include members of the user community as well as scientists, programmers, and project managers. In addition, we have developed plans for training, outreach, and publication of additional guidance.

2. INTRODUCTION

This document presents a plan for developing CADDIS. The primary goal of the CADDIS project is to help investigators in EPA Regions, states, and tribes determine the most likely causes of aquatic impairments so that appropriate remedial, regulatory, restoration, or protective actions can be taken. CADDIS will be a computer-based system based on the *Stressor Identification Guidance Document* (U.S. EPA, 2000); however, it will make that causal assessment methodology easier to use and will provide access to the required information. As discussed below, the SI methodology was developed primarily to support the development of TMDLs under the CWA. However, both the SI methodology and CADDIS will be useful in

effluent permitting, watershed assessments, contaminated site assessments, and other situations in which a biological impairment is observed and corrective action is being considered.

This development plan presents the process for building CADDIS and describes the key functions that will make up the system. It also describes the development of the knowledge base that is needed to conduct defensible causal evaluations, including databases that make relevant research more accessible and case studies that illustrate the application of the SI process. The plan will serve as a guide for the system development team, its managers, and research coordination representatives, and will provide the user and research communities with information on the type of decision support system currently envisioned by the U.S. Environmental Protection Agency (EPA, or the Agency). The plan is based on the assumption that sufficient resources will be available and maintained.

2.1. BACKGROUND

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. Use of biological assessments and criteria as a tool to identify impaired water bodies in the United States began in the early 1970s; currently, biological assessment programs are in place in all the states and at least two tribal nations (U.S. EPA, 2002b). When an aquatic assemblage has been identified as impaired, an accurate and defensible assessment of the cause can help ensure that appropriate actions are taken (text box).

The lack of scientific tools with which to properly diagnose causes of biological impairment has been identified by States as a major impediment to the further use of biological assessments and criteria in their water quality programs (GAO, 2000). Methods for identifying cause are crucial for defensible development of TMDLs, which

> **Stressor Identification and CADDIS Build on a Foundation of Sound Biological Assessment Programs**
>
> Stressor identification relies on data gathered from sound biological assessment programs in the states and tribes. These data are not used only to identify the biological impairments that trigger the process; they can also provide key insights into the specific changes that are occurring. The use of biological assessments and biocriteria in state and tribal water quality standards programs is a top priority for EPA (U.S. EPA, 2003a). As such, one of the Agency's objectives is to ensure that all states and tribes develop water quality standards and supporting programs that
>
> - produce accurate, comparable, comprehensive, and cost-effective monitoring data and assessments that are capable of meeting the goal of supporting all water quality management programs (e.g., Yoder, 1998),
> - use bioassessment information to evaluate the condition of aquatic life in all waterbodies,
> - establish biologically based aquatic life use designations,
> - protect aquatic life use standards with narrative or numeric biocriteria,
> - regulate pollution sources and assess the effectiveness of water quality management efforts, and
> - communicate the condition of their waters.

are required under the CWA for waterbodies that do not meet their designated uses. Citizen action groups initiated lawsuits in the mid 1990s charging that states and EPA were not fulfilling the CWA requirements regarding TMDLs. Because of the lawsuits and associated settlements, nationwide attention is now directed toward an accelerated cleanup of impaired waters through the TMDL process.

To meet the need for a consistent, defensible method for assessing cause, the Office of Research and Development (ORD) and Office of Water (OW) developed SI guidance (U.S. EPA, 2000). The guidance, summarized briefly in Appendix A, provides a methodology for identifying the causes of observed biological impairments in aquatic systems.

The SI process is prompted by biological assessment data that indicate that biological impairment has occurred. The general SI process entails critically reviewing available information, forming possible stressor scenarios that might explain the impairment, analyzing those scenarios, and producing conclusions about which stressor (or stressors) is causing the impairment. SI facilitates the development of defensible, high-quality, and effective remedial strategies for impaired surface water by ensuring that management actions are directed at the real cause of the impairment.

2.2. WHY CADDIS IS NEEDED

The primary goal of the CADDIS project is to facilitate the determination of causes of aquatic impairments. Although the SI guidance document provides a useful methodology for addressing the need for causal analysis, it can be difficult to implement. First, the organization and analysis of information and the formulation of logical inferences can be complex. Second, the time required for acquisition and interpretation of the volumes of information needed for defensible causal analyses is not available to TMDL developers, policy makers, and decision-making officials who are faced with demanding court-ordered schedules. CADDIS will alleviate these problems by guiding state, Regional, and tribal investigators through the SI process and providing easy access to relevant information and knowledge. Specifically, CADDIS will: (a) provide easier access to supporting information, (b) provide support in organizing and combining site-specific data with other information, (c) assist in performing the logical inferences, and (d) facilitate the sharing of experiences and knowledge gained in conducting causal evaluations. Using CADDIS, defensible, transparent, and repeatable SI results will be much more readily attained.

The need for a system to aid in causal assessment was noted by the authors of the SI guidance document. It was first articulated in the recommendations of the OW/ORD Strategic Planning and Research Coordination (SPRC) Diagnostics Group, which met in May 2000, and it is reflected in the diagnostics section of *Aquatic Stressors Framework and Implementation Plan*

for Effects Research (U.S. EPA, 2002c). The National Academy of Sciences recently highlighted the need for methods and models that link stressors to biological responses in order to effectively use biological criteria in the TMDL program (NRC, 2001). The need for a system such as CADDIS has also been indicated in two recent reports: the General Accounting Office's (GAO's) *Water Quality: Key EPA and State Decisions Limited by Inconsistent and Incomplete Data* (GAO, 2000) OW's *Twenty Needs Report: How Research Can Improve the TMDL Program* (U.S. EPA, 2002d).

The GAO report indicates that there is strong evidence of the need for additional analytical methods and technical assistance to help states analyze complex pollution problems and develop TMDLs. A survey of state water quality data managers suggested that EPA:

> ...should develop sample or standardized approaches, such as templates, that states could use to guide them through certain types of TMDLs. In addition, several states pointed out the need for efficiency in developing TMDLs. For example, one state noted that states should be benefitting from others' experiences in developing TMDLs, rather than "reinventing the wheel."

The GAO report also describes how the diversity of approaches to identifying water quality impairments and culpable stressors used by the different states results in inconsistent reporting. The use of CADDIS to identify stressors should improve the consistency in reporting by the states, thereby making comparisons across states and classes of stressors more reliable.

OW's Twenty Needs Report summarizes areas identified by TMDL practitioners where research efforts can improve the TMDL program (U.S. EPA, 2002d). Of the needs specified in the report, CADDIS will aid in addressing the following:

- Develop "state-of-the-science" syntheses in several high-priority subject areas to aid busy TMDL practitioners and decisionmakers,

- Mutually improve networking and access to expertise in ORD, OW, and EPA Regions,

- Provide ORD technical support and technical information transfer,

- Increase the quantity and quality of completed TMDLs,

- Improve the science base concerning all stressors (pollutants and pollution) and their impacts, and

- Address numerous stressor-specific issues identified through the SPRC process.

The existing stressor-specific protocols for TMDL development and the SI guidance document address some issues identified by the GAO and OW reports. However, CADDIS will meet additional identified needs and will greatly increase the efficiency of technical support by linking investigators with state-of-the-science information, conceptual model building tools, and instructive case studies that would otherwise be unavailable or difficult to access.

2.3. THE CADDIS DEVELOPMENT PROCESS

The development process described here emphasizes three components that will be developed in roughly parallel time frames and interwoven to form the complete system (Figure 2). The first component is the development of the CADDIS platform, which provides the user interface; the logical structure for analyzing, organizing, and synthesizing causal evidence; and the mechanisms for accessing relevant information. The second component is the development of the knowledge base that, when combined with site-specific information, provides the substantive basis for causal inference. The third component is the development of case studies that provide concrete illustrations of the process and a mechanism for sharing experiences and increasing investigator expertise.

An essential attribute of our proposed approach to development is the solicitation and incorporation of user input throughout the process. To this end, the development team will include members of the user community along with scientists, programmers, and project managers. Depending on time availability, input from user community members may include detailed reviews of interim products and drafts, comments on design features and priorities, and development of case examples. The CADDIS team will also solicit feedback from the broader user community and incorporate its experiences into refining the system design and the underlying SI process. Plans for outreach, training, and technical support are discussed in Chapter 7.

