Next Steps – Continuing Discussion to Describe Guidelines For Building the CCL Universe
Revision 2 – From April 14, May 1, 2003 Data AG Meetings

At the March 27-28, 2003 NDWAC CCL CP Work Group meeting, the Work Group adopted the conceptual framework for building the CCL Universe (see the March 27, 2003, *Building The CCL Universe*, discussion paper). The framework outlines a principles-based approach; a process for defining the CCL Universe on the basis of a set of fundamental premises regarding the nature of the contaminants that should be considered. This recognizes the NRC's (2001) fundamental recommendations for the contaminants that should be included in building the CCL Universe. The two basic principles that form the foundation of the CCL Universe are:

- The CCL Universe should include those contaminants that have demonstrated or potential occurrence in drinking water; and
- The CCL Universe should include those contaminants that have demonstrated or potential adverse health effects.

The *Building The CCL Universe* discussion paper also laid out the following definitions pertinent to the CCL Universe discussion:

- contaminants: any physical, chemical, or biological agent that does, or may, occur in water.
- known contaminants: physical, chemical, or biological agents that have been identified in the technical literature and adequately characterized to enable a judgment regarding their inclusion in the CCL Universe.
- emerging contaminants: a subset of known physical, chemical, or biological agents previously evaluated as not requiring inclusion in the CCL, for which new information becomes available which heightens concern and triggers re-evaluation.
- new contaminants: physical, chemical, or biological agents that are or may be newly-discovered or synthesized, for which little is known about their potential occurrence or adverse health effects.

The purpose of this discussion paper is to expand on the *Building the CCL Universe* consensus:
- to establish further guidelines on how to apply the principles;
- to outline specific examples of the application and implications of the principles-based approach; and
- to enable the next stages of discussion and analysis of data sources to progress.

Background

Building the CCL Universe would involve a multi-step process, including:
1. constructing a database of known contaminants;
2. a surveillance process for new and emerging contaminants;

3. a nomination process for new and emerging contaminants; and
4. an expedited process (to enable EPA to move a contaminant to a higher priority of assessment, as situations warrant).

This paper primarily focuses on discussion issues related to constructing the database of known contaminants, to help clarify application of the principles for selection of data sources. For "Constructing the CCL Universe" the *Building the CCL Universe* concept paper notes:

> Data sources would be identified that provide relevant information about known contaminants that may be potential drinking water contaminants. Data from these sources would be merged or recombined, using the data source compilation approach, from discrete databases to compile novel sets of records with multiple criteria. An example would be merging databases for High Production Volume Chemicals and the Registry of Toxic Effects of Chemical Substances (RTECS) to identify and combine relevant information for a specific contaminant of interest. The physical process of the known contaminants approach would involve exporting output from various databases into a new database.

The following data source discussion primarily pertains to constructing a database of known contaminants but some of the principles apply to other Universe issues. For example, in building the CCL Universe bibliographic data sources would primarily be used in the Surveillance and Nomination Process for New and Emerging Contaminants. However, because of the lack of database-type sources for microbiological contaminants, it may be necessary to use bibliographic sources and primary literature to compile data and information for microbes for the immediate future CCL needs.

Discussion – Clarifying Issues For Selecting Data Sources for the CCL Universe

Various issues are identified below that may further clarify the intent and implications of the Data Activity Group (AG). The discussion attempts to frame these in example statements, followed by discussions of the implications, and example applications. Discussion and consensus on these issues will further clarify data source selection guidelines. The discussion assumes that the data/data sources will also meet guidelines for data quality that are yet to be clarified.

Inherent in the multi-step, data source compilation approach, and the discussions of the Data AG, is that there would be a hierarchy in which data sources would be merged into the CCL Universe. As part of this process, some data sources would provide both names of contaminants and pertinent data elements (to be defined), associated with the contaminants, to the CCL Universe database. These would be the data elements needed for later screening and processing of the CCL Universe to the PCCL, and possibly to a CCL. Other data sources that do not contain data elements relevant to the CCL principles and process may only provide names to the CCL Universe. In addition, other sources which are not directly relevant to the principles of occurrence and/or health effects may still provide data/information to fill gaps in the data

elements. In this regard, the Data AG recognized that EPA will need to use its expert judgment to assess relevant information. (See the discussion and examples for stages 2 and 3, below.)

The multi-step process is summarized as four stages below. Sources that met any one of the four "ovals" of NRC's Venn diagram would be included. Each stage involves iterative steps to include multiple sources of information. The four stages incorporate the guidelines discussed by the Data AG. Also, the CCL Universe process is not complete until all four stages are completed. Furthermore, the Surveillance and Nomination processes may still add to the CCL Universe beyond these stages.

1. Compile the most relevant data sources. Following the principles for Building the CCL Universe, data sources that have either relevant occurrence and/or relevant health effects data would be compiled into the CCL Universe. While SDWA puts more weight on occurrence information, the AG recommends that equal weight should be given to occurrence and health effects information.

