

September 22, 1998

EPA-SAB-DWC-ADV-98-004

Honorable Carol M. Browner
Administrator
U.S. Environmental Protection Agency
401 M. Street, SW
Washington, DC 20460

Subject: An SAB Advisory on the National Drinking Water
Contaminant Occurrence Database

Dear Ms. Browner:

On June 18, 1998, the Drinking Water Committee (DWC) of the Science Advisory Board (SAB) met to review the design phase considerations of the National Contaminant Occurrence Data Base (NCOD). Review of the NCOD is required in the 1996 Amendments to the Safe Drinking Water Act (SDWA). The review was conducted in a public session under the provisions of the Federal Advisory Committee Act (FACA).

This SAB advisory provides advice on an Agency work-in-progress. The goal of an SAB advisory is to provide suggestions to the Agency for mid-course corrections that will refine the ultimate product. In this case, the agency is engaged in the development of the NCOD in response to a firm deadline contained in the SDWA 1996. The SAB expects to conduct an additional review after the agency has completed this initial phase of database development and has considered updates to the system. At that time, a significant number of new participants will be added to the reviewing panel -- by changes in DWC membership and/or inclusion of additional consultants -- to ensure independent assessment of the Agency's work.

The materials provided to the SAB for review consisted of: a) a set of briefing charts titled *National Contaminant Occurrence Data Base Design Phase Considerations; Briefing to and Questions for the Science Advisory Board* dated June 18, 1998; b) the *National Drinking Water Contaminant Occurrence Data Base - Development Strategy* dated December 1997; and the c) *NCOD Attribute Type List* dated April 30, 1998.

1. Background

Section 1445(g) of the SDWA requires the EPA Administrator to “...assemble and maintain a national drinking water contaminant occurrence data base, using information on the occurrence of both regulated and unregulated contaminants in public water systems obtained under subsection (a)(1)(A) or subsection (a)(2) and reliable information from other public and private sources.” In addition the act states that “In establishing the occurrence data base, the Administrator shall solicit recommendations from the Science Advisory Board, the States, and other interested parties concerning the development and maintenance of a national drinking water contaminant occurrence data base, including such issues as the structure and design of the data base, data input parameters and requirements, and the use and interpretation of data.”

The Agency intends to use the NCOD to help it identify contaminants for future Contaminant Candidate Lists, to select contaminants for future regulation, to develop new national primary drinking water regulations for selected contaminants, to revise existing national primary drinking water regulations, and to provide information to the public in a readily accessible form. The Agency intends to build the NCOD on existing data sources (e.g., Safe Drinking Water Information System--SDWIS and Storage and Retrieval of U.S. Waterways Parametric Data--STORET) and to build in refinements later. NCOD data could potentially include information on regulated and unregulated contaminant occurrence, ambient monitoring data, and other data from research and special studies. Both historical and future data are to be included in the data base.

Since May 1997, the Agency has worked to develop an NCOD strategy and has interacted with stakeholders and other groups on technical issues associated with SDWIS, STORET, microbiological contaminants, data quality, sample test results, public health, environmental factors, public access, reporting standards, and database design.

The Agency plan for completing the development of the NCOD includes the following:

- a) Decision on data elements (spring 1998)
- b) Decision on the electronic platform (spring 1998, plus 1 month)
- c) Design and development (April 1998 to April 1999)
- d) Develop analytical plan to apply to listing, selection, and regulation
- e) NCOD operational test (April to July 1999)
- f) Guidance on data submission to NCOD (June 1999)
- g) Plan NCOD long-term maintenance and data analysis (June 1999)
- h) Public access (August 1999)

The charge to the Committee asked:

- a) Are the data elements included for Sample Test Results adequate for scientific analyses, recognizing that more detailed data will still be stored by the laboratory?
- b) What types of results should be reported for peer review by the scientific community relative to regulatory decisions? How should these results be reported?

