

SYSTEM FOR CAPTURING, STORAGE, RETRIEVAL, AND SHARING OF TOXICOLOGICAL INFORMATION REQUIRED IN THE EVENT OF RELEASE OF CBRN MATERIALS IN THE ENVIRONMENT

Tonya Nichols, Ph.D. and Cynthia Sonich-Mullin, Ph.D.
National Homeland Security Research Center, Office of Research and Development
U.S. Environmental Protection Agency

Michael Taylor, Ph.D.; Michele Ritondo, Ph.D.; Laurie Singer, Ph.D.; James Rogers, Ph.D.;
Daniel Chappie; Jennifer Price; Emily Fleming; Anastasia Torok, Ph.D.; and Donald McGonigle;
Battelle
505 King Avenue
Columbus, OH 43201 United States

INTRODUCTION

At the 6th Chemical Biological Medical Treatment Symposium (CBMTS VI) held in Spiez, Switzerland in April 2006, we presented a description of a database being developed by the Threat and Consequence Assessment Division in the Environmental Protection Agency's National Homeland Security Research Center (1). As envisioned, this database will provide immediate access to information extracted from articles and reports pertaining to risk assessment retrieved from the world's scientific literature and could serve as a "tool" that can be used to address the threat of terrorist attacks as well as other catastrophic events. The purpose of this paper is to provide an update of the status of the development of the database and to indicate progress made toward incorporating it into a web-based application.

MATERIALS AND METHODS

Approach being utilized for the development of the database

Information (literature citations) pertaining to biological materials of interest (e.g. biological threat agents, biological materials that pose an especially significant threat to the health of exposed populations and/or the environment) is initially sought by performing key word searches using several publicly available databases such as *PubMed*. Retrieved citations are assessed by Subject Matter Experts (SMEs) and articles that appear to be pertinent to the assessment of risks posed by biological materials to human health and the environment are acquired. Following acquisition, SMEs perform in-depth assessments of each article, extract information and data, and enter these data into the Knowledge Base (KB) of the database. Electronic versions of the acquired articles are placed into the literature repository (LR) of the database. Along with the article, a completed document criteria form is also entered into the LR on which the SME has checked off answers to interrogatories that provide a description of the article and its contents. The overall process for the development of the database is presented in Figure 1. In the following paragraphs we present additional information on the KB and LR compartments in the database including a description of current status of the database contents.

Knowledge base (KB)

The KB includes nine major information nodes each of which are directly pertinent to assessing the risk posed by a biological material. As the SME assesses an article, he or she is looking for information applicable to one or more of the 9 nodes (*i.e.*, the nodes guide the SME in the quest for pertinent information). [Note that the goal in KB development is to develop a picture of the current “state of knowledge” for characteristics of each biological material necessary to conduct microbial risk assessment. The goal is not to catalogue all articles that have been published on the particular biological material.] Additionally, SMEs add notations in the KB information regarding source documents or research gaps and are also able to insert relevant additional commentary to further clarify extracted information.

The nodes define the taxonomy of the KB and serve also to compartmentalize the knowledge for quick reference and retrieval. These nodes are: “General Information”; “Disease Characteristics and Host Interaction”; “Disease Transmission”; “Dose-Response Information”; “Prophylaxis and Treatment”; “Sampling, Detection, and Risk Assessment Techniques”; “Agent Delivery, Transport, and Fate”; “Inactivation, Decontamination, and Remediation”; and “Personal Protection”. All content in the KB is linked with its source Portable Document Format (pdf) document in the LR. Figure 2 shows an actual screen shot of a segment of the KB that is in the Routes of Transmission section of the “Disease Transmission” node in the portion of the KB dedicated to *Francisella tularensis*.

It should also be noted that the primary emphasis in this initial version of the database was the identification of content to populate the nodes “Disease Transmission”; “Dose-Response Information”; “Agent Delivery, and Transport and Fate”; and “Sampling, Detection and Risk Assessment Techniques.” While information for other nodes was collected and entered into the KB as it was available, the other nodes may not yet contain equivalent content with respect to the available literature. Initially information was sought for two biological organisms - *Bacillus anthracis* and *Francisella tularensis*. And, at this point in time, review of the literature for the nodes of primary interest for these two organisms is complete. Of course, additional work will be required to continuously update the KB contents and to, as resources are available, fill in any existing information or research gaps.