   These most relevant sources would be the first to be uploaded into the CCL Universe data base. These include sources directly relevant to evaluating the potential for human exposure via drinking water (distinguished from sources relevant to exposure via other media, or of less relevance to human health effects). They would form the first entries to the list of contaminants in the CCL Universe and they would begin to populate the relevant data elements in the CCL Universe database. Foe example, this first stage would include occurrence and health effects data sources, such as:

   - National Contaminant Occurrence Database (NCOD)
   - EPA's (developing) Unregulated Contaminant Monitoring Regulation database (which will become part of the NCOD)
   - USGS's National Water Quality Assessment program
   - High Production Volume (HPV) Chemical Lists
   - Toxics Release Inventory (TRI)
   - ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Priority Lists
   - High Production Volume Master Summary Table
   - EPA's Health Advisory Tables
   - EPA's Integrated Risk Information System
   - Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs)
   - California's Office of Environmental Health Hazard Assessment (CA OEHHA) Toxicity Criteria Database (cancer values only)
   - World Health Organization (WHO) Drinking Water Quality Guidelines
   - World Health Organization's Classification of Pesticides by Hazard (CPH) Database
   - The International Agency for Research on Cancer (IARC) lists of carcinogens
   - FDA's Generally Recognized as Safe (GRAS) Notices
   - TERA's International Toxicity Estimates for Risk

These sources may, in part, be redundant for adding contaminants to the CCL Universe listing. But some will have occurrence data for the contaminant, some will have production data, some health effects data, and so forth, adding unique data elements to the CCL Universe database. As they are compiled some will add new contaminants, as well. For example, in compiling 23 data sources as a test of building the universe, the sum of contaminants from all sources was 17,891, however these correspond to about 8,750 unique contaminants.

2. Compile less directly relevant data sources. In the second stage of compilation, data sources would be used that may not be directly relevant to drinking water occurrence or human health effects, but that have a reasonable link (pathway) to drinking water concerns. These sources may be used for additional lists of contaminants or for specific data elements. This could include sources such as:

    - National Sediment Inventory (NSI). The contaminants listed in this source might only contribute new "names" to the CCL Universe. Contaminants registered in the NSI would be compared with the list of contaminants comprising the CCL Universe. Any contaminants on the NSI that were not already in the CCL Universe would then be added to the listings in the CCL Universe. The NSI might not contain any relevant data elements to be included in the CCL Universe database. A similar approach might be used with air-deposition data sources.
    - OSHA work place hazard information could also be reviewed at this stage. As noted by the Data AG, at this stage, EPA would apply expert judgment to filter for relevant information. OSHA data sources might be used for adding new names to the listings in the CCL Universe and might contribute needed data elements, as well.
    - Some data sources with purely ecological endpoints (e.g., AQUIRE) may or may not be appropriate and will require further expert review to assess if relevant information is available.

3. Use supplementary data sources to fill data gaps remaining from stages 1 and 2. In a third stage, data sources would be used whose purpose or development may not be relevant to the occurrence and health effects principles, but that can provide needed data elements and fill data gaps. As an example:

    - The High Production Volume Chemical data sources provide lists of contaminants that would be compiled into the CCL Universe in stage 1. In stage 3, the Chemical Abstract Service (CAS) Registry (STN) database could be used as a supplementary source to fill information gaps for needed data elements, such as solubility (or other properties).

4. Use surrogates to fill important information gaps. Following the data source compilation of stages 1-3, data gaps will still remain. For some contaminants it may be necessary to use surrogate information, or to "model" or "estimate" "potential" occurrence or health effects end points. EPA's Office for Prevention, Pesticides, and Toxic Substances (OPPTS) routinely uses quantitative structure activity relationships (QSARs) (e.g., from models such as EPIWIN) to fill data gaps for new chemicals as part of the Pre-Manufacturing Notification (PMN) program (under the Toxic Substances Control Act, sec. 5). Examples might be:

    - Use estimated water solubility from OPPTS PMN QSAR modeling to fill data gaps for relevant chemicals.
    - Use estimate of microbial pathogenicity from the virulence factor activity relationship (VFAR) approach (if available) or other surrogate process.

Again, as noted, the multi-step, data source compilation approach for known contaminants for the CCL Universe is not finished until all four stages are completed, and this process is repeated over time. In addition, the Surveillance and Nomination processes may still add to the CCL Universe beyond these stages.

Next Steps – Data Quality Guidelines

The Data AG has also touched on data quality issues. Initial discussion suggested that the data quality bar should be low for acceptance to the Universe. The suggestion was simply that some description of the origin of the data must be available, including minimal information, such as:

- Contact name;
- Description of the data elements;
- How the data were obtained; and
- Meaningfulness and relevance of the data.

The Data AG also noted in discussion that original data sources should be used to the extent feasible. There are further details and clarification to be considered. EPA also has new information quality guidelines to consider. An issue paper on data quality considerations is in preparation for next step discussions with the Data AG.