PBPK Literature Visualization Project: ModelMap

Authors:

Susan Ritger Crowell[1], Scott Dowson[2], and Justin Teeguarden[1]

[1] Systems Toxicology and Exposure Science, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

[2] Visual Analytics, Fundamental and Computational Sciences, Pacific Northwest National Laboratory, Richland, WA

Background and Objectives

The development and evaluation of PBPK and QSAR models remains a relatively new discipline; therefore readily accessible and high-quality resources are necessary to support increasing model use. The resource needs are broad, ranging from parameter values, to model archives and publication archives. A common thread is the need to organize the available PBPK model related literature in a fashion that supports rapid access and organization according to data needs and more detailed analysis.

Dr. Rocky Goldsmith, U.S. EPA, presented Battelle with a general plan to create novel information resources that will support the chemoinformatic needs and or data organization and archiving in support of the U.S. EPA software program PRePaAE specifically and PBPK modeling in general. The series of ideas proposed by Dr. Goldsmith related to the development of an approach to automate the extraction of data related to PBPK modeling from a body of literature pre-identified as being related to PBPK modeling (or metabolism, etc.).

Battelle has developed, for other government clients, a number of informatics tools that support the kind of automated analysis and organization sought by the U.S. EPA for the PBPK modeling literature. In this work assignment, Battelle modified existing computational and informatics tools to develop and expand a first generation tool, ModelMap, that can be used to automatically organize and give rapid access to the body of PBPK modeling literature. The application can be accessed and utilized at http://pbpk.labworks.org/.

Progress and Workflow

PBPK literature naturally lends itself to hierarchical organization and categorization, features that we have exploited to create a powerful literature navigation interface. Tree mapping is a method of information visualization and navigation for hierarchical data such as the PBPK literature: data are organized thematically and can be navigated by the user choosing a theme of interest, and then drilling down through sub-themes to the level of individual data or documents.
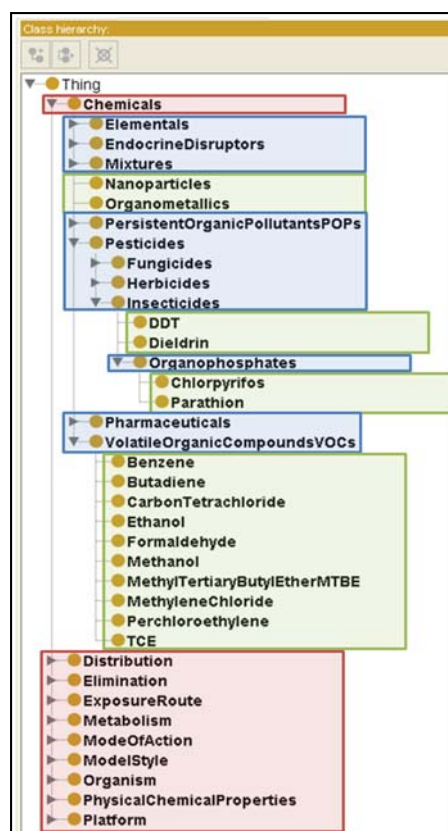
Hierarchy Development

Literature associated with PBPK and biological modeling was identified using a specific search string (Appendix) in the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed), yielding an initial database of > 2,800 MEDLINE abstracts. These data were organized into a logical hierarchy (i.e., taxonomy) by subject matter experts (i.e., toxicologists and modelers) using Protégé, a publicly available ontology editor (http://protege.stanford.edu/).

The taxonomy (Figure 1) used to create the tree map reflects the native ways in which users engage the PBPK related literature. Major categories of interest, called fundamental nodes, were identified, as well as sub-nodes and leaf nodes, the smallest and most specific unit of the taxonomy. For example, the fundamental node "Chemicals" contains sub-nodes such as "Volatile Organic Compounds (VOCs)" and "Pesticides." The sub-node "Pesticides" contains additional sub-nodes ("Fungicides", "Herbicides", and "Insecticides"), while the sub-node "VOCs" contains leaf nodes representing individual compounds (e.g., benzene).

Subsequent to an initial round of taxonomy development by one subject matter expert, the taxonomy was reviewed first by an internal panel of subject matter experts at PNNL, and then by external experts in the field. The internal panel was comprised of Jordan Smith, Justin Teeguarden, Susan Crowell, Rick Corley, and Chuck Timchalk (Systems Toxicology and Exposure Science, PNNL); Dennis McQuerry (Visual Analytics, PNNL); and Torka Poet (formerly Systems Toxicology and Exposure Science, PNNL; now Summit Toxicology). The panel members were briefed on the ModelMap application, individually reviewed the taxonomy, and then discussed the taxonomy's structure as a group. The combined comments of the internal panel were incorporated into the taxonomy, which was then dispersed to external reviewers. The external reviewers were Eva McLanahan and Paul Schlosser (US E.P.A), Jerry Campbell (The Hamner Institute), Hugh Barton (Pfizer), and Jeff Fisher (NCTR/FDA). Reviewers were again briefed on the ModelMap application and underlying taxonomy via webinar, and then invited to submit comments on the taxonomy. These comments were incorporated into the final version of the taxonomy delivered to the clients and used in the ModelMap application.

