ABSTRACT
Toxics Release Inventory (TRI) Explorer is a web-based tool that enables TRI data users to compile their own reports on-line. Under the Emergency Planning and the Community Right-to-Know Act, EPA is committed to making TRI data easy to access. TRI Explorer was developed to support this goal. TRI Explorer allows users to generate their own reports on specific chemicals, and on chemical releases by industry sectors, by environmental media, by geographic area, and by individual facilities. With the help of the TRI Explorer, users can easily determine what toxic chemicals are present in their neighborhood, how the releases are changing over time, and how their own situation compares to other communities around the country. TRI Explorer provides data for all reporting years (since 1988) and the data are synchronized with the published Public Data Release documents.

The Explorer is available on the web at www.epa.gov/triexplorer.

The presentation will include a demonstration of TRI Explorer and the role of various SAS products used in developing this application will be described.

BACKGROUND
The Toxics Release Inventory (TRI) is a publicly available EPA database that contains information on specific toxic chemicals and other waste management activities reported annually by certain covered industry groups as well as federal facilities. This inventory was established under the Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA), which requires facilities to use their best readily available data to calculate their releases and waste management estimates. If facilities do not have actual monitoring data, submitted values are derived from various estimation techniques.

The TRI Explorer provides access to the TRI data to help communities identify facilities and chemical release patterns that warrant further study and analysis. Combined with hazard and exposure information, the TRI Explorer can be a valuable tool for risk identification.

The original version of TRI Explorer was provided as a downloadable executable and dBase files. The application had limited capabilities. However, the design was excellent and the potential benefits of providing access to the data in an electronic form were realized. In collaboration with the original developer of TRI Explorer Version 1.0, William P. Smith, EPA, we were assigned the task of developing a Web based version of TRI Explorer. Version 2.0 of TRI Explorer was first deployed on May 11, 2000, and due to delays in the printing process it became, technically, the first electronic only TRI Public Data Release (PDR).

The capabilities of TRI Explorer are incrementally enhanced to meet user requests, and refine business rules used in processing and presenting the information. Version 4.0 is currently in development and will add further detail to the geographic area search capabilities.

THE TECHNOLOGY
Or, perhaps an often asked question is; Why SAS? Had this project taken a software engineering approach such as Object Oriented Analysis and Design (OOAD) that separates the desired product from the underlying technology other tools may have been used. However, often in reality, developers are limited to a specific set of resources, time constraints, and domain knowledge. Customer preference for SAS products played a significant role in this decision. However, this preference is well grounded in experience with EPA data system and an understanding of the level of manipulation required to produce meaningful products. In this case SAS has proven and continues prove itself as an excellent tool for developing data driven Web content.

At this time a Web based user interface accesses a report engine developed in SAS through the SAS/IntrNet product. Data stored in SAS tables on the server is queried to produce tabular reports that are returned and displayed in the user’s Web browser. All communication is stateless to avoid the use of cookies.

Admittedly some components of the server side programming will benefit for reverse engineering in order to improve maintainability and adaptability of TRI Explorer. The product was developed in a reactive, ad hoc development environment. Original specifications were grown from existing products and the design architecture was modeled from prototypes. Product acceptance, proposed enhancement, and performance requirements will dictate the need for a more controlled software engineering approach. Ultimately the benefits will be realized in component reuse for additional EPA Web based data exploration tools.

We had tested SAS AppDev Studio at Version 1.0. The potential at that point was very promising although most our focus was on client side development. As look towards a longer term approach to product development may provide the development environment that can support implementation in an OOAD approach.

THE PROBLEM DOMAIN
TRI data is not simple. The rules that determine whom and what must be reported vary from year to year and because of these changes specific rules must be applied when presenting and analyzing the information. These rules, or TRI business rules, are applied in TRI Explorer.

For example: TRI explorer can produce trend reports showing the amount of chemicals reported though the life of the program. In order to produce a consistent view from year to year, TRI business rules are applied so that the implied trends are meaningful. That is, trends are only reported for those chemicals and industries that would have been required to report in every year.

The source data for the TRI Explorer is contained in a highly normalized Oracle database. In order to produce analysis results in reasonable response times TRI Explorer must work with denormalized tables. Also, certain levels of pre-summarization are used to enhance performance were indexing
We found PROC SQL and the Data step advantageous to extract and denormalize TRI data from the Oracle database. The alternative to this approach would involve the use of cursors in Oracle PL/SQL to perform the denormalization. Which would have required adding an addition skill and possibly slowing progress (or perhaps I’m just getting too old to learn one more language).

CONCLUSION
EPA’s Analytical Computing Center (ACC) was formed as a center of excellence for environmental computing, and to provide quick access to senior IT professionals for rapid application development and response. The ACC provides mixed platform development environment and technical support to the Office of Environmental Information. The ACC provides expertise in Information Profiles, Data Analysis, Data Warehousing, Smart Tools, Geospatial Intelligence, Credible Information products, and Data Mining. The Goals will continue but the environment is changing and we must learn to adapt.

SAS products have been proven effective in delivering results. EPA’s ACC must mature and adopt the principles of software engineering to continue developing and providing useful products. Until now the ACC development process has been adaptive and in the work of software engineering this approach declares “The code is the design”. For a small ad hoc environment like the ACC this has been effective in meeting the customer’s need but it is unlikely to be efficient in terms of cost as the product matures.

As we adapt to our new home we hope to adopt the OOAD process with the Unified Modeling Language. In development we hope to capture the advantages of both adaptive and model-drive approaches as suggested by Steve Hudson, META Inc. Doing this SAS tools will provide our customers with an incremental and iterative process to product development.

REFERENCES


Teague, Cynthia, Cynthia Zender, SAS Web Tools: Querying and Updating Data with htmSQL Course Notes, SAS Institute, October 1998.

ACKNOWLEDGMENTS
The author would like to thank the US EPA Office of Environmental Information, which supports this work.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at: