NDWAC
CCL Classification Process Work Group
Work Plan

Discussion Draft with Revisions from December 16-17, 2002 Meeting

Objectives

1. Provide guidelines and recommendations for developing a database that encompasses contaminants that may have the potential to occur in drinking water. This database shall serve as the "Universe" from which potential contaminants will be selected in the CCL.
2. Provide recommendations on criteria and a process to screen contaminants from the Universe Database to the PCCL.
3. Develop an approach to score contaminant attributes that reflect their potential for drinking water occurrence and/or adverse health effects. Evaluate scoring process.
4. Identify, evaluate, test and make recommendations on a process to select contaminants from the PCCL for the CCL.
5. Conduct pilot studies and make recommendations for research plan for VFAR.
6. Prepare a common vocabulary and glossary of terms for data, criteria, attributes, and characteristics of databases, protocols and classification systems to allow meaningful and understandable conversations among work group members.

Tasks

Task 1 - Review available data sources ~~databases~~
Activity Group Assignment – ~~Universe~~ Data

1A. Microbial data sources

The work group will review databases recommended by the NRC and other existing sources of information on microorganisms. The NRC recommended many ~~databases as~~ sources of information for the Universe. The data sources vary relative to the nature and extent of the data included in those sources, ranging from groups of potential contaminants (e.g., pesticides) with occurrence and/or health effects data to simple lists with no additional data. ~~Both chemical and microbial databases will be reviewed.~~ The work group will familiarize themselves with these ~~data~~ sources to gain an understanding of the nature and extent of available data.

The work group will evaluate data availability relative to the ability to appropriately apply screening and classification processes to the Universe Database and the PCCL. The work group will consider the purpose of previous data collection efforts to better understand the relevance and applicability of the data to a screening process, particularly with respect to matching data to screening criteria.

The work group will perform an analysis to identify methods to fill data gaps for potential contaminants as well as data unavailable, but desirable, for future CCL preparation. Occurrence and health effects data can be expressed in a number of ways and the strengths and weaknesses of these methods vary. When data are lacking for a specific chemical, some estimate of

occurrence or health effects information may be obtained from other data (e.g., QSARs for chemical/physical properties or topicologic parameters from structures or occurrence data from release or production data). The work group will identify and evaluate methods to estimate data. These methods can help to populate the Universe to increase its usefulness in identifying emerging contaminants and candidates for the PCCL and possibly CCL.

Information Required

1. Alphabetical list of data sources, with both a short and detailed text description.
2. Tabular summaries of data sources, including information on data elements (e.g., type of occurrence information, type of health effects endpoints).
3. A compilation of ~~the~~ available databases and their data elements, in a format that can be queried.
4. Evaluation of occurrence data sources and data elements.
5. Discussion paper on data sources for the Universe framing issues and questions surrounding the construction of the Universe.
6. Input on Universe to PCCL screening from Task 4.
7. Input on attributes from Tasks 7 and 8.

Deliverables and Schedule

1. Characterization of available data sources ~~bases~~ – February 2003.
2. Evaluation of extent of data – March 2003.
3. Gap analysis – March 2003.
4. Identify and evaluate methods for estimating parameters – March 2003.
5. Ideas on process for addressing emerging potential contaminants – March 2003.
6. Draft data elements needed for populating the universe – February 2003.
~~5.~~7. Recommendation of appropriate data for the Universe Database – May 2003.

1B. Chemical data sources

The work group will review databases recommended by the NRC and other existing sources of information on chemicals. The NRC recommended many databases as sources of information for the Universe. The data sources vary relative to the nature and extent of the data included in those sources, ranging from groups of potential contaminants (e.g., pesticides) with occurrence and/or health effects data to simple lists with no additional data. ~~Both chemical and microbial databases will be reviewed.~~ The work group will familiarize themselves with these ~~data~~ sources to gain an understanding of the nature and extent of available data.

The work group will evaluate data availability relative to the ability to appropriately apply screening and classification processes to the Universe Database and the PCCL. The work group will consider the purpose of previous data collection efforts to better understand the relevance and applicability of the data to a screening process, particularly with respect to matching data to screening criteria.

The work group will perform an analysis to identify methods to fill data gaps for potential contaminants as well as data unavailable, but desirable, for future CCL preparation. Occurrence and health effects data can be expressed in a number of ways and the strengths and weaknesses of these ways vary. When data are lacking for a specific chemical, some estimate of occurrence or health effects information may be obtained from other data (e.g., QSARs for chemical/physical properties or topicologic parameters from structures or occurrence data from release or production data). The work group will identify and evaluate methods to estimate data. These methods can help to populate the Universe to increase its usefulness in identifying emerging contaminants and candidates for the PCCL and possibly CCL.

Information Required

1. Alphabetical list of data sources, with both a short and detailed text description.
2. Tabular summaries of data sources, including information on data elements (e.g., type of occurrence information, type of health effects endpoints).
3. A compilation of ~~the~~ available databases and their data elements, in a format that can be queried.
4. Evaluation of occurrence data sources and data elements.
5. Discussion paper on data sources for the Universe framing issues and questions surrounding the construction of the Universe.
6. Input on Universe to PCCL screening from Task 4.
7. Input on attributes from Tasks 7 and 8.

Deliverables and Schedule

1. Characterization of available data sources ~~bases~~ – February 2003.
2. Evaluation of extent of data – March 2003.
3. Gap analysis – March 2003.
4. Identify and evaluate methods for estimating parameters – March 2003.
5. Ideas on process for addressing emerging potential contaminants – March 2003.
6. Draft data elements needed for populating the universe – February 2003
7. Recommendation of appropriate data for the Universe Database – May 2003.

Task 2 - Evaluate the quality of available data and data sources to construct the Universe Database

Activity Group Assignment – Data~~Universe~~

2A. Microbial data

The work group will evaluate data quality relative to the ability to appropriately and confidently populate the Universe Database with reliable data. The work group will consider the methods and protocols for collecting the data so that a relative understanding of accuracy, reproducibility, and overall reliability can be determined. The quality of data will be ~~compared to~~determined by screening criteria to better understand the usefulness of the data in subsequent protocols (e.g., screening from the Universe Database to the PCCL).

The work group will provide recommendations as to the appropriateness of including the available data in the Universe Database.

Information Required

1. Database review from Task 1.
2. Input on Universe to PCCL screening from Task 5.
3. Input on alternative prototype classification approaches, attributes, and scoring from Task 8.

Deliverables and Schedule

1. Evaluation of data quality – March 2003.
2. Recommendation and rationale of what quality of data is necessary for the Universe – May 2003
3. Identification of the best available data sources for each element – May 2003
4. Summary of appropriateness of data for subsequent analysis – May 2003.

2B. Chemical data

The work group will evaluate data quality relative to the ability to appropriately and confidently populate the Universe Database with reliable data. The work group will consider the methods and protocols for collecting the data so that a relative understanding of accuracy, reproducibility, and overall reliability can be determined. The quality of data will be ~~compared to~~determined by screening criteria to better understand the usefulness of the data in subsequent protocols (e.g., screening from the Universe Database to the PCCL).

The work group will provide recommendations as to the appropriateness of including the available data in the Universe Database.

Information Required

1. Database review from Task 1.
2. Input on Universe to PCCL screening from Task 5.
3. Input on alternative prototype classification approaches, attributes, and scoring from Task 8.

Deliverables and Schedule

1. Evaluation of data quality – March 2003.
2. Recommendation and rationale of what quality of data is necessary for the Universe – May 2003
3. Identification of the best available data sources for each data element – May 2003.
4. Summary of appropriateness of data for subsequent analysis – May 2003.

Task 3 - Recommend the composition and structure of the Universe Database

Activity Group Assignment – ~~Universe~~Data

3A. Microorganisms

Based upon the efforts conducted in Tasks 1, 2, 5, and 8, the work group will recommend the composition and structure of the Universe Database. The Universe Database should be in a format that allows it to be updated with new information, readily identifies data gaps and needs, and facilitates the process of screening to arrive at a PCCL and classification to a CCL. The types of elements that can be considered for occurrence include spatial scale (national vs. local), water type, extent of detection, population served, production/release/use data, release medium, etc. The types of elements that should be considered for health effects include type of study, health effect outcome, contaminant level producing the outcome, potential population affected (e.g., sensitive subpopulations).

Information Required

1. Data review and gap analysis from Task 1.
2. Input on Universe to PCCL screening from Task 5.
3. Input on attributes from Task 8.
4. Example database structure.

Deliverables and Schedule

1. Recommended composition and structure for the Universe Database – May 2003.
2. Recommend process for addressing potential emerging pathogens – May 2003.
3. Recommend process for updating Universe for microorganisms – May 2003.