Figure 1. Taxonomy of Major PBPK Themes. Red shaded items were identified as "Fundamental Nodes" or themes. Blue shaded items are "Sub-Nodes" and green shaded items are "Leaf Nodes."

Literature Search Strings

IN-SPIRE, a unique and powerful data visualization tool that facilitates identification of linguistic patterns and major themes in data sets that might not be readily apparent (http://in-spire.pnnl.gov/), was used to analyze abstracts for key terms for further development and refinement of the taxonomy, as well as to develop and test search strings to isolate specific documents relevant to each leaf node of the hierarchy (Figure 2). For example, the search string used to identify abstracts under the leaf node "Sprague Dawley Rat" (Taxonomy path: Organism > Rodent > Rat > Sprague Dawley) is:

("SD (rat rats)" "sprague dawley" sprague-dawley)

Entering this search string into IN-SPIRE isolates the 196 abstracts focused on PBPK models or relevant research conducted using Sprague Dawley rats (Figure 3). Developers and subject matter experts worked together with IN-SPIRE to generate appropriate search strings for each of the leaf nodes identified in the taxonomy. This iterative process involved testing proposed search strings against the database of MEDLINE abstracts, assessing the returned documents for relevance, refining the search strings based on these assessments, and so forth.

Once developed, search strings were incorporated directly into the taxonomy within Protégé, the ontology editor.
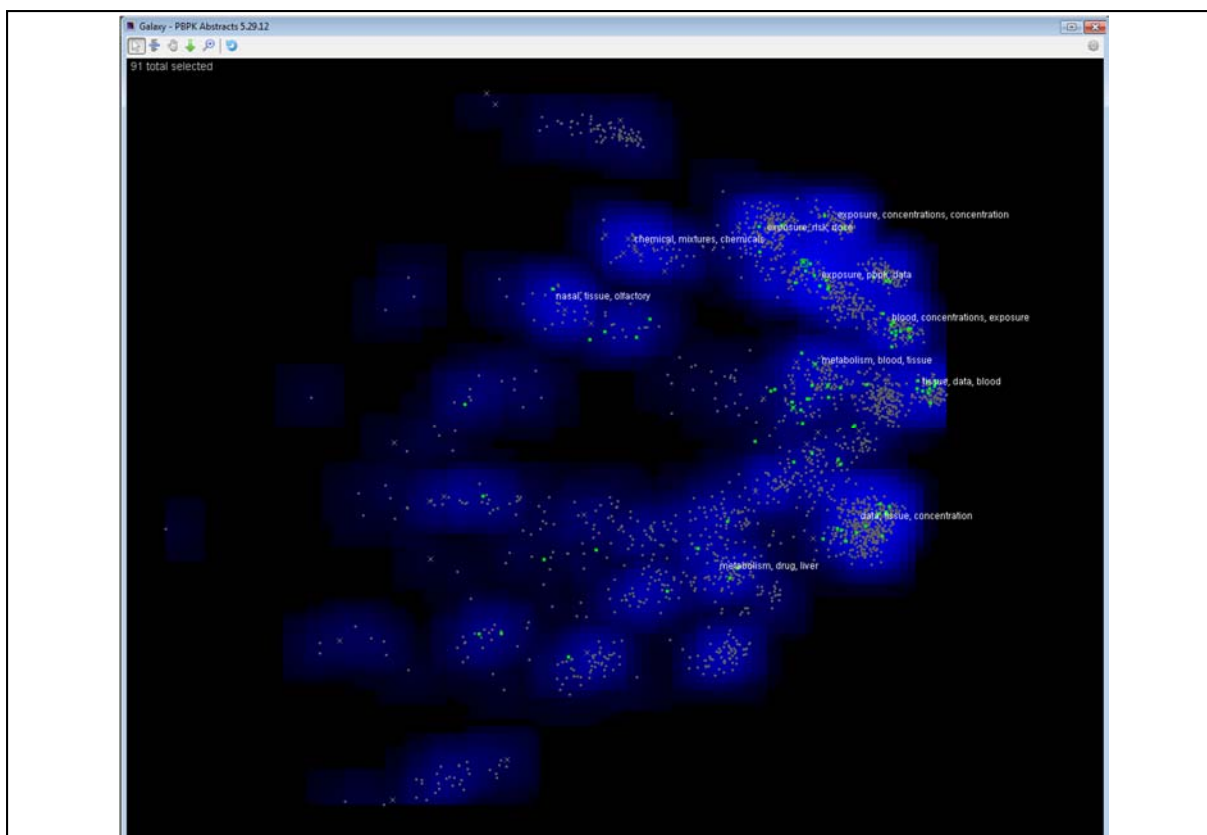