In addition to information on *Bacillus anthracis* and *Francisella tularensis*, segments of the KB have been opened for information on H5N1 avian influenza virus, and *Yersinia pestis*. These data are currently being acquired in the manner as described above and relevant information is currently being entered into the appropriate section of the KB. Information on additional biological agents will be sought in the future.

Literature repository (LR)

The citation list of the LR contains electronic copies of all articles acquired to date (3,838 articles have been acquired however not all were used in the construction of the KB). Thus the total number of articles is somewhat higher than the total number of articles assessed. To date, 1,650 documents have been assessed for relevant information regarding *B. anthracis* and 577 for *F. tularensis*. In addition, a total of 1,060 documents have thus far been assessed for relevant

information on H5N1 avian influenza virus and 255 have been assessed for *Yersinia pestis*. As mentioned earlier, the literature review for these latter two agents is still in process.

Documents included in the LR were each characterized according to Document Type, Literature/Scientific Content Level, Experimental Methods, Statistical Analysis, Technology/Methods Used, Model Organism(s) Used, Controls, and Distribution. Drop-down menus were used for the characterization of each article in the LR. Figure 3 shows an example of the resulting information provided to users. Document category selections are displayed as unique identifiers (*i.e.*, colored dots) for the respective documents in the LR. Assignment of category selections to each article decreases the subjectivity of the document characterization process, while providing pertinent quality and relevance indications to the end user. The document characteristics provide a basis for various end users to interpret and implement data for a broad range of scenarios.

Internet access

The database described above (formulated using Oracle^R) has been coupled with a custom-designed worldwide web application and the database coupled with the web application has been termed “Support for Environmental Rapid Risk Assessment” or “SERRA”. By making the database web-accessible, of course, the access is greatly expanded. The features of SERRA, as well as the Oracle database are summarized in Figure 4. Currently, SERRA is undergoing beta-testing and will next subsequently undergo EPA internal review. SERRA is hosted on the CBRNIAC system with the eventual goal of making SERRA accessible through US EPA’s intranet resources. Internet accessibility is critical for maximizing portability and accessibility for use during an incident.

SUMMARY

SERRA and the associated methodologies to derive its content provide a dynamic approach for capturing and accessing critical information for the assessment of risk posed by CBRN releases. Currently, SERRA content reflects assessments of over 3,000 publications selected from over 10,000 citations. The KB currently contains information on four organisms: *Bacillus anthracis*, *Francisella tularensis*, H5N1 avian influenza virus, and *Yersinia pestis*. Literature review and assessment of pertinent articles have been completed for *B. anthracis* and *F. tularensis*. Per their policy, EPA will be conducting an internal peer review during Fall 2007 as the first step toward allowing external access to SERRA.

KEY WORDS

risk assessment, microbial agents, biological agents, CBRN, databases, literature search, data mining

FIGURES AND TABLES

Figure 1. Database workflow.

