Background and Introduction

The PRISM e-Label initiative is being undertaken to modernize the label approval process between registrants and the Environmental Protection Agency's (EPA) Office of Pesticide Programs (OPP). Label approval will move from a manual, document centric paper-based process to an electronic stream of data. In this paradigm electronic data will be passed through all stages of the labeling process, from the creation of the label through review by regulatory staff to distribution of the approved label information.

Around 2003, the Registration Division began an effort where registrants were encouraged to submit labels as text-based PDF files in addition to the required paper copies. This allowed for a comparison of label content of the newly submitted label with that of the previously approved label. The comparison process used in Adobe Acrobat, however, proved to have its limitations and difficulties. In February 2008, the Information Technology and Resources Management Division (ITRMD) of OPP initiated a project to look into the structuring of pesticide label content using eXtensible Markup Language (XML). An internal workgroup was formed to address whether a pesticide label could be broken down into constituent parts and to further identify what those parts or elements would be.

Based in general on the FDA's Structured Product Labeling, OPP's e-Label would define the content of the pesticide labels submitted to the EPA in a structured, XML format. Using a specified XML schema, all label content would be identified as discrete elements which would allow for a common mechanism to exchange the label content between registrants, EPA, and other stakeholders. Once the label content is identified as individual fields, the structured content could then be easily compared, parsed, searched, stored, or rendered.

As an aid to the registrant community, an e-Label Builder application would allow for the easy creation of a structured label. By including business rules and validation logic into the e-Label Builder, the application would serve as an automated reminder to include basic textual requirements and encourage registrants to use standard precautionary text and approved language. The original idea was to have the Builder function as a "wizard" application similar to that used by commonplace software products such as TurboTax. More recent thinking is the creation of a smart Word template that would allow the saving of the label content in the defined XML format yet maintain the familiar functionality of Microsoft Word. The use of a Word document may better integrate into registrant's existing business processes with minimal impact yet still allow for the creation of the desired XML data file.

Because the label content submitted in XML format is "formless", one or more standard renderings will need to be created to display the label information. It is expected that a style sheet for each of the major pesticide types (insecticide, herbicide, etc.) would be required in order to address particulars of each type. The style sheets would render the label information in a standard look so that reviewers and other users of the master label would find the same information in the same place from label to label. The common rendering would only apply to the master label submitted to the EPA. End use product labeling would remain completely with in the domain of the registrant-state regulatory authority relationship. It is thought, however, that a common master label format may facilitate some state review processes.

Based on discussions with the internal workgroup and outside stakeholders the following are some of the envisioned benefits of the e-label project.

1. Process labels more efficiently by allowing EPA resources to focus on critical tasks.
2. Improve data quality and consistency across labels.
3. compare label content against current rules, requirements, guidance, and laws.
4. Provide a level playing field for registrants.
5. Easily identify label changes and make the comparison process more efficient.
6. Reduce data entry burden.
7. Improved search and retrieval of label information.
8. Facilitation of web distributed labeling.

Details of structure analysis

The label content has been categorized into six major organizational categories.

1. *General information* would include overall product and company information
2. *Ingredient information* would include details about the active and inert ingredients and overall diluent information.
3. *Precautionary statements* would include restrictions regarding the environment, human, and user safety
4. *Directions for use* would include instructions on how to mix and apply the product
5. *Additional information* would include warranty and marketing statements and information
6. *Regulatory information* would include EPA tracking and processing information. This last category would be mostly used for communications back to the registrant or across third-party stake holders.

The six categories are divided into 31 label sections to further categorize the content. For example, the *Ingredient Information* category consists of the *Active Ingredient, Density/Total Product Concentration, Inert Ingredient of Concern,* and *Packaging* sections. However, some individual data fields not included with in any specific section such as “Other Ingredient Percentage.” Sections are further broken down into 25 subsections. For example, the *Active Ingredient* section includes the “concentration” subsection where the amount of active ingredient is identified (e.g. “product contains 5.7 pounds AI per gallon”). Overall, 280 individual data fields have been identified as having some merit of being captured. Any field identified as requiring the user to choose from a list of values as the allowed content will also have an “other” option to account for values not currently available to the user. In addition, the six major categories have an “other” field where the registrant may include any label content that can not be fit into the predefined fields.

Individual data fields could take the form of discrete values, large free-text blocks, or a combination of both. Examples of discrete data would be the identification of the active ingredients and their corresponding percentages. An example of the other extreme, the free-text blocks, would be fertilizer mixing instructions. A hybrid field would be a free-text block with an embedded discrete value such as the REI contained with in the Agricultural Use Requirements. Additionally, what may otherwise appear as a free-text block may actually be made up of a group of standard discrete statements. An example of this would be the Environmental Hazard block.

By use of an e-Label Builder, business rules and logic would suggest label content based on certain conditions. The registrant could use the suggested content or choose to make changes to fit their particular needs. The fact that the "standard" text was or was not used would be recorded and submitted as part of the e-label. Once received by EPA, an information technology solution would display the e-label content for review and identify where non-standard content was used. This would allow regulatory staff to focus their attention on areas of concern and potential problem text. Additionally, a comparison of label content with previously approved labels would be more seamless and less problematic than with current methods.

In addition to the actual label textual content, the Builder application would capture non-label information such as the LD50 and similar values. This information would be required to perform some of the validation such as ensuring that the appropriate precautionary statements are included. Future integration with an electronic confidential statement of formula (e-CSF) may also reduce form preparation on behalf of the registrant and create more reliable data.

Questions to address

1. Will a "Builder" application be useful?
2. What style Builder would be more appropriate – Wizard or smart Word template?
3. Can labels be formatted using standard style sheets?
4. Has all of the possible content been accounted for?
5. Are the data fields too detailed, not detailed enough, or just right?