Figure 2A. IN-SPIRE galaxy view of PBPK literature space, with documents relevant to Sprague-Dawley rats highlighted in green.
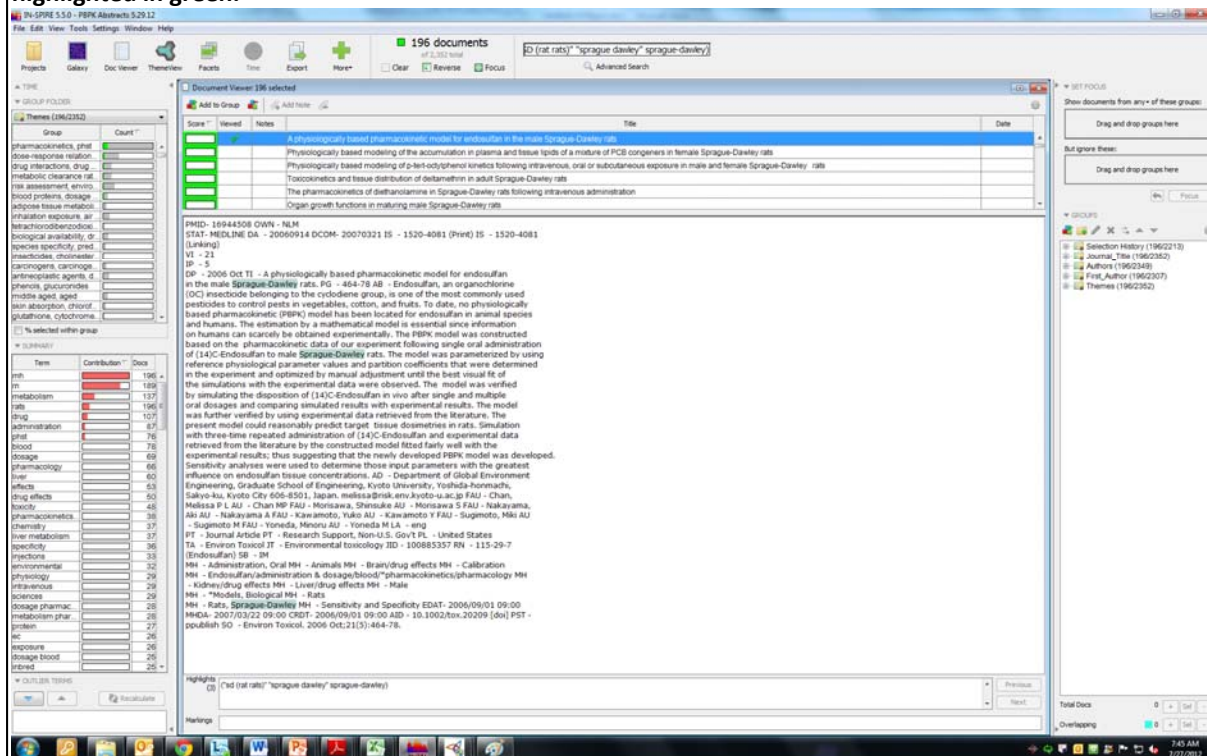

Figure 2B. IN-SPIRE document viewer, focused on documents relevant to Sprague-Dawley rats.

Tree Map Visualization

The publications from PubMed are obtained through email in a text format:

> PMID- 24671884
> OWN - NLM
> STAT- Publisher
> DA - 20140327
> …

This data is converted to XML and goes through a tagging process based on the search strings defined for each leaf node in the taxonomy. A utility that is a part of IN-SPIRE, called Taggert, processes an XML document, and analyzes it in the context of a supplied taxonomy. The text of each publication is extracted and tested against each node's search string. When a match is detected, for example:

> **Node**: (Organism > Rodent > Rat > Sprague Dawley)
> **Search String**: ("SD (rat rats)" "sprague dawley" sprague-dawley)
>
> **Abstract**:
>
> … PET radiotracer for mGlu4, and characterized its biological properties in **Sprague Dawley** rats. [(11)C]3 was synthesized from …

That abstract will be annotated with the path (Organism/Rodent/Rat/Sprague Dawley). As the tagging process is solely driven by the structure and search strings in the taxonomy, a publication may by tagged with multiple taxonomy paths, or none. Based on all the tags captured in the set of publications, a tree map visualization is then used to organize and display the set of abstracts in the context of the organizational structure of the taxonomy.