3B. Chemicals

Based upon the efforts conducted in Tasks 1, 2, 5, and 8, the work group will recommend the composition and structure of the Universe Database. The Universe Database should be in a format that allows it to be updated with new information, readily identifies data gaps and needs, and facilitates the process of screening to arrive at a PCCL and classification to a CCL. The types of elements that can be considered for occurrence include spatial scale (national vs. local), water type, extent of detection, population served, production/release/use data, release medium, etc. The types of elements that should be considered for health effects include type of study, health effect outcome, contaminant level producing the outcome, potential population affected (e.g., sensitive subpopulations).

Information Required

1. Data review and gap analysis from Task 1.
2. Input on Universe to PCCL screening from Task 5.

3. Input on attributes from Task 8.
4. Example database structure.

Deliverables and Schedule

1. Recommended composition and structure for the Universe Database – May 2003.
2. Recommend process for addressing potential emerging chemical contaminants – May 2003.
3. Recommend process for updating Universe for chemicals – May 2003.

Task 4 - Evaluate processes for screening contaminants from the Universe to the PCCL

Activity Group Assignment – Methods~~Screening~~

Concurrent with Tasks 1, 2, 8 and 9, the work group will evaluate a process to screen contaminants from the Universe to a Preliminary Contaminant Candidate List (PCCL). The PCCL will serve as the basis to apply the recommended prototype classification process. The Subgroup will identify the information that will be desirable to allow for screening from the Universe to the PCCL. The NRC recommended a Venn Diagram approach using screening criteria for:

- Demonstrated occurrence in drinking water (e.g., measurements in tap water, distribution systems, finished water in treatment plants, source waters).
- Demonstrated adverse health effects (e.g., epidemiological studies, toxicological laboratory animal studies).
- Potential occurrence in drinking water (e.g., observations in watershed and aquifers, historical contaminant release data, chemical production data).
- Potential adverse health effects (e.g., other sources of health effects information such as predictive biological activity or effects models).

Information Required

1. Summary of the NRC conceptual approach for identifying contaminants for inclusion on a PCCL.
2. Issue paper on screening options.

Deliverables and Schedule

1. Proposed alternatives for screening from the Universe to the PCCL – March 2003.

Task 5 - Evaluate and recommend criteria and data elements for Universe to PCCL screening process

Activity Group Assignment – Data ~~Screening~~

In order to adequately evaluate the efficacy of various screening approaches, the work group will develop screening criteria for the NRC's Venn Diagram approach or any other approach that the work group proposes. These screening criteria will be of a nature that they can be applied to the data in the Universe Database to develop preliminary contaminant candidates for the PCCL.

The work group will identify data elements that must accompany screening criteria in order for those criteria to be used effectively. The extent and quality of these data elements will be evaluated in Tasks 1 and 2.

Information Required

1. Summary of the types of information that contribute to criteria for demonstrated occurrence/adverse health effects and potential occurrence/health effects.
2. A summary of potential data elements for the contaminant candidates on the PCCL.

Deliverables and Schedule

1. Criteria for screening from the Universe to the PCCL – March 2003.
2. Data elements for screening criteria – March 2003.

Task 6 – Recommend a process to screen the Universe to the PCCL

Activity Group Assignment – Methods (plenary) ~~Screening~~

The work group will evaluate alternative PCCLs based upon the screening process defined in Task 4 and the Universe Database constructed in Task 3. The work group will test the screening process and review example outputs for the proposed approach(es). Based upon this review, the work group will recommend an approach, screening criteria to be used with that approach, and any weighting or scoring to be applied to the screening criteria.

Based upon a review of these analyses, the work group will recommend a process to screen the Universe Database to develop a PCCL. The work group will review and comment on EPA's draft PCCL using the recommended process.

Information Required

1. Universe Database from Task 3.
2. Screening criteria from Task 5.
3. Sensitivity analyses of the impact of various screening protocols on the composition of the resulting PCCL.

Deliverables and Schedule

1. Review of sensitivity analysis of alternative methods – May 2003.
2. Recommended process for screening from the Universe to the PCCL – July 2003.
3. Review of EPA draft PCCL using the recommended process – September 2003.

Task 7 – Identify decision methods and prototype approaches

Activity Group Assignment – Methods~~Classification~~

The work group will evaluate applicable decision methods (e.g., rule-based, expert judgment, prototype algorithm) and will recommend an overall decision approach for developing the CCL from a PCCL. The NRC recommended the prototype classification approach for this purpose and the work group will evaluate alternative decision ~~prototype~~ approaches (e.g., neural network). During this evaluation, the work group will discuss other decision methods so that the ultimate recommendation is transparent and reproducible.