2. General Comments

The Science Advisory Board (SAB) is pleased that the Agency is organizing drinking water data to facilitate its effective use. The principal recommendation of the SAB is that the Agency should consider and clearly articulate the intended uses of this data, and the methods that will be used for data analysis and presentation, before the NCOD design is completed. This will enable EPA scientists to more effectively identify those data elements that are essential for inclusion within the data base. The Committee also recommends that the EPA pay special attention to the collection and organization of high quality data in the future and not to invest heavily in previously collected data of less well-defined quality that was gathered before the NCOD was designed.

3. Specific Charge Questions

- 3.1 Are the data elements included for Sample Test Results adequate for scientific analyses, recognizing that more detailed data will still be stored by the laboratory?**

The Agency provided the DWC with a list of attributes for possible NCOD inclusion. Some of the attributes would be reported while others would be obtained from automated reference tables within the system. The list includes over 120 separate attributes: about 10 percent of these are labeled as Sample Test Results (STR). These include the following:

- a) Concentration measure
- b) Units of measure for concentration
- c) Dead counts
- d) Live counts
- e) Detection limit measure
- f) Detection limit type
- g) Detection limit unit of measure
- h) Lower 95% confidence measure
- l) Upper 95% confidence measure
- j) Percent recovery
- k) Percent recovery standard deviation
- l) Sign of the result
- m) Validity indicator

Other attribute categories provide information on the sampling location, water source, chemical identity and applicable drinking water standards, distribution system, laboratory conducting the analysis, nature of the sample collected, analytical techniques used, treatment techniques used on the water, and zip code.

The Agency asked whether the NCOD attributes labeled as “Sample Test Results” are adequate for the scientific analyses needed to identify contaminants for future Contaminant Candidate Lists (CCL), to select contaminants for future regulation, to develop new national primary drinking water regulations for selected contaminants, to revise existing national primary drinking water regulations, and to provide information to the public in a readily accessible form.

To support these uses, the Agency stated that the data base should be designed to answer the following questions:

- a) What is the contaminant?
- b) At what concentration is the contaminant found?
- c) Where and when is the contaminant found?

- d) What is the type of water source?
- e) Is water treatment associated with the occurrence?
- f) At what concentration is the contaminant a health concern?
- g) What number of people are exposed?
- h) Is there co-occurrence with other contaminants?
- i) Why was the sample collected?
- j) What is the level of confidence in the measure of concentration?

These are important questions to ask, however, the situation under which the STR data will be used is quite complex. Until the Committee has a clear understanding of how the data will be applied to answer these questions in support of the regulatory purposes noted above, it is not possible to fully comment on whether these attributes are, or are not, adequate. Although some reaction is possible as a result of observing elements in the list provided (see Appendix A for some examples), the SAB does not feel that it is now useful to dwell on these issues because its comments would only reflect a fragmented picture of the uses intended for the database. Instead, the SAB recommends that the Agency explicitly examine the intended uses of the contaminant occurrence data. Doing so will lead the agency to a systematic approach to define the specific data elements that need to be included in the NCOD.

The SAB understands that the Agency is currently developing a plan for the analysis of data from the NCOD that will be broadly applied during contaminant listing, selection, and regulation. This plan is now scheduled for Agency review later in 1998. This analysis plan will lay out how the information from the NCOD is to be reported and how it will be accessed by the public. Overlapping the development of this plan, the Agency is settling on NCOD design and development issues and will conduct an operational test of the system during the April to July 1999 time frame. The SAB recommends that the Agency move up its time-table for the development of the analysis plan.

As used in this report, this analysis plan would describe the use of data from the NCOD and it should include at least:

- a) a clear and formal statement of the purposes to which the data will be put.
- b) a formal statement of what the objectives of the data collection are to be relative to its representativeness (i.e. representative of a single water supply, representative of the nation as a whole, determining whether contaminants are derived from the source

water or introduced in treatment and distribution, etc.). This will translate directly into a sampling plan and decisions about what data can or should be included in the database.