The finished ModelMap application can be found at http://pbpk.labworks.org/. The tree map style of visualization shows the distribution of publications based on their associated tags at a specified node in the taxonomy, and allows the user to navigate rapidly through subjects of interest to the associated abstracts. Initially the application opens to shows the distribution at the fundamental node, i.e., the highest level of categorization. Each colored [fundamental node] box (e.g. 'Chemicals', 'Model Style', 'Exposure Route') is sized so that the area is proportional to the number of abstracts tagged as relevant to that category. The tree map visualization also shows the sub-nodes of each fundamental node, using various shades of the fundamental node's color (Figure 4.). This provides more insight into the distribution of the data beyond the parent node. The visualization supports navigation and drill-down through the taxonomy by allowing users to click on a particular [fundamental] node which resets the visualization anchoring on the selected node, and then showing the distribution of data on the sub-node and leaf node level.
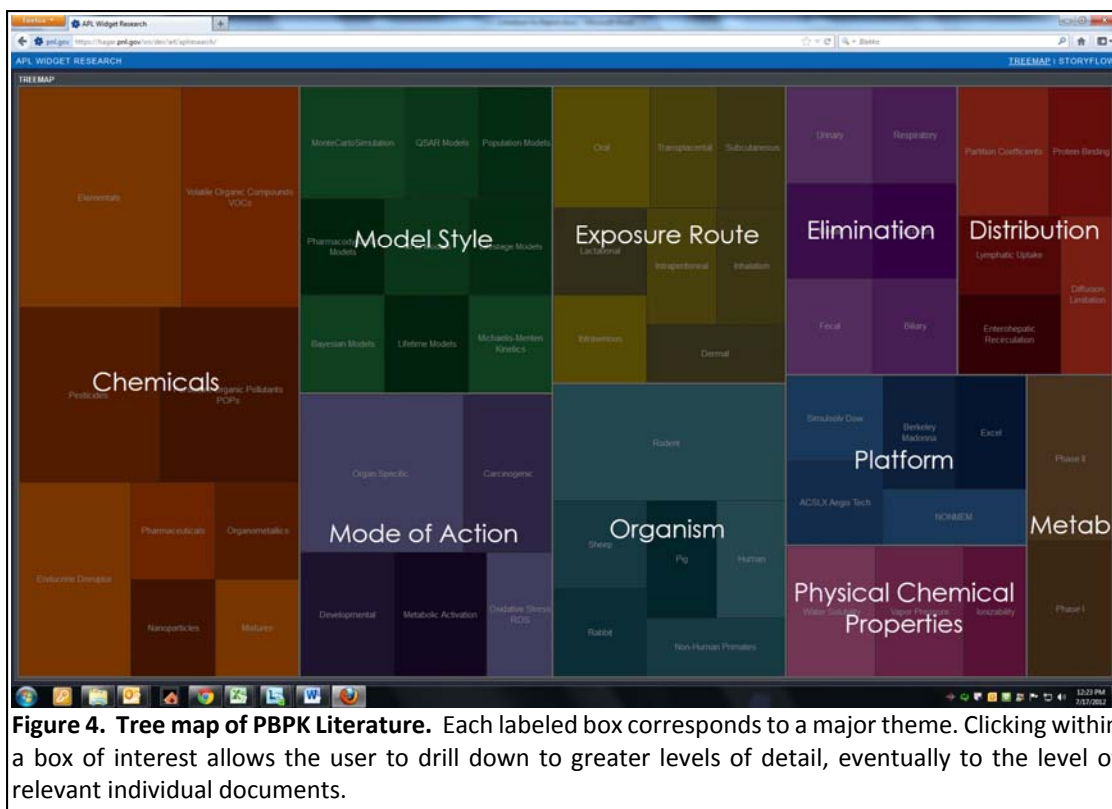
Figure 4. Tree map of PBPK Literature. Each labeled box corresponds to a major theme. Clicking within a box of interest allows the user to drill down to greater levels of detail, eventually to the level of relevant individual documents.

Autoupdate Features

The search string used to identify the initial literature database was saved in PubMed using a dedicated email address (pbpkharvest@pnnl.gov). Every day, PubMed emails a text report of all new publications that match the saved search. An automated process downloads these emails, parses the reported abstracts into an XML format, tags them according to the existing ModelMap ontology, and finally updates the final collection of all archived publications.

Additional Documentation

Installation details can be found within the delivered code, and at http://srs.pnnl.gov/wiki/index.php/Admin:installing-SRS.

Appendix

Search string used to identify relevant literature in PubMed:

pbpk[All Fields] OR pbtk[All Fields] OR ((physiologic[Text Word] OR physiologically[Text Word] OR physiological[Text Word]) AND based[All Fields] AND (pharmacokinetic[Text Word] OR toxicokinetic[Text Word] OR pharmacokinetics[Text Word]) AND (model[Text Word] OR models[Text Word] OR modelling[Text Word])) AND English[lang]