The CADDIS platform will be developed using a phased and modular approach (described in greater detail in Chapter 3). A phased CADDIS development process has many benefits, including the ability to: (a) provide some support quickly, (b) solicit early user and management feedback, (c) keep up with changing technology (e.g., operating system changes, browser upgrades), (d) allow the development team to showcase successes quickly and often, and (e) keep the system focused on, and relevant to, changing user needs. In addition, specific functions of CADDIS will be developed as modules that can be accessed independently. A modular CADDIS will: (a) make the system easier to update or upgrade, (b) allow a broader user
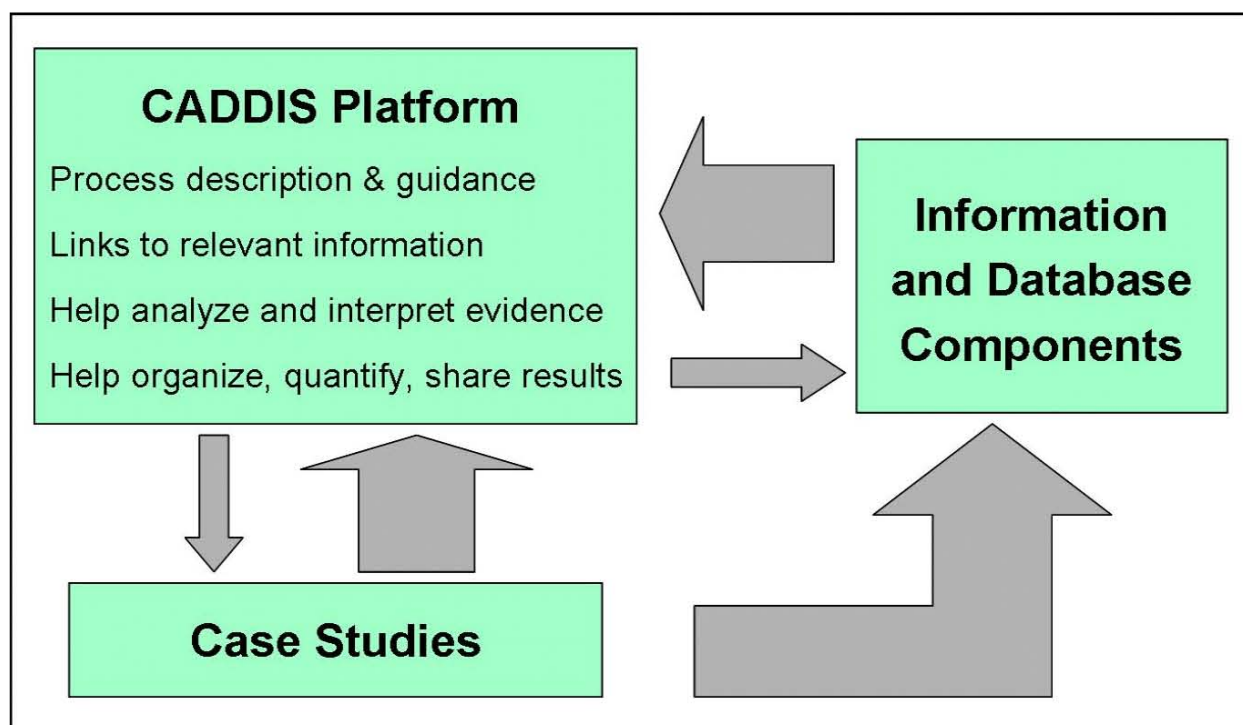
Figure 2. The CADDIS development process emphasizes three components (shown in the green squares) that will be interwoven to form the complete system. Concurrently with the three components, training will be developed and the stressor identification guidance document (U.S. EPA, 2000) will be revised.

base to take advantage of individual components of the system (e.g., a conceptual model builder would be useful in risk assessment as well as causal analysis), and (c) make it easier to match the strengths of different system platforms with functional needs.

The development of the knowledge base, described further in Chapter 4, will also proceed in a phased, modular approach, targeting the highest priority needs first. For example, the development of stressor-response information has consistently been ranked as a high priority need. The development of the CADDIS knowledge base will build on information and expertise that already exists or is being developed within ORD and OW. The CADDIS development team plans to form partnerships with others within and outside EPA in order to allow CADDIS users easy access to databases developed for other purposes but containing information useful for causal evaluations. All elements of the knowledge base will be easily and readily accessible in each version of CADDIS through an Internet portal. Through a query interface, the investigator will be able to search for and extract data useful for a particular case. As CADDIS becomes more advanced, the system will have the capability to perform the query on the basis of the user specifications and extract the relevant information directly into a useable format for the

investigator. Every effort will be made to ensure that the information imported and exported through CADDIS will be compatible with the user's own data management and processing tools.

Development of case studies is another important component of CADDIS and is needed to continue the refinement of the SI process (Chapter 5). Case studies will be developed to illustrate the application of the process to different impairments, stressors, and systems. In early versions, the case studies would be examples provided by EPA. However, as the functionality of CADDIS improves and as outreach activities increase the community of users, CADDIS will provide the option of uploading and storing user cases in a searchable database. This Internet-accessible, central repository of case studies is seen as a way to enable the community as a whole to improve expertise in causal evaluations. Making the database searchable will allow users to find situations similar to the ones they are evaluating.

Attention to quality assurance (QA) throughout the different aspects of development will ensure confidence in the results obtained using the system, increase transparency in the quality of information that underlies causal evaluations, and help users describe the uncertainty in their conclusions. QA discussed in Chapter 6. Finally, we close the circle of user involvement with communication and outreach activities to solicit ad hoc input to the development process, to train users, to obtain feedback once the system is in use, and to advance the development of causal assessment methods (Chapter 7).

3. THE CADDIS PLATFORM

Participants in the CADDIS development workshop provided fundamental input to the system's design, platform, and architecture, including identification of key functions, tools, and information resources. This chapter describes the system's platform development. The platform is defined as the portion of CADDIS that provides the user interface; the logical structure for analyzing, organizing, and synthesizing causal evidence; and the mechanisms for accessing relevant information. The CADDIS team will consult with both ORD and other EPA information technology staff early and often throughout the development process to ensure that the product adheres to all Internet guidelines put forth by EPA (U.S. EPA, 2003b), that it is compatible with EPA system requirements, and that it is in the public domain.

The system will be offered for use in three main versions: CADDIS 1, CADDIS 2, and CADDIS 3. The three versions build on each other sequentially in a resource-efficient path forward for development and implementation. The system is intended to evolve from a tool that provides streamlined access to information in a static environment to an adaptive framework applied by the user to support and document the decision-making process for individual sites.

Phased development also allows for the incorporation of user feedback into the next stage of development. Because of this iterative development plan, the final product may differ somewhat from what is described here.

3.1. CADDIS 1

The first version of CADDIS will be a simple, Internet-based resource that provides step-by-step prompts and advice. The user will also be assisted in recording site data onto blank forms and tables and will be provided with warnings about pitfalls at critical stages. In addition to walking the user through the SI process, CADDIS 1 will enhance the user's ability to conduct a causal evaluation by providing access to relevant supporting information. Because CADDIS 1 will include links to outside sources of information, platform development will include a query interface that allows users to precisely collect the information most useful to their site-specific SI (e.g., information on specific stressors, effects, or geographic locations). CADDIS 1 will also provide examples in the form of completed worksheets and reports. These examples will come from EPA-developed case studies that represent a variety of biological effects, stressors, regions, and flowing-water systems, and they will seed a larger database that will eventually include information and knowledge contributed by CADDIS users in the SI community.

3.2. CADDIS 2

The second version of CADDIS will enable investigators to use site-specific information more interactively with the system. This version will provide step-by-step prompts and advice in a question-and-answer format, interact with user-supplied information from sites, and automatically fill out forms and tables. Supporting information will be collected from EPA databases queried by the user, through CADDIS 2. Conceptual models will be included that may be customized by the user and a searchable library of case studies will be provided. The final output of CADDIS 2 will include supporting text, QA statements, and a bibliography, in addition to the tables and forms.

The functional aspects of CADDIS 2 will likely require a PC-based platform with Internet access. Migrating to a PC-based system for CADDIS 2 will provide greater security for user input information by allowing users to save their SI project locally and will allow users to generate custom reports. A Web-based system may also be considered if security concerns and user access needs can be addressed. In either case, CADDIS 2 will have a more complex query interface due to the inclusion of the wider array of data modules that will be incorporated into the system. In addition to the enhanced query interface, a simple rule base may be incorporated into the platform of CADDIS 2. A rule base is the component of a system that contains the rules describing the problem-solving knowledge (i.e., a collection of IF [conditions]–THEN

[hypothesis] statements). Rule bases may be used to move users through the SI process or to access information in databases not constructed primarily for the system. In the latter case, the rule base would be used to process the data in order to make them directly applicable to causal evaluations. Any rule base that is built into the platform will need to undergo a peer review process that is separate from the QA conducted on the program logic or basic system design.

3.3. CADDIS 3

Improvements to the CADDIS system will be heavily based on user feedback from the initial versions; therefore, it is difficult to predict the key functions of CADDIS 3. However, it is likely that the platform will be PC-based with Internet access because of data security and access issues similar to those mentioned for CADDIS 2. We suspect that the final version of CADDIS will be similar to CADDIS 2 in that it will provide step-by-step prompts and advice in a question-and-answer format and automatically fill out forms and tables. Advancements are anticipated in the form of increased automation. As an example, CADDIS 3 could generate and update conceptual models on the basis of site-specific input. These custom conceptual models can serve as the basis for calculating probabilities that a stressor is associated with the impairment. In addition, CADDIS 3 could generate custom species sensitivity distributions using site data and stressor-response data harvested from databases (see Section 4.2, Figure 5). Quantitative diagnostics models are an area of active research (e.g., U.S. EPA, 2002c), and CADDIS 3 will be designed to capitalize on advancements. Quantitative decision analysis modules would likely require the development of an advanced rule base or decision logic, which will carry significant peer review and QA requirements above and beyond what is needed for the basic platform design.

4. INFORMATION AND DATABASE COMPONENTS

A collection of databases that provide key information for investigators are the engines that will make CADDIS a truly useful product. The development and synthesis of this knowledge is a primary function of the CADDIS project.

The following sections summarize key information and database needs. Whenever possible, CADDIS will take advantage of existing databases. In these cases, the role of CADDIS is to create links (e.g., query interfaces) that allow this information to be more easily used for SI. The development team also will prepare state-of-the-science syntheses to assist investigators in finding and summarizing relevant information. Sometimes, information is available, but it has not been brought together in a form that can be queried. In these cases, the CADDIS team will

develop databases and encourage the complete reporting of data from research projects to increase utility to CADDIS users. Additionally, functionality may be developed to allow users to import or export spatial information between geographic information systems or spatial models and CADDIS 2 and 3. The CADDIS team will also synthesize data into summaries that provide easy access to pivotal results.

The CADDIS team will identify promising ongoing and future research projects using existing communication avenues such as OW/ORD joint planning meetings and ORD's multi-year planning process to communicate the objectives and needs of CADDIS. We will encourage complete reporting of data from research projects to increase utility to CADDIS users. The development of CADDIS will reveal information needs and spur additional applied research useful for causal assessments.

Within each category of information, the CADDIS team will first target the highest priority stressors and responses for development. High-priority stressors identified through case studies and needs of the TMDL program include clean sediments, habitat alteration, metals, dissolved oxygen (DO), nutrients, and temperature. High-priority responses include changes in fish and macroinvertebrate assemblages and stakeholder-valued fish populations.

4.1. CONCEPTUAL MODELS

A Connecticut Department of Environmental Protection stakeholder defined conceptual models as "exactly what you would need to put in front of managers/stakeholders." A diagram that depicts the connection from types of source(s) to candidate cause(s) and intermediate effects to the specific biological impairment can be invaluable for identifying the types of data that might be useful. Conceptual models communicate the ecological basis for suspecting a candidate cause, organize the evidence, and ultimately illustrate findings of the causal characterization to stakeholders and managers (Figures 3 and 4).

CADDIS 1 will provide downloadable models in a form that can be easily adapted for a particular case. The files will also contain citations supporting the different pathways in each model. CADDIS 2 and 3 may enable users to interactively modify conceptual models and may include the capability to link users directly from the model to relevant data, case studies, or other resources. In addition, these later versions of CADDIS may integrate the conceptual models with tabular summaries of the evidence, providing a more visual presentation of the case.

CADDIS will provide conceptual models that describe causal pathways for high-priority sources, stressors, and biological impairments. For example, conceptual models will be developed for commonly recognized causes listed in OW reports and biological assessment results used by the states and tribes as the bases for their biological criteria. These conceptual models will build on existing generalized models (e.g., Karr et al., 1986). The CADDIS team

11
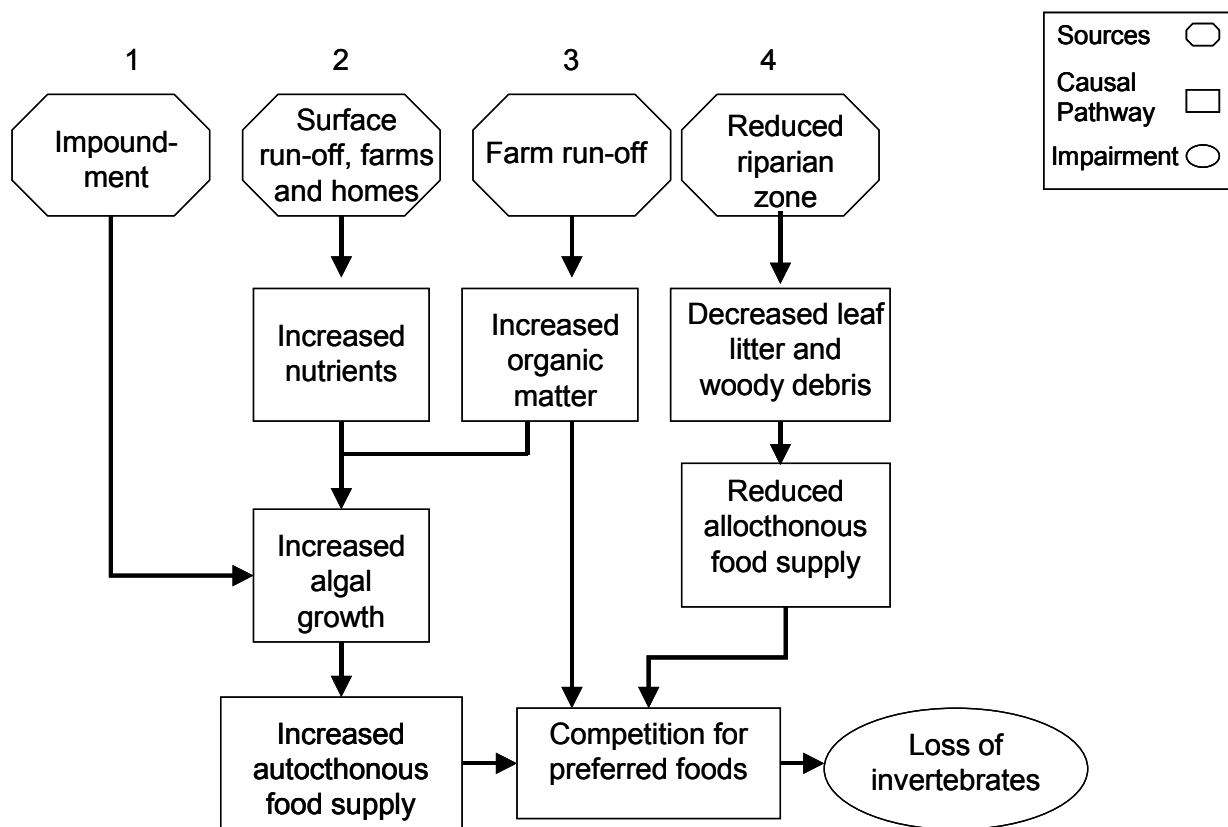
Figure 3. Example of conceptual model for altered food composition that causes the loss of invertebrates in a particular case. Sources are indicated in octagons, the causal pathway by boxes and arrows, and the biological effect in the oval.

will also aim to develop conceptual models for impacts on specific life stages of highly valued species, such as spawning, juvenile survival, or migration.

4.2. STRESSOR-RESPONSE RELATIONSHIPS

Tools for describing stressor-response relationship have been consistently mentioned by stakeholders as a high-priority need. These relationships often provide key evidence when comparing the strength of evidence for different candidate causes.

The CADDIS team will capitalize on existing databases, build additional data modules to fill data gaps, and synthesize stressor-response information for use in causal evaluations. The existing databases we intend to build on are ECOTOX (U.S. EPA, 2003c) and the Toxicity/Residue database (U.S. EPA, 2003d), which are invaluable sources of stressor-response information for chemicals. CADDIS will assist users in extracting and using information from ECOTOX for SI. The team is currently designing a database for information that is not covered in existing resources, specifically information from observational studies, empirical stressor-

12

Figure 4. The conceptual models above depict two individual causal pathways that may contribute to loss of aquatic invertebrates in a particular case. Evidence is shown in the rectangles with lifted corners. The causal pathway shown on the left (loss of suitable habitat from increased fines) is not strongly supported (X), in part because a site located between the source (farm run-off) and the impaired site did not exhibit a loss of invertebrates. The right-hand figure depicts how the loss of riparian vegetations might cause the loss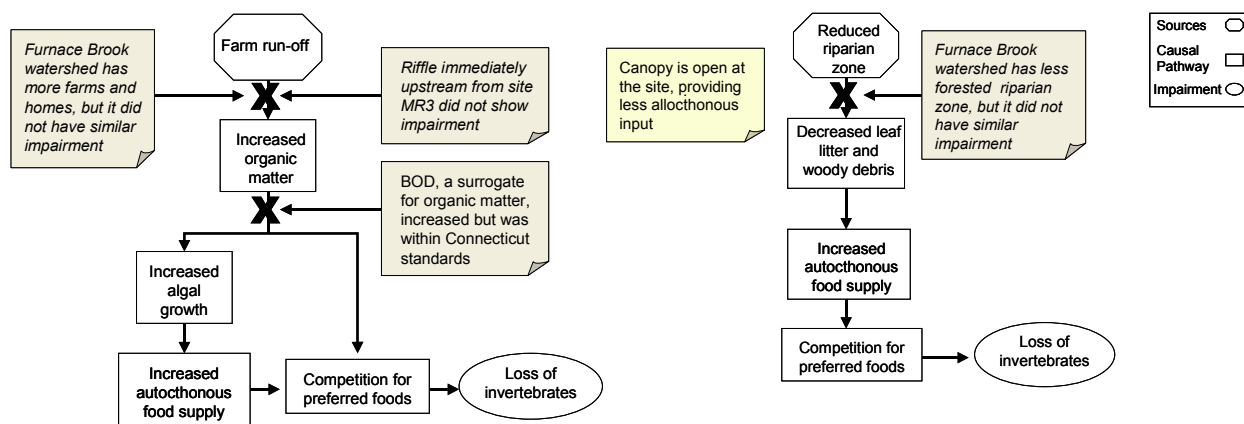 of invertebrates. In this case, evidence is presented that supports and refutes this source and causal mechanism. (BOD = biological oxygen demand).

response models, and on nonchemical stressors such as sedimentation, DO temperature, and habitat alteration. Collection and syntheses of data on additional stressors of high priority, such as pesticides, ammonia, and nutrients, will rely on future funding and manpower.

The CADDIS team will also contribute to the syntheses of stressor-response information that will facilitate its comparison with site-specific stressor data. Exposure-response data will be distilled into synthesis documents, including concentration-response relationships for organisms and species sensitivity distributions (SSDs). An SSD is a cumulative distribution function that describes the variation in toxic responses among a set of species to a certain compound or mixture (Posthuma et al., 2002) (Figure 5). Where data are available, SSDs available through CADDIS will address chronic exposure data as well as specific effects data, including sensitive endpoints and life stages. The syntheses will include the annotated SAS code and the Microsoft Excel template used to generate SSDs. Functionality in CADDIS 2 or 3 might include automated generation of site- or species-specific SSDs that combine user-entered data with information gathered from available databases. Selecting individual points in the CADDIS-generated SSD would allow the user to view specific information about how the data were collected, and, in cases where the data point originated from dose-response data, the dose-response curve could be provided. Additionally, CADDIS could allow the user to modify the SSD figure or even evaluate the impact of individual data points.
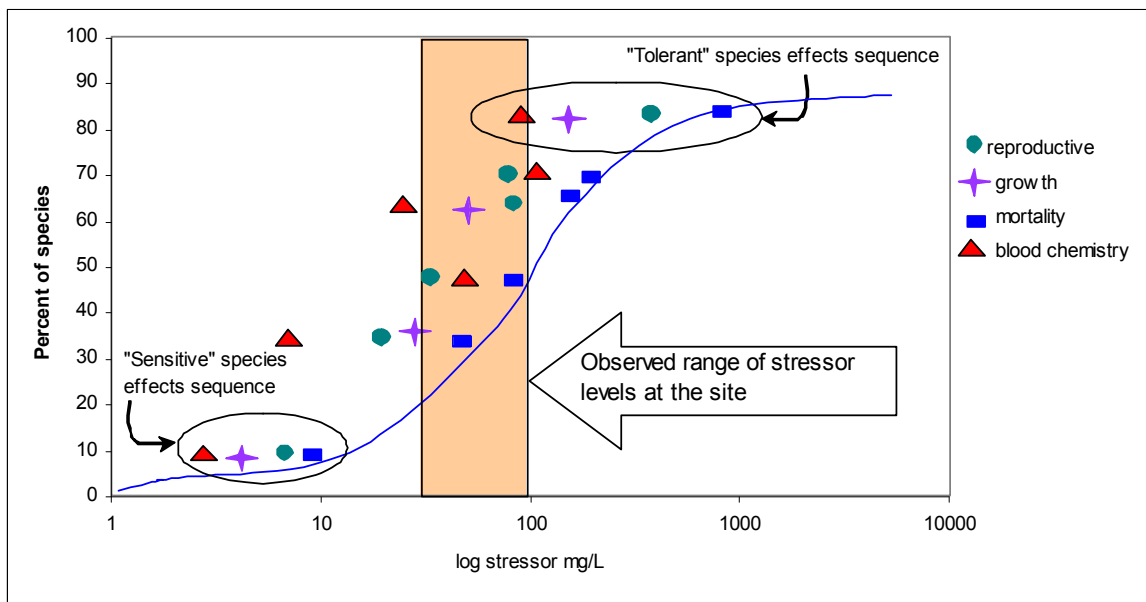
Figure 5. The species sensitivity distribution (SSD) plot illustrates multiple effects and the stressor intensity at which they were observed. A custom SSD plot would illustrate multiple effects, the stressor intensity at which they were observed, and the disposition of site specific data relative to SSD data (shaded area).

4.3. TOLERANCE VALUES

Assigning tolerance values to organisms has been a useful approach for developing indices (e.g., Hilsenhoff, 1987; Karr et al., 1986) that can then be used to estimate the degree of environmental degradation at sites. In the best cases, tolerance values can be used to develop empirical stressor-response relationships on the basis of assemblage composition that then can be used to evaluate whether effects would be expected given the level of a stressor. For example, a metric developed by the Idaho Department of Environmental Quality uses taxon-specific temperature tolerances to relate the benthic macroinvertebrate assemblage composition to stream temperature (Brandt, 2003) (Figure 6).

Ideal tolerance values are quantitative, stressor-specific, and reproducible. There are currently no agreed-upon methods for developing tolerance values, and no repository of tolerance values has been developed to date. An EPA regional tolerance value workgroup is providing the focal point for this work. A workshop planned for fall 2003 will begin the process of developing consensus on appropriate methods for tolerance value development and identifying tolerance values for compilation in a database. We anticipate that the CADDIS team will provide assistance in populating the database and developing queries so that the information can be brought into causal evaluations. Additional work may include cross-validating tolerance values

14
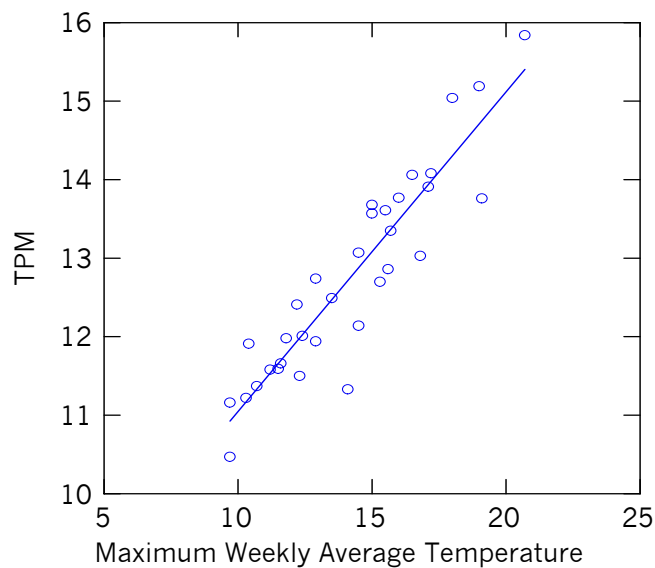
Figure 6. The relationship between Idaho's temperature preference metric, based on species tolerance values and maximum weekly temperature.

Source: Brandt (2003).

with the stressor-response information discussed above and confirming tolerance values in laboratory studies.

4.4. MITIGATION AND RESTORATION RESULTS

Providing access to information about what, where, and when mitigation or restoration activities have been effective can be valuable in assessing causality in similar situations. Specifically, experimental manipulations are used as a line of evidence when comparing or eliminating candidate causes. For the purposes of SI, an experiment is defined as the manipulation of a candidate cause by eliminating a source or altering exposure for the purpose of evaluating its relationship to an effect (i.e., mitigation and restoration results).

Several sources of mitigation and restoration information currently exist. For example, the National Risk Management Research Laboratory has developed an inventory of restoration projects within the mid-Atlantic region. This inventory has been incorporated into the OW's National River Corridor and Wetland Restoration database (U.S. EPA, 2003e). Another major effort is underway to develop a National Riverine Restoration Science Synthesis, which will provide information about restoration practices and their effectiveness. This effort is being led by American Rivers in partnership with several academic, governmental, and research center organizations. An additional example is the National Stormwater Best Management Practices Database (UWRRC, 2003) developed by the Urban Water Resources Research Council under a

cooperative agreement with EPA. Finally, mitigation information is provided to EPA under the TMDL program, Non-Point Source program, and other programs. It would be very useful to collect this information and make it accessible from CADDIS.

The CADDIS team will monitor the progress of these databases and facilitate their development when possible. CADDIS 1 will link users to these information sources; later versions will aim to help users extract and use this information for SI.

4.5. DIAGNOSTIC INFORMATION

In human and veterinary medicine, the causes of diseases are diagnosed by examining symptoms and determining which cause is indicated by the observed symptom set. Similarly, CADDIS 1 will provide worksheets to guide users through the diagnostic steps and direct them to references currently available for fish kills (e.g., Meyer and Barclay, 1990), fish diseases (U.S. Fish and Wildlife Service, 2003), and wildlife diseases (U. S. Geological Survey, 1999). CADDIS 2 and 3 may have more interactive systems that will prompt the user to make specific observations or tests that will enable the cause of aquatic impairments to be determined on the basis of symptomology.

Diagnostic keys have not been developed for some of the more common types of biological impairments, for example, those identified through changes in assemblages. Diagnosis of these impairments is more difficult because stressor-specific symptoms are unknown or the observed changes are common to many causes. For instance, an impairment that is frequently described is a decrease in the abundance of Ephemeroptera (mayflies). Unfortunately, there are too many possible causes of this effect for any one to be proven by diagnostic methods. Nevertheless, there is hope that future research may reveal symptomatic characteristics from more than one taxon and at different levels of organization, such as a combination of genetic markers, biochemical properties, physical deformities, sensitive taxa or species, or other characteristics that will provide a diagnostic set of symptoms. A number of researchers have made inroads into these emerging capabilities (e.g., Norton et al., 2000, 2002; Yoder and Rankin, 1995; Riva-Murray et al., 2002), and CADDIS will be designed to accommodate the latest developments in this area.

4.6. TOXICITY IDENTIFICATION EVALUATION RESULTS

Toxicity identification evaluations (TIEs) put field-collected samples through a battery of physical/chemical manipulations coupled with toxicity tests (U.S. EPA, 1991, 1992, 1993a, 1993b). Determining which physical/chemical manipulations affect toxicity of the samples provides useful evidence for identifying the causative agent. Developing a database of TIEs would have utility for similar cases. Users could determine whether the TIE approach had been applied successfully in a similar situation to the one under investigation, or they could use past

16

TIE results for similar cases to identify potential candidate causes (e.g., toxic components of similar effluents).

The CADDIS team will monitor the progress of TIE applications and field validation efforts and facilitate the development of a TIE database.

4.7. STATISTICAL AND PROCESS MODELS

Statistical and process models have many potential uses for SI. Some model results, such as stressor-response relationships or tolerance values, will be incorporated into the databases described in the sections above. However, statistical and process models have many other potential applications, ranging from quantifying consistency of association for particular regions to exploring possible causal mechanisms. Statistical models can describe how often an effect and candidate cause occur together in different places or times, and they can be used to develop surrogates for causes that can be difficult to measure. For instance, estimates of flow can be generated on the basis of rainfall accounts, nearby gauging stations, cover, soil type, and topography. Water temperature extremes might be estimated on the basis of some water temperature data, air temperatures, and information about cover and rain events. Future possibilities include classifying watersheds according to their vulnerability to different stressors or estimating the likelihood of observing particular stressors on the basis of landscape level information. Process models based on knowledge of ecological mechanisms (e.g., Aquatox, BASS) can be used to explore the consequences of different stressor scenarios. When compared with data from the site, they can be used to evaluate whether observations are consistent with ecological theory.

Statistical and process models are an area of active research. Because of the wide variety of potential models and uses, CADDIS will direct users to specific models when it can be shown how they may be applied in SI (e.g., indicate how they were used in a case study or describe the specific systems for which model results have been verified). Model results will be incorporated into appropriate databases so that they can be queried. Finally, the CADDIS team will monitor promising research, facilitate model development and application to SI, and aim to accommodate the latest model developments.

5. CASE STUDIES

Case studies that provide concrete illustrations of the process and a mechanism for sharing experiences and increasing investigator expertise are an essential component in the continuing refinement of the SI process. We plan a close interaction between the development of these case studies and CADDIS. The case studies will help identify critical information

needs, help test CADDIS functions (e.g., report generation), and provide a way to obtain feedback on CADDIS and the SI process from users. Incorporating a database of case studies into CADDIS will help to create a community that shares and builds a knowledge base of causal analyses and mitigation information. Making the database searchable allows users to find situations similar to the ones they are evaluating and reduce the effort required to perform causal analyses that are relatively similar. Depending on feedback, resources, and QA requirements, a future feature may include a mechanism by which users can input their own case studies into the database.

Four case studies that have either been developed or are scheduled for activity will be included in CADDIS to illustrate how SI is being used to determine the cause of impairment:

- Little Scioto, OH: The Little Scioto River was determined to be impaired on the basis of results of macroinvertebrate and fish surveys. Candidate causes investigated included habitat alteration from channelization, toxic substances from historical and current industrial uses, and excess nutrients from agricultural land uses and municipal waste treatment.

- Willamantic, CT: The Willamantic River in Connecticut was determined to be impaired, based on the results of macroinvertebrate surveys. Candidate causes investigated included toxic substances from historical and current industrial uses, nutrients from upstream agriculture and a municipal waste treatment plant, and flow and sediment changes from impoundments.

- Long Creek, ME: Long Creek is located in a rapidly urbanizing watershed and no longer sustains a brook trout population. Candidate causes of the impairment under investigation include excess nutrients, high temperature, low DO, flow regime changes, and toxic substances.

We plan to develop several additional case studies during a problem-solving workshop in which waterbodies identified as impaired by EPA Regions, states, and tribes will undergo causal analyses by teams of interested state, tribal, Regional, federal, and academic environmental professionals. This effort will serve the dual purpose of training scientists to do causal analyses and fostering a scientific community approach towards addressing environmental protection.

In selecting case studies, the team has sought to represent a variety of biological effects, stressors, regions, and flowing-water systems. Additional case studies may be developed, perhaps extending the concepts to lakes, estuaries, coastal systems, wetlands, and terrestrial ecosystems. We also will seek opportunities to coordinate and collaborate with case studies under development by other groups within EPA.

6. QUALITY ASSURANCE

This chapter provides the first steps toward outlining a CADDIS Quality Assurance Project Plan (QAPP) and discusses issues affecting each CADDIS component. A single general QAPP will be developed for CADDIS to ensure cohesiveness and consistency in system quality. However, documentation of QA is required for the major components of CADDIS, necessitating individual QA plans for each element of the system as it is developed. Section 6.2 contains a description of QA issues associated with platform development, defined here as the design, programming, and rule bases needed for queries or interaction with user-supplied data. QA issues associated with data modules, which encompass the modules developed by the CADDIS project (i.e., databases of refereed literature and gray literature), data referred to by CADDIS (e.g., ECOTOX), and data input into CADDIS by users (e.g., case study examples), are discussed in section 6.3. Because not all system and data elements will be used in all versions of CADDIS, the level of QA needed will increase with increasing system complexity. This will be reflected in both the general QAPP and the individual component-specific QAPPs.

6.1. THE OVERARCHING QUALITY ASSURANCE PROJECT PLAN

General issues of QA that affect all aspects of CADDIS will be addressed in the overarching QAPP. Data quality objectives selected for CADDIS components can strongly influence the eventual magnitude, complexity, functional capabilities, and reliability of CADDIS. The generic QA principles that should guide the QAPP development are:

- Every function designed into CADDIS should be periodically verified as performing to specifications that are defined in individual QAPPs,

- The level of effort devoted to QA of CADDIS components will be commensurate with the purpose of each component and the consequences of encountering error,

- Satisfactory performance of the system at large will be ensured through periodic checks of the specifications list in the QAPP,

- 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,

- CADDIS will provide users with information on the quality and source of data supplied for use in their analyses, and

- Information resource selection criteria will be developed objectively and stated clearly to be eliminate any appearance of bias.

The QAPP will contain information on project management, data acquisition, assessment/oversight, and data validation and usability. The project management section will identify the project's organization, problem definition, quality objectives, and necessary documentation (e.g., individual component QA plans).

The data acquisition section will demonstrate that the intended data acquisition and processing methods are appropriate for achieving project objectives; it will include design, specifications, and performance and acceptance criteria for each component. Data acquisition for CADDIS will consist of the gathering, synthesis, and analysis of secondary data. Because the acquisition varies significantly with each system version, this will be covered in detail in the individual component-specific QA plans. However, it will be important to document how the components will be incorporated and managed into the project's data management system. This may be described through "requirements documentation," which describes the characteristics and behaviors that the system must possess to function adequately for its intended purpose, including scientific defensibility of the approach. The documentation should be written with sufficient detail so it can be used as the foundation for design and testing.

In order to ensure consistent quality across all components, this section of the QAPP will include a description of the CADDIS design, specifications, and performance and acceptance criteria for each component. The design and specification documentation is a description of how each component will meet its "requirements." Specifically, the section on design describes design layout, and the specification documentation describes how information is stored and transferred between components created by different developers.

The assessment and oversight section of the QAPP will include basic information on the assessments, response actions, and reports to the CADDIS management team. Assessments discuss how the data can be used to address the project objectives. Lists of planned assessments, performance and acceptance criteria, and methods for each assessment will not be included in the general QAPP because they are dependent on the system component or database. The QAPP may, however, identify which of the following general reports to management are required:

- Technical project and budget reports,
- QA Project Plan deviations and impacts,
- Need for and results to corrective actions,
- Calibration reports,
- Assessment reports,
- Peer review reports, and
- System evaluation reports.

The data validation and usability section of the QAPP identifies the assessment procedures needed for confirming that the data are of the type and quality needed and ensures that the data use limitations are defined. Data validation methods should be written in a test plan, which might include peer review, code verification, validation of input data, or assessment of system output. This section also includes a reconciliation of final output assessment results, with user-defined requirements to accept, reject, or describe the uncertainty of the output for its intended use.

6.2. QUALITY ASSURANCE CONSIDERATIONS FOR INDIVIDUAL SYSTEM PLANS

This section describes the two major categories of system QA considerations: 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, and peer review will be conducted for the system design document and any related prototypes. System programming QA requirements will vary with the functionality and tools that eventually make up CADDIS. These may be incremental, and thus, compatibility with pre-existing CADDIS components and functions will be important.

Issues that the system-specific QA plans may address include:

- The use of automated unit test and systems tests that can mechanize some of the QA on an ongoing basis;

- Configuration management, or how components interrelate; and

- The use of paired programming to integrate 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 utilized by CADDIS. This type of QA will not be necessary until the latter versions of CADDIS, and its development may rank among the most complicated QA tasks. Both a peer review of the causal analysis assumptions and decisions and an assessment of the program functionality would be needed, although use of existing expert system shells may assist in the latter.

6.3. QUALITY ASSURANCE CONSIDERATIONS FOR DATA MODULES

Three major types of data will be used in CADDIS: data within CADDIS, data referred to by CADDIS, and data input into CADDIS by users. The knowledge base within CADDIS will be in the form of databases built primarily by gathering and synthesizing data from peer-reviewed and gray literature. QA requirements for these databases include not only QA of the

secondary data, but also QA of the database structure itself. Database developers will need to be cognizant of the structure and content of related databases to ensure consistency of assumptions and data transfer between systems. QA requirements for the data referred to by CADDIS and data input into CADDIS by users will focus heavily on documentation and ensuring user access to metadata records.

The largest component of CADDIS data may be causal analysis literature search and retrieval. For the most part, the QA responsibilities lie with the originators of this information, and the CADDIS QA role is more in characterizing than performing 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 keywords, external search engine, annotated bibliographies) will have additional QA implications that need to be addressed in module-specific QA plans.

Rules for characterizing QA will be necessary and should be included in the general CADDIS QAPP to ensure consistency of QA across information databases. Peer-reviewed sources will likely be the core of CADDIS literature data, but numerous gray literature sources may also be useful due to their often greater detail. Such sources are typically difficult to obtain, making their inclusion in CADDIS valuable. Peer review and publication are no guarantee of relevance or quality if the data or methods relevant to causal analysis were not specifically QA'ed. Gray literature, on the other hand, despite lacking refereed publication, may still contain well QA'ed causal information, but it will need to be verified and documented.

User input data are likely to be a growing source of data in CADDIS, and significant QA responsibilities will be associated with this resource. However, as with the QA considerations for literature, the QA burden lies primarily with the originator. CADDIS is responsible for characterizing and documenting its QA and making the metadata available to users. Further QA requirements, such as acceptance/rejection criteria, will need to be addressed in a module-specific QA plan.

7. OUTREACH, TECHNICAL SUPPORT, AND GUIDANCE

We plan to identify and reach out to representative user groups during all phases of CADDIS development. During the early phases, communication with users will help refine our understanding of their needs as well as encourage system use. After the system has been developed, we will add training and technical support components.

7.1. OUTREACH

Participants from regions, states, and local governments at the CADDIS workshop provided valuable input regarding all aspects of system development (U.S. EPA, 2002a). We will use various forms of communication, ranging from electronic interactions to training sessions, to expand our access to as broad a cross-section of the potential user community as possible. Anticipated activities include:

• Speaking and holding informational or training sessions at professional and coordinator meetings (e.g., EPA Region meetings, and meetings of Water Coordinator groups, the Council of State Governments, Society of Environmental Toxicology and Chemistry, North American Benthological Society, American Water Resources Association, the Bioassessment and Biocriteria Academy);

• Writing articles in newsletters and trade journals (e.g., OW's Watershed Events and Nonpoint Source Newsnotes, Ohio EPA Newsletter);

• Posting messages on an established listserv, e-mail distributions, and discussion databases (e.g., U.S. EPA Techloops, the Biological Assessment Discussion Database, STORET and NHD listservs, Water Quality Standards listserv);

• Speaking on established conference calls held by various groups;

• Creating a training module for the OW "Watershed Academy" website;

• Creating a feedback form on the current EPA SI Guidance website; and

• Developing a briefing page or presentation that demonstrates to the EPA Regions, states, and tribes how OW programs and ORD programs such as EMAP and CADDIS all fit together.

7.2. TECHNICAL SUPPORT

Once CADDIS is developed, users will need technical support. This will require a long-term commitment. The group has thus far considered three different approaches to technical support. The first, and most resource intensive, would be establishing a dedicated technical support team that could work on a one-on-one basis with users who call or e-mail. An alternative approach would be to establish a listserv or discussion database with a searchable archive function to allow users to help each other. The technical support team would monitor the listserv/database and respond to questions that could not be answered by the user community and ensure that the help offered by users was correct. The third method of technical support involves publishing an ongoing list of known system problems and solutions and any planned system improvements.

7.3. GUIDANCE

We anticipate that our experiences in developing CADDIS, conducting case studies, and interacting with users will lead to substantial advances in our knowledge of causal evaluation. This will be captured in a revision of the current SI guidance document.

8. TIMELINE AND OVERVIEW OF PRODUCTS

This development plan describes the development of CADDIS using a phased, modular approach. This approach will allow for early release of products while additional development takes place (Table 1). The products shown in Table 1 generally coincide with the annual performance measures (APMs) described in the multi-year plans for ORD research. As of the writing of this strategy, relevant APMs are included in the Multi-Year Plans for Water Quality (Goal 2) and Ecological Research (Goal 4). The detailed APM list is provided in Appendix B.

Table 1. Overview of CADDIS products and timeline

Product	2003	2004	2005	2006	2007	2008	2009
The CADDIS Platform							
CADDIS Development Plan	■						
Release CADDIS 1	■	■	■				
Release CADDIS 2	■	■	■	■	■		
Release CADDIS 3	■	■	■	■	■	■	■
Database and Information Components							
Methods/indicators for diagnosing impairments due to sediment	■	■					
Methods/indicators for diagnosing impairments due to toxic metals	■	■	■				
Conceptual models for high-priority causal pathways	■	■	■	■			
Case Studies							
Case study: causes of biological impairment (Willamantic)	■						
Case study: urban setting with non-point source (Long Creek)	■	■	■	■			
Collection of case studies				■	■	■	
Guidance and Training							
Watershed Academy website	■	■	■	■			
Problem-solving workshop: determining the causes of biological impairment		■	■	■			
Guidance on causal evaluation				■	■	■	■

9. PROJECT RISKS

The CADDIS project faces many challenges to becoming a widely used, effective product. We have identified some general risks and our intended approaches to minimizing their impact.

Risk: CADDIS is too difficult to use.
Mitigation:

- Release the system soon so users can identify areas/features that are too complex and provide feedback early,

- Provide simple interfaces through which users can report problems as they are using the system,

- Continue training workshops, and

- Allow investigators to use CADDIS with their own data analysis and word processing tools.

Risk: Users are unaware of CADDIS or cannot obtain it. Alternatively, value to resource-limited state and tribal users is insufficient to justify the time and effort necessary to learn a complex process. This is also known as "*We build it but no one comes.*"
Mitigation:

- Provide functionality that significantly improves on existing approaches to causal evaluation;

- Build linkages between CADDIS and Agency programs that use causal assessments (e.g., 303d listing efforts);

- Conduct outreach through existing lines of communication and incorporate CADDIS into an outreach program to provide assistance with site assessments;

- Continue outreach to: (1) identify critical gaps and information needs, (2) provide those critical functions and data, (3) make sure that users are aware that they can obtain the information/functions from CADDIS, and (4) make it easy for the users to obtain that information;

- Continue case studies to demonstrate the value-added and new features from CADDIS; and

- Become familiar with user requirements to determine the best method(s) of accommodating investigators' computing systems, such as:

  - retaining and updating simpler versions of CADDIS to accommodate a variety of user needs and computing requirements;

  - providing system as software that can be downloaded updated either online or on CDs mailed to the user; and

  - housing CADDIS on a server outside of EPA to avoid firewall, data security, and down-time issues.

Risk: EPA is perceived as giving a stamp of approval to poor-quality information if CADDIS is used to conduct causal analyses based on poor quality information.
Mitigation:

- Include metadata with records in the data modules.

- Increase awareness of the need for high-quality information and a sound biological assessment program (see the text box in Section 2.1).

Risk: Key team members leave and are not replaced.
Mitigation:

- Verify to those with intellectual investment that their efforts are resulting in constructive improvements in environmental decision making and that the value of these efforts are recognized by both management and the SI community;

- Seek management backing for project at the Assistant Administrator level and on the Research Coordination Teams;

- Provide avenues for continued growth and innovation;

- Recruit and maintain a quality CADDIS workgroup committed to excellence in environmental decision making and state-of-the-science approaches to environmental problem solving; and

- Expand the scope of the CADDIS program so that it does not become simply a maintenance activity; for example, add terrestrial and marine ecosystems.

Although the risks we have identified are not trivial, they are also not insurmountable, and the potential benefits of CADDIS are substantial. By defensibly directing management

actions toward the true cause of impairment, remedial and restoration strategies will be more effective in improving the quality of the nation's surface waters.

APPENDIX A

STRESSOR IDENTIFICATION PROCESS OVERVIEW[1]

[1]Modified from the U.S. EPA (2000)

A.1. THE STRESSOR IDENTIFICATION (SI) PROCESS

Figure A-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). Decisionmaker and stakeholder involvement is shown along the left-hand side; this 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 these data is indicated 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 tools and designs.

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); however, some of the criteria presented for evaluating evidence may be specific 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 core of the SI process is shown within the bold line of Figure A-1 and consists of three main steps:

Step 1. Listing candidate causes of impairment.

Step 2. Analyzing new and previously existing data to generate evidence for each candidate cause.

Step 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.

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
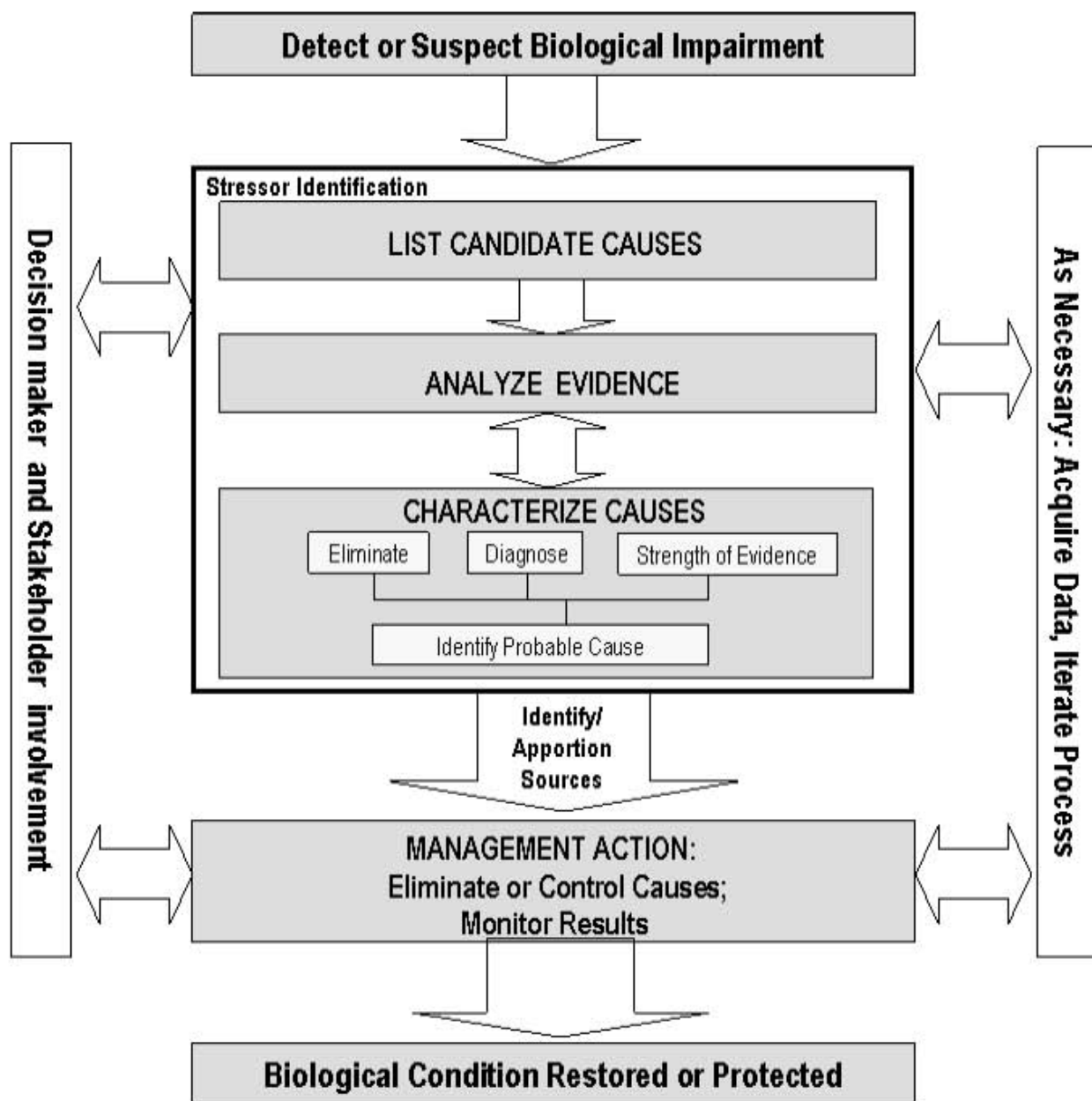
Figure A-1: The management context of the SI process. The SI process is shown in the center box with a bold line. SI is initiated with the detection of a biological impairment. Decision maker and stakeholder involvement is particularly important in defining the scope of the investigation and in listing 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.

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 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, and

- Cause and effect in deliberate manipulations of field situations or media.

These associations make up the body of evidence used to characterize the cause.

In Step 3, characterize causes, the investigator uses the evidence to eliminate, diagnose, and 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.

A.2. 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.

A.3. USING THE RESULTS OF SI

SI is only one of several activities required to improve and protect biological condition (Figure A-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 provide valuable feedback to the SI process (e.g., Yoder and Rankin 1998).

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.

APPENDIX B

**ANNUAL PERFORMANCE MEASURES (APMS) IN THE
MULTI-YEAR PLANS**

Table B-1. List of CADDIS APMs in the Ecological Research MYP[a]

Ecological Research Multi-Year Plan (Goal 4: Communities and Ecosystems) **LTG 2: Managers and researchers understand links between human activities, natural dynamics, ecological stressors, and ecosystem condition.**							
	2003	2004	2005	2006	2007	2008	2009
	Sub LTG D1: Environmental managers are able to characterize ecosystem properties and processes in a manner that supports diagnosis of current condition 2008						
	APG 55: Environmental managers are able to implement new, more efficient methods for stressor identification and characterization 2005						
Causal tools: Suter lead		*APM 199: Report on methods/ indicators for determining when biological impairments of rivers and streams are due to sediment Water Quality MYP, APG 9, 2004[b]*					

Table B-1. List of CADDIS APMs in the Ecological Research MYP[a] (continued)

Ecological Research Multi-Year Plan (Goal 4: Communities and Ecosystems) LTG 2: Managers and researchers understand links between human activities, natural dynamics, ecological stressors, and ecosystem condition.							
	2003	2004	2005	2006	2007	2008	2009
	Sub LTG D3: Diagnostic methods exist that incorporate knowledge of and contributions of multi-stressor interactions 2006						
	APG: Risk assessors can attribute causes of impairment when caused by more than one stressor input 2007						
Case Studies: Cormier lead				*Case study determining the causes of biological impairment in an urban setting with non-point source impacts so that states and tribes will have prototypes to facilitate completion of TMDLs (Water Quality MYP)[c]*			
Guidance and Training				*Watershed Academy website training for causal analysis (Water Quality MYP)[c]*			

Table B-1. List of CADDIS APMs in the Ecological Research MYP[a] (continued)

	Ecological Research Multi-Year Plan (Goal 4: Communities and Ecosystems) LTG 2: Managers and researchers understand links between human activities, natural dynamics, ecological stressors, and ecosystem condition.						
	2003	2004	2005	2006	2007	2008	2009
	Sub-LTG D4: Protocols are available to regional, state and watershed scientists and managers that provide the scientific basis for determining the cause of observed ecological effects 2009						
	APG: Methods are demonstrated that allow regional, state and watershed scientists to determine the causes of observed ecological effects 2009						
Case Studies: Cormier lead	*APM 95: Case study demonstrating the Stressor Identification Process that identifies the causes of biological impairment in the nation's waterbodies (EERD) Water Quality MYP, APG 16[b]*						

Ecological Research Multi-Year Plan (Goal 4: Communities and Ecosystems) LTG 2: Managers and researchers understand links between human activities, natural dynamics, ecological stressors, and ecosystem condition.							
	2003	2004	2005	2006	2007	2008	2009
CADDIS: Norton lead	Parameters and the architecture for a Causal Analysis and Diagnosis Decision Information System (CADDIS), that helps users identify causes of biological impairment in the nation's waterbodies[d]		Release Level 1 Causal Analysis and Diagnosis Decision Information System (CADDIS[1])[d]		Release Level 2 Causal Analysis and Diagnosis Decision Information System (CADDIS[2])[d]		Release Level 3 Causal Analysis and Diagnosis Decision Information System (CADDIS[3])[d]
	APG: Guidance on causal evaluation is made available to regional, state and watershed scientists and managers 2009						
Guidance and Training							Guidance on causal evaluation is made available to regional, state and watershed scientists and managers (Proposed as a synthesis document)[d]

[a]   Italicized APMs are cross-referenced in the Ecological Research MYP and the Water Quality Research MYP and are listed again in Table B-2. All text taken from the April 1, 2003, MYP Draft, with revisions as submitted December 2003.

[b]   Responsible laboratory = NERL

[c]   Responsible laboratory = NERL/NCEA

[d]   Responsible laboratory = NCEA/NERL

Table B-2. List of CADDIS APMs in the Water Quality Research Program MYP[a]

Water Quality Research Program Multi-Year Plan Timeline (Goal 2: Water) **LTG 2: Provide the tools to assess and diagnose sources and causes of impairment in aquatic systems**						
	2003	2004	2005	2006	2007	2008
	APG: Equip EPA Regions, States and Tribes with knowledge, skills and tools to determine the causes of impairments for freshwater and coastal systems required in various regulations 2007					
Case Studies: Cormier lead	Case study demonstrating the Stressor Identification Process that identifies the causes of biological impairment in the nation's waterbodies[c,]			Case study determining the causes of biological impairment in an urban setting with non-point source impacts so that states and tribes will have prototypes to facilitate completion of TMDL's.[c]		Case study focusing on the special needs to perform causal analysis in biologically impaired aquatic systems rivers.[c] Collection of case studies determining the causes of biological impairment, the scientific basis, tools and applications toward improving stream quality.[c]
Causal Tools: Suter lead		Report on methods/indicators for determining when biological impairments of rivers and streams are due to sediment[c]	Report on methods/indicators for determining when biological impairments of rivers and streams are due to toxic metals[d]	Conceptual models for high priority causal pathways[b,e]		

Table B-2. List of CADDIS APMs in the Water Quality Research Program MYP[a] (continued)

	2003	2004	2005	2006	2007	2008
Water Quality Research Program Multi-Year Plan Timeline (Goal 2: Water) **LTG 2: Provide the tools to assess and diagnose sources and causes of impairment in aquatic systems**						
Guidance and Training			Training and problem solving workshop: determining the causes of biological impairment, the scientific basis, tools and applications applied to state-listed 303d streams [c]	Watershed Academy website training for causal analysis[c]		

[a] All text taken from the April 28, 2003 MYP Draft, with revisions as submitted December 2003
[b] Responsible laboratory = NERL
[c] Responsible laboratory = NERL/NCEA.
[d] Responsible laboratory = NCEA/NERL.
[e] Not currently listed in MYP; proposing as new APM.

REFERENCES

Brandt, DT. (2003) A procedure to determine the biological gradient of macroinvertebrates to stream temperature. Presented at: National Biological Assessment and Criteria Workshop: Advancing State and Tribal Programs; March 31–April 4; Coeur d'Alene, ID.

GAO (General Accounting Office). (2000) Water quality: key EPA and state decisions limited by inconsistent and incomplete data. GAO/RCED-00-54. Washington, DC.

Hilsenhoff, WL. (1987) An improved biotic index of organic stream pollution. Great Lakes Entomol. 20:31–39.

Karr, JR; Fausch, KD; Angermeier, PL; et al. (1986) Assessing biological integrity in running waters: a method and its rationale. Illinois National History Survey Special Publication 5. 28 p.

Meyer, FP; Barclay, LA. (1990) Field manual for the investigation of fish kills. Fish and Wildlife Service Resource Publication 177. Washington D.C.: United States Department of the Interior, Fish and Wildlife Service.

Norton, SB; Cormier, S; Smith, et al. (2000) Can biological assessments discriminate among types of stress? A case study from the Eastern Corn Belt Plains ecoregion. Environ. Toxicol. Chem. 19(4):1113–1119.

Norton, SB; Cormier, S; Smith, M; et al. (2002) Predicting levels of stress from biological assessment data: empirical models from the eastern corn belt plains. Environ. Toxicol. Chem. 6:1168–1175.

NRC (National Research Council). (2001) Assessing the TMDL approach to water quality management. National Academy Press, Washington DC.

Posthuma, L; Suter, SW II; Traas, TP. (2002) Species sensitivity distributions in ecotoxicology. CRC Press LLC, Boca Raton, FL.

Riva-Murray, K; Bode, RW; Phillips, PJ; Wall, GL. (2002) Impact source determination with biomonitoring data in New York State: concordance with environmental data. Northeastern Naturalist, 9(2):127–162.

U.S. EPA (Environmental Protection Agency). (1991) Methods for aquatic toxicity identification evaluations, phase I, toxicity characterization procedures. 2nd ed. Office of Research and Development, Washington, DC, EPA/600/6-91/003.

U.S. EPA. (1992) Methods for aquatic toxicity identification evaluation: characterization of chronically toxic effluents, phase I. Office of Research and Development, Washington, DC, EPA/600/6-91/005F.

U.S. EPA. (1993a) Methods for aquatic toxicity identification evaluations, phase II, toxicity identification procedures for samples exhibiting acute and chronic toxicity. Office of Research and Development, Washington, DC, EPA/600/R-92/080.

U.S. EPA. (1993b) Methods for aquatic toxicity identification evaluations, phase III, toxicity confirmation procedures for samples exhibiting acute and chronic toxicity. Office of Research and Development, Washington, DC, EPA/600/R-92/081.

U.S. EPA. (2000) Stressor identification guidance document. Office of Water, Office of Research and Development, Washington, DC, EPA 822-B-00-25.

U.S. EPA. (2002a) Summary report for the workshop on the causal analysis/diagnosis decision information system (CADDIS). Office of Research and Development, Washington, DC, EPA/600/R-02/078.

U.S. EPA. (2002b) Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: streams and wadeable rivers. Office of Environmental Information and Office of Water, Washington, DC, EPA-822-R-02-048.

U.S. EPA. (2002c) Aquatic stressors: framework and implementation plan for effects research. Office of Research and Development, National Health and Environmental Effects Research Laboratory, Research Triangle Park, NC, EPA/600/R-02/074.

U.S. EPA. (2002d) The twenty needs report: how research can improve the TMDL program. Office of Water, Washington, DC, EPA 841-B-02-002.

U.S. EPA. (2003a) U.S. EPA biocriteria web page <http://www.epa.gov/waterscience/biocriteria/> (accessed August 12, 2003).

U.S. EPA (2003b) Web guide, updated June 18, 2003, <http://www.epa.gov/webguide/index.html> (accessed July 24, 2003).

U.S. EPA. (2003c) ECOTOXicology database. National Health and Environmental Effects Laboratory, Duluth, MN; <http://www.epa.gov/ecotox/>.

U.S. EPA. (2003d) Toxicity/residue database. National Health and Environmental Effects Laboratory, Duluth, MN; <http://epa.gov/med/databases/tox_residue.htm>.

U.S. EPA. (2003e) River corridor and wetland restoration database. Office of Water and National Risk Management Research Laboratory, Washington, DC; <http://www.epa.gov/owow/wetlands/restore/>.

U.S. Fish and Wildlife Service. (2003) National wild fish health survey database. U.S. Fish and Wildlife Service Division of Fish Hatcheries and Montana State University Environmental Statistics Group, Bozeman, MT; <http://wildfishsurvey.fws.gov/>.

U.S. Geological Survey. (1999) Field manual of wildlife diseases: general field procedures and disease of birds. Washington, D.C., U.S. Geological Survey Information and Technology Report 1999-001.

UWRRC (Urban Water Resources Research Council). (2003) National stormwater best management practices database. Urban Water Resources Research Council of the American Society of Civil Engineers, Denver, CO; <http://www.bmpdatabase.org/>.

Yoder, CO. (1998) Important concepts and elements of an adequate state watershed monitoring and assessment program, pp. 615–628. In: Proceedings of the NWQMC National Conference Monitoring: Critical Foundations to Protecting Our Waters. U.S. EPA, Washington, DC, 663 pp. + app.

Yoder, CO; Rankin, ET. (1995) Biological response signatures and the area of degradation value: new tools for interpreting multi-metric data. In: Davis, WS; Simon, TP, eds. Biological assessment and criteria: tools for water resource planning and decision making. Boca Raton, FL: CRC Press, Inc.; p 263–86.

Yoder, CO; Rankin, ET. (1998) The role of biological indicators in a state water quality management process. Environ. Mon. Assess. 51(1-2):61–88.