- c) expectations of precision and accuracy that will be needed to meet the stated objectives of the data collection activity.
- d) Sample test cases should be used by the Agency to insure that all the data attributes required for the specified uses have been identified.

Sample test cases should address the Agency's array of goals (e.g., regulatory development, exposure assessment, etc.), which are among the most important questions to the Agency and its stakeholders. These test cases should also provide a framework for developing quantitative statistical and geographic procedures and facilitate the definition of specific input parameters and sample and contaminant information needed to support scientifically defensible statistical and geographic analyses. Sample test cases would also help to identify a set of relevant data quality objectives pertaining to the input parameters and contaminant measurement values used in the statistical algorithms and geographic procedures. For example, some of the important categorical factors uncovered by the test cases might be related to: a) treatment processes, b) sample characteristics, and c) methods used for measuring the contaminants and how missing information would be handled in the analysis among other things. Each of these factors would have specific attributes identified in the sample test cases or mock exercises.

An extension of the example may be illustrative of the utility of sample test cases. If one wanted to evaluate the effect of a treatment process on a contaminant, it would be important to capture changes in process from one sampling episode to the next. At least two additional attributes would be needed for the analysis, the location of the sampling points (i.e., source water and treated water) and the detection limits for the analyses. Indicators of precision and bias (in the measurement values, e.g., how non-detects were handled) would be important data elements for each contaminant measurement reported and included in the database. These factors, and others like them, would have to be included in the data base to make a sensible analysis. The sample test cases or mock exercises should make it clear whether one or more important sample attributes that would be critical to the desired analyses have been inadvertently omitted.

Finally, the development of an analysis plan should involve consultation on major issues with experts such as engineers for treatment processes, with analytical chemists for sampling and contaminant analysis, and with microbiologists for sampling and analysis of microbes.

The Committee expects several positive outcomes from the analysis plan that will ensure that the NCOD provides data for regulatory analyses that meet the highest scientific standards. It is the Committee's opinion that the NCOD will produce such benefits for the agency and the regulated community if it is properly developed. Specifically, the Committee would like to point out some obvious benefits relating to data quality.

- a) Establishing a database that has defined standards for data quality and completeness will have major benefits. The SAB recommends that the Agency bias its effort toward influencing and collecting good quality contemporary data first and only invest in the inclusion of older data as a secondary priority. Furthermore, such attributes will allow casually submitted data that may be of poorer quality to be segregated from good quality data that will be needed for certain types of analyses.
- b) Data taken at "standard stations" like water treatment plant intakes, water treatment plant outlets, water wells, and designated sampling points in the distribution piping of drinking water systems and at designated ambient sampling points used by the United States Geologic Survey (USGS) will prove most useful and should have first priority. One-of-a-kind sampling programs or sampling programs that do not have fixed sampling points should receive secondary attention. The Agency should take the opportunity provided by the NCOD to apply existing and emerging technologies for presenting data and the Agency's analyses of the data to the public (e.g., geographic information systems).
- c) Sample data submitted by states, with analyses conducted by certified drinking water laboratories using standard or draft standard methods, will prove most reliable and should have priority over sample data submitted from one-of-a-kind surveys and/or analyses conducted by laboratories that are not certified.

- d) Sample compositing requires special attention because it is only appropriate for contaminants whose effects are associated with total dose consumed over extended periods of time. It is not appropriate for sampling which measures microbial contaminants, chemicals with primary effects on development, or chemicals that may lead to acute effects, such as, nausea, vomiting, or diarrhea. Where appropriate (e.g., carcinogenic chemicals), compositing can be done in particular places or at specific times or over different magnitudes of space or time. There are a variety of techniques that can be used for compositing. Though composited data are potentially powerful in certain circumstances, their interpretation and their comparison with other data on the same contaminants can be quite difficult.