Figure 2. Screenshot of KB Information

Figure 3. Document Criteria Report for LR Content

Figure 4. SERRA

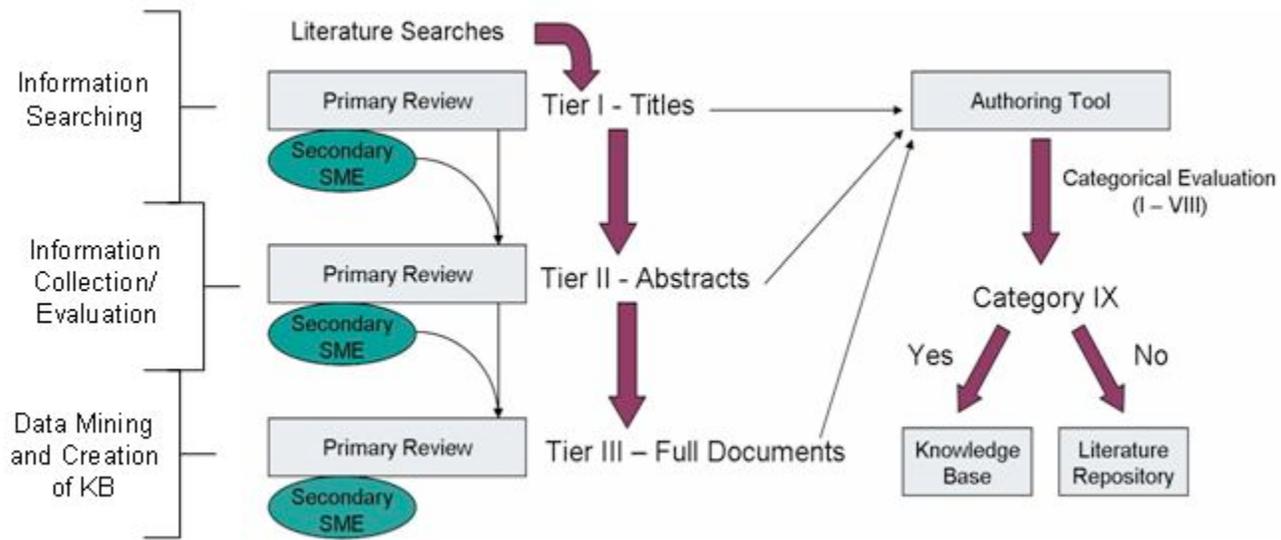


Figure 1 Overall process for development of the database.

1.1 [Routes of Transmission](#)

1.1.1 [Ingestion](#)

- Ingestion of **contaminated food or water** can cause glandular tularemia.²⁷
- *F. tularensis* can be transmitted to humans via **contaminated water** and **inadequately cooked meat**.¹²⁵
- Ingestion of **contaminated food** may cause clinical disease, although with less frequency than mucosal contact with tissue or fluids of infected animals and the bite of infected arthropods.²⁵
- Sporadic tularemia cases in the United States have been associated with **contaminated drinking water**.²³

1.1.2 [Inhalation](#)

- Breathing **contaminated air** can cause pneumonic tularemia.²⁷
- The inhalation of **contaminated dusts** may cause clinical disease.²⁵
- Tularemia has been associated with **lawn mowing** or **cutting brush** and various **laboratory exposures**.^{23 122}
- It has been hypothesized that *F. tularensis* could be aerosolized if a **nest harboring tularemia-afflicted rabbits** were passed over with a lawn mower.²⁴³
- A laboratory technician who handled the **blood cultures** of a patient eventually diagnosed with tularemia developed antibodies for *F. tularensis* 10 days after sniffing one of the culture plates.¹⁵¹
- An outbreak of pulmonary tularemia was proposed to be from the inhalation of contaminated particles from a serologically positive dog's fur disseminated by the **dog shaking itself**.⁶¹

Figure 2. Screenshot of selected information available in KB for Routes of Transmission section of “Disease Transmission” node.

SERRA

Support for Environmental Rapid Risk Assessment

Document Criteria Report

Respiratory pathogenicity of Bacillus anthracis spores. II. Genetic variation in respiratory pathogenicity and invasiveness of colonial variants of B. anthracis
 Zelle, M.K., Lincoln, R.E., Young Jr., G.A.,;

Category	SME Assessment
Document Type	Peer-Reviewed English Language Literature, 1925-1975 ●
Literature / Scientific Content Level	Primary Research Publication ●
Experimental Methods	Incomplete/Missing Important Information ●
Statistical Analysis	Not Applicable ●
Technology / Methods Used	All Established Technology/Methods ●
Model Organism(s) Used	No Surrogate Agent Used, Original Agent Used for All Studies ●
Controls	Controls Not Indicated/Described ●
Distribution	Unclassified, Open-Source, Distribution Unlimited ●

[Key for Criteria Report](#)

Figure 3. Document criteria report available to SERRA users for each document in the LR.

What Is SERRA?

- * Custom web application and Oracle database
 - SERRA permits:
 - Fine-grained management of user access
 - Browsing, searching, and annotating of expert-mined content
 - Access to reference documents and document assessments
- * Comprised of two parts
 - Knowledge Base (KB)
 - Knowledge compartments of risk-related information for biological contaminants, chemical contaminants, and radiological/nuclear materials
 - Literature Repository (LR)
 - Currently contains 2,159 electronically-formatted articles including quality assessments
- * Accessible to risk assessment specialists and emergency response staff through EPA Intranet



Figure 4. A description of SERRA's web application, Oracle database, and main components.