Characteristics (e.g. prediction, interpretation) that are desirable for a decision approach will be identified and weighed, including the ability of a given method to address data gaps. These characteristics will be compared to the characteristics of each alternative decision method.

Based upon this information, the work group will recommend an overall decision method and associated prototype approach(es) for further evaluation.

Information Required

1. Summary of possible decision methods.
2. Summary of characteristics for identified decision methods and comparison.
3. Summary of model information, requirements and software for recommended decision method and prototype approach(es).

Deliverables and Schedule

1. Desired characteristics of a decision approach, available decision methods, approaches, advantages/ disadvantages of various approaches – February 2003.
2. Recommended decision method and associated prototype approach(es) – March 2003.
3. Summary of prototype models selected – May 2003.

Task 8 – Identify attributes for recommended decision method and prototype approach

Activity Group Assignment – Data ~~Classification (with support from Screening)~~

The NRC recommended magnitude, prevalence and persistence/mobility as attributes for occurrence; severity and potency as attributes for health effects. The work group will identify and evaluate attributes to be used with the recommended decision-making approach. The work group will identify data elements associated with the attributes. The work group will ensure that

the data elements required for the recommended approach can be drawn from, or estimated through, the data elements associated with the candidate contaminants in the PCCL.

Information Required

1. Summary of NRC recommendations for attributes.
2. Summary of potential data elements for attributes.

Deliverables and Schedule

1. Summary of attributes and data elements for the recommended decision method and prototype approach(es) – May 2003.

Task 9 – Develop attribute scoring approach

Activity Group Assignment – Data ~~Classification (with support from Screening)~~

The work group will recommend a strawperson approach for scoring attributes as a part of the decision making approach. The work group will summarize the relationship between the approach, the required data elements, and the attributes.

Information Required

1. Output from Tasks 1, 2, 3 and 8.

Deliverables and Schedule

1. Discussion draft of attribute scoring approach and example scoring – May 2003.
2. Proposed initial attribute scoring – July 2003.

Task 10 – Prepare training dataset(s)

Activity Group Assignment – Plenary (Data)

The work group will evaluate training dataset(s) for the recommended prototype approach(es). The training datasets will be sufficiently robust to reveal variation in prototype model behavior (e.g. prediction, interpretation) based upon the attribute scoring and the composition of the training datasets. Training dataset(s) will include a) compounds not expected to be contaminants of concern, and b) compounds expected to be contaminants of concern. This dataset(s) will be used for evaluating the impact of scoring on prototype output and will be used for training of the prototype model.

The work group will identify additional technical experts to assist in preparing the datasets, if required.

Information Required

1. Output from Task 1, 2, and 8.

Deliverables and Schedule

1. Training dataset(s) – May 2003.

Task 11 – Evaluate scoring approach
Activity Group Assignment – Plenary (Data) ~~Classification~~

The work group will use the strawperson scoring approach from Task 9 and a training set from Task 10 to evaluate the impact of the attribute scoring on prototype model output. The attributes in the approach and/or scoring may be varied to determine the sensitivity of the output. The basis for scoring attributes will be reviewed using expert judgment. The required data elements may be revisited, depending upon the output. The work group will consider the overall evaluation and recommend a final scoring approach for the prototype model.

Information Required

1. Output from Tasks 9 and 10.

Deliverables and Schedule

1. Attribute scoring – September 2003.

Task 12 – Build, test and perform sensitivity analysis on decision ~~prototype~~ approach(es)
Activity Group Assignment – Methods ~~Classification~~

The work group will evaluate a sensitivity analysis performed with the recommended classification approach(es). The work group will review and evaluate output from the prototype model using the training dataset. Model output will be compared to expectations to evaluate the efficacy of the approach.

It may be necessary to obtain additional expert judgment to evaluate the output from the sensitivity analysis. The work group will identify additional technical experts to assist in evaluating model output, if required.

Information Required

1. Output from Tasks 9, 10 and 11.

Deliverables and Schedule

1. Prototype model – May 2003.

2. Summary of sensitivity analysis and model capability – September 2003.

Task 13 – Recommend decision method and prototype approach
Activity Group Assignment – Plenary

Based upon the work performed in Tasks 7 through 12, the work group will recommend a decision method and prototype approach for developing a CCL from a PCCL. The recommendation will be based upon the sensitivity analysis in Task 12 and the work group's expert professional judgment. The recommended decision method and prototype approach will be carried forward for developing the next CCL.