- e) The Agency should consider how it will report data with many non-detects (NDs) determined by different methods and by different laboratories, each with their own detection limit. For example, one could indicate for non-detects, one or more of the following. The:
 - (1) number of samples analyzed
 - (2) range of values for chemical contaminants
 - (3) reporting of microbes (yes/no presence, too many to count)
 - (4) number of samples with quantifiable levels
 - (5) number of N.D.s < 1st MDL - 1st Method
 - (6) number of N.D.s , 2nd MDL - 2nd Method
 - (7) 50% value determined by Maximum Likelihood Methods
 - (8) 90% value determined by Maximum Likelihood Methods

In summary, the Committee recommends that the following steps will help it confirm that the most appropriate data elements are included in the NCOD. It should determine exactly how the data elements will be used in the regulatory process, exposure assessment, etc. by developing a detailed Analysis Plan as a critical step in database design; design report forms to address each of their needs and consider how the reports can be organized to make the results user friendly; and build the database requirements using this information with additional assistance from experts in the field.

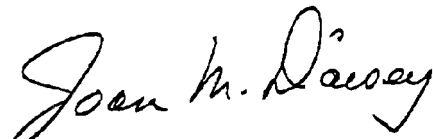
3.2 What types of results should be reported for peer review by the scientific community relative to regulatory decisions? How should these results be reported?

Again, there is no simple answer to this question; however, as indicated above the development of an analysis plan will be important in responding to this question. Once the intended use of the data are described and the possible results from using the data are identified, the need for peer review and the manner in which the results are to be reported can be addressed.

Peer review will undoubtedly occur in the context of a particular use of the data. Part of the review will be focused on data quality, but peer reviewers will also be interested in whether the data are sufficiently representative to accomplish their intended purpose. The definition of uses of the data should be explicit, not simply be couched in terms of “regulatory uses.” With that further level of specificity, data attributes would be identified with the uses of the data rather than being defined *a priori* and without reference to specific uses. Using this approach it is probable that many of the attributes listed may be found to be unimportant. More importantly, such an organized approach could minimize the number of important attributes that might be overlooked.

In conclusion, the SAB appreciates the opportunity to review and comment upon this needed data base. The Agency has made substantial progress in developing this tool that will be important to future drinking water regulations. The SAB is confident that once the Agency’s analysis plan showing how the data are to be used in supporting future regulatory analyses is completed that it will be possible to determine the final attributes that will need to be included in the NCOD. The SAB would be pleased to review that plan and to provide additional advice on the adequacy of the Sample Test Results included in the NCOD in supporting the stated regulatory needs.

Sincerely,



Dr. Joan Daisey, Chair
Science Advisory Board



Dr. Richard J. Bull, Chair
Drinking Water Committee
Science Advisory Board

APPENDIX A

Examples of Attributes Listed That Will Impact Data Use

To conduct exposure assessments, there is a need for representative data. The sample test results attributes (STR) included do not seem to allow for an assessment of representativeness. A data point does not represent a concentration for a specific contaminant year round for all populations. How the data are to be manipulated and used needs to be clear and this would affect the attributes and data elements needed.

For example, the listed attributes will only provide answers to the concentration (with upper and lower 95% confidence bounds), locations, and point in time. However, we would also want to know an exposure level representative of a longer term exposure concentration in other locations, or exposure levels that would be applicable to all states, if a federal regulation is to be developed that is to be applied to all states. Would the consideration of factors such as the frequency of sampling, sample size, and number of sampling location (and how they are distributed) needed for this purpose already be incorporated and expressed in final values in the attributes such that no more data manipulation is needed? The critical point is, how representative is the 'concentration measure' that would be reported? If other information is needed to make this 'concentration measure' representative, then it should be added to the attributes.

Another example is whether the Agency would like to identify a concentration at which the contaminant is a health concern" How is this concentration determined? The item 'applicable drinking water standards' is included as an attribute that might address this question, but MCLs are not solely health based. Many are based simply on quantitation limits. The basis for deriving the health concern concentration should be included in the data base if this question is of interest to the Agency.