Information Required

1. Sensitivity analysis in Task 12.

Deliverables and Schedule

1. Summary of recommended decision-making method, advantages and disadvantages, data requirements, and limitations – November 2003.
2. Final training set for the approach – November 2003.

Task 14 – Evaluate Virulence Factor Activity Relationships (VFARs)
Activity Group Assignment – VFAR

The work group will evaluate the efficacy of using VFARs to estimate the potential of a microorganism to exhibit similar characteristics to microorganisms with known health impact. A series of pilot projects will be identified to evaluate the VFAR concept and how the concept can be included in a CCL classification process. For example, keyword and genomic sequence searches using available databases will initially be evaluated to determine the extent and quality of data available to match sequences of genes and proteins.

The work group will conceptualize an approach to use VFAR as a part of the CCL process, summarize the advantages/disadvantages of the VFAR approach and recommend a timeline for potential implementation.

Information required

1. Keyword/sequence searches of known pathogens using available databases.
2. Results of keyword searches and nucleotide sequences identified.
3. Summary of potential pilot projects.

Deliverables and Schedule

1. Clarification of databases available to use and the benefits and limitations of each
2. Summary evaluation of the feasibility of the VFAR approach - December 2002.

3. Outline pilot projects and evaluations to test the viability of the VFAR approach – February 2003.
4. Time frame for potential VFAR implementation that identifies VFAR components and milestones to be accomplished - February 2003.

Task 15 – Develop a Common Vocabulary
Activity Group Assignment – All

The work group will develop a common vocabulary and glossary of terms to ensure that communication between various subgroups is well understood. This will be particularly important when discussing screening processes and prototype classification systems.

Information Required

1. Preparation materials provided by EPA and their consultants.

Deliverables and Schedule

1. Glossary of terms. First draft for December 16, 2002 meeting.

Task 16 – Risk communication and public involvement (new language drafted by RESOLVE)
Activity Group Assignment – All and plenary

The work group will continually look for ways to communicate with stakeholders and the public in an understandable and meaningful way about work group deliberations and products. For example, each activity group (AG) will have a "transparency leader" to help ensure that the group addresses the need for transparency as it works on its other tasks. The transparency leaders are Dan Wartenberg for the Methods AG and Benson Kirkman for the Data AG. At each plenary meeting, each AG will report on how it is addressing risk communication and public involvement. In addition, the work group will prepare recommendations for the NDWAC on risk communication and public involvement for the CCL process.

Deliverables and Schedule

1. Options for increasing transparency of the work group process – March 2003.
2. Recommendations for the NDWAC on risk communication and public involvement for the CCL process – July 2003.

Task 17 - Evaluate the quality of available data and data sources to classify potential contaminants from the PCCL to the CCL
Activity Group Assignment – Data

2A. Microbial data

The work group will evaluate data quality relative to the ability to appropriately and confidently classify potential microbial contaminants from the PCCL to the CCL with reliable data. The work group will consider the methods and protocols for collecting the data so that a relative understanding of accuracy, reproducibility, and overall reliability can be determined. The quality of data will be determined by classification attributes to better understand the usefulness of the data in classifying from the PCCL to the CCL.

The work group will provide recommendations as to the appropriateness of including the available data.

Information Required

Database review from Task 1.
Input on Universe to PCCL screening from Task 5.
Input on alternative prototype classification approaches, attributes, and scoring from Task 8.

Deliverables and Schedule

1. Evaluation of data quality – March 2003.
2. Recommendation and rationale of what quality of data is necessary for classifying from the PCCL to the CCL – May 2003.
3. Identification of the best available data sources for each data element – May 2003.
4. Summary of appropriateness of data for subsequent analysis – May 2003.

2B. Chemical data

The work group will evaluate data quality relative to the ability to appropriately and confidently classify potential chemical contaminants from the PCCL to the CCL with reliable data. The work group will consider the methods and protocols for collecting the data so that a relative understanding of accuracy, reproducibility, and overall reliability can be determined. The quality of data will be determined by classification attributes to better understand the usefulness of the data in classifying from the PCCL to the CCL.

The work group will provide recommendations as to the appropriateness of including the available data.

Information Required

1. Database review from Task 1.
2. Input on Universe to PCCL screening from Task 5.

3. Input on alternative prototype classification approaches, attributes, and scoring from Task 8.

Deliverables and Schedule

1. Evaluation of data quality – March 2003.
2. Recommendation and rationale of what quality of data is necessary for classifying from the PCCL to the CCL – May 2003
3. Identification of the best available data sources for each data element – May 2003.
4. Summary of appropriateness of data for subsequent analysis – May 2003.