The fields of the NCOD that pertain to the microbial contaminants seem to be significantly underdeveloped. The fields reflect the very narrow viewpoint of data to come from the Information Collection Rule (ICR). The SAB assumes that the NCOD will be used for purposes other than analyzing the ICR data, therefore the database needs to be developed within a broader context. For example, the

attribute "Sample Result-Percent Recovery" is to be reported for protozoan analyses only. It will be just as important to have information on the percent recovery for other microorganisms, such as viruses. In addition, there is no way to report results of qualitative analyses, such as those from PCR (polymerase chain reaction) or MPN (most probable number) analyses.

The SAB could develop some additional microbiological attributes for the NCOD, however, this would not be the most effective way to compile a complete set of attributes. The SAB recommends that the EPA convene a group of experts (internal, external, or both) to consider the issue of microbial attributes needed to support regulation once the Analytical Plan is completed.

APPENDIX B

ABBREVIATIONS

CCL	Candidate Contaminant List
DWC	Drinking Water Committee
ICR	Information Collection Rule
MCL	Maximum Contaminant Level
MDL	Minimum Detection Level
MPN	Most Probable Number
NCOD	National Contaminant Occurrence Database
ND	Non-Detects
PCR	Polymerase Chain Reaction
SAB	U.S. EPA Science Advisory Board
SDWA	Safe Drinking Water Act Amendments (1996)
SDWIS	Safe Drinking Water Information System
STORET	Storage and Retrieval of U.S. Waterways Parametric Data
STR	Sample Test Results
USGS	U.S. Geologic Survey

NOTICE

This report has been written as part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

**ENVIRONMENTAL PROTECTION AGENCY
SCIENCE ADVISORY BOARD
DRINKING WATER COMMITTEE**

ROSTER

CHAIR

DR. RICHARD BULL, Batelle Pacific Northwest Laboratories, Molecular Biosciences, Richland, WA

MEMBERS/CONSULTANTS

Dr. JUDY A. BEAN, Department of Epidemiology and Public Health, University of Miami, School of Medicine, Miami, Florida 33136

DR. LENORE S. CLESCERI, Rensselaer Polytechnic Institute, Materials Research Center, Troy, New York

DR. MARY DAVIS (consultant), Professor of Pharmacology & Toxicology, Department of Pharmacology & Toxicology, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, WV

DR. YVONNE DRAGAN, McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin

DR. JOHN EVANS, Harvard Center for Risk Analysis, Boston, Massachusetts

DR. ANNA FAN-CHEUK, California Environmental Protection Agency, Berkeley, California

DR. CHRISTINE MOE (consultant), Department of Epidemiology, University of North Carolina, Chapel Hill, NC

DR. LEE D. (L.D.) MCMULLEN, Des Moines Water Works, Des Moines, IA

DR. CHARLES O'MELIA, Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, MD

DR. EDO D. PELLIZZARI, Research Triangle Institute, Research Triangle Park,
NC

DR. VERNE A. RAY, Medical Research Laboratory, Pfizer Inc., Groton,
Connecticut

DR. GARY A. TORANZOS, Department of Biology, University of Puerto Rico,
San Juan, Puerto Rico

DR. RHODES TRUSSELL, Montgomery Watson, Pasadena, CA

DR. MARYLYNN V. YATES, Department of Soil and Environmental Sciences,
University of California, Riverside, CA

SCIENCE ADVISORY BOARD STAFF

MR. TOM MILLER, Designated Federal Official, US EPA/Science Advisory
Board, 401 M Street, S.W. (1400), Washington, D.C. 20460

MS. MARY WINSTON, Staff Secretary, Science Advisory Board, 401 M Street,
S.W. (1400), Washington, DC 20460

DISTRIBUTION LIST

Administrator
Deputy Administrator
Assistant Administrators
Deputy Assistant Administrator for Science, ORD
Director, Office of Science Policy, ORD
EPA Regional Administrators
EPA Laboratory Directors
EPA Headquarters Library
EPA Regional Libraries
EPA Laboratory Libraries
Library of Congress
National Technical Information Service
Congressional Research Service

G:\USER\SAB\REPORTS\98REPORT\DWCADV.004