

Toxics Release Inventory Data and Tools

Steve Witkin

U.S. EPA, Office of Information Access and Analysis, Environmental Analysis Division
1200 Pennsylvania Avenue NW, MC 2842T, Washington, DC 20460

witkin.steve@epa.gov

ABSTRACT

This paper will answer two main questions: What does the Toxics Release Inventory (TRI) offer to a data user today? And Why a user discouraged by previous limitation may want to take a new look at TRI. This inventory is mandated by the Emergency Planning and Community Right-to-Know Act (EPCRA) and is to inform communities and citizens of chemical hazards. The TRI is publicly available and containing information on toxic chemical releases, transfers and waste management activities on nearly 700 chemical. Data is submitted annually becomes available starting seven months after the close of the reporting cycle. The database is made available through several dissemination points including Envirofacts, TRI Explorer, and TRI.NET to name a few. Envirofacts is EPA's primary data warehouse containing powerful query capability. TRI Explorer provides quick access to the TRI data in an easy to use and understand format. TRI.NET is a high performance data engine for querying TRI data. This highly interactive tool has a number of advanced query capabilities using ad hoc queries and mapping technologies. TRI.NET features several value added data layers including Tribal Lands, US/Mexico Border and Environmental Justice. The TRI data set is continually growing to add more data and linkages to other datasets. Internal enhancement such as implementation of an Extract, Transform, and Load tool will bring faster availability of the data with improve overall data quality. Several new projects to enhance the availability and usefulness of the TRI data are currently underway and are discussed in more detail.

HISTORY AND BACKGROUND

In 1984, an industrial accident in Bhopal, India resulted in the released of a deadly cloud of methyl isocyanate killed thousands of people and injuring over half a million. Shortly thereafter, another industrial accident in Institute, West Virginia resulted in leaks of toxics chemicals. These and other events spurred public interest and environmental organizations around the country to accelerated demands for information on toxic chemicals being released "beyond the fence line" -- outside of the facility. Against this background, the [Emergency Planning and Community Right-to-Know Act \(EPCRA\)](#) was enacted in 1986.

One of EPCRA's primary purposes is to inform citizens of toxic chemical releases in their areas. EPCRA Section 313 requires EPA and the States to collect data annually on releases and transfers of certain toxic chemicals from industrial facilities and make the data available to the public through the Toxics Release Inventory (TRI).

In 1990 Congress passed the [Pollution Prevention Act](#) which requires facilities to report additional data on waste management and source reduction activities to EPA under TRI. The goal of the Toxics Release Inventory Program is to provide communities with information about toxic chemical releases and waste management activities and to support informed decision making at all levels by industry, government, non-governmental organizations, and the public.

Following the inventory's rapid start and growth, it is now undergoing a phase of maturing with modest growth. The first public data release of TRI data was for the chemicals released in the 1987 calendar year. This data release took nearly 2 year from data receipt to data availability, was only

available in print form, only reported chemical releases and transfers of manufacturers and covered approximately 300 chemicals. Since that beginning much has been added or improved with the only substantial elimination being the annual multi inch thick paper document of charts and tables. Today EPA can turn much of the data around from receipt to electronic availability in as little as 1 to 2 months, data is now available in a wide variety of electronic data files and computer based access tools including availability on smart phones, Data users can now generate the tables and graphs which are of interest to them at their convenience, data on chemical waste management activities has been added, facilities which feed raw materials or energy to or receive wastes from the manufacturing sector must now report to the inventory, and the number of chemicals covered has more than doubled. With the size of the data universe under TRI still growing, albeit at a somewhat more conservative pace, the focus the last several years has shifted towards streamlining the data quality process and improving the usefulness of the data.

DISCUSSION

How has TRI changed recently?

Build in Data Quality at the Source

Starting in early 2008, the TRI-ME web application became available to all facilities. This single tool allows EPA to process the incoming data faster, reduces data processing costs, and improves data quality. TRI-ME web is an online data submission tool using a question driven approach to assist data submitters. Drop down pick lists are available where appropriate, eliminating spelling errors and ensuring users conform to standard naming conventions for items such as chemical names. When a pick list is not a suitable option, the reporting tool increasingly has been using pre-populated data fields for items such as Facility Name to increase year to year consistency in the text. Each year additional quality checks are being added to alert users of entries which may be atypical for the specific submitter.

Over the last three or four years, increase attention has been given to data elements which are not pounds of a chemical. Two such fields are parent company name and RCRA ID number, required to identify the destination of off-site transfers. Although these items are not release data, they are very important to various data users. In the past, the data user had to perform additional data cleansing steps to maximize the usefulness of all the data elements within the inventory. Again, by assisting the data submitter with pick lists and data checks against other EPA data systems, we have improved the data quality at the point of submission.

Unique data issues arise when data gaps exist in data external to the TRI program itself. One of the most significant data elements in this category is the facility spatial data. At the inception of the TRI, spatial data, specifically latitude and longitude, of a facility was provided by the facility during the data submission process. Due to the large quantity of bad data being submitted, the TRI program moved to using the spatial data in EPA's Facility Registry System (FRS). During the past several months and working with the Office of Information Collection (OIC) within the Office of Environmental Information (OEI) we have shrunk the size of the data gap in half simply by altering the logic used to determine what is "good" versus "bad" data. These corrections are in place today and benefit all EPA programs which utilize the FRS data system. The remaining gaps must use a manual facility by facility process and in many cases may require a site visit to the facility. We expect to have this last batch updated within the next several months.

Data Integration

For many years there has been a desire to integrate the TRI data with other data sets. These other data sets broadly fall into two categories: Non-regulatory data associated with a facility's location, such as watershed, proximity to Indian lands, metropolitan statistical areas (MSAs) and EPA programmatic data sets such as the Discharge Monitoring Report (DMR) and the Air Facility System (AFS). Today a few of the TRI data access tools contain some non regulatory data which can be used either simply as a

displayed element or as a data filter. In fact the very positive response to tribal filters in some TRI tools has led to a project developing a tribal filter for EPA's Envirofacts system. This filter is due out to production within the next few months or sooner. The latest integration of the TRI data with other EPA datasets has occurred in the form of stand alone comparative tools and will be discussed in more detail later in this paper.

More Data

To round out the latest changes, the TRI data set is also growing in size. With the release of the reporting year (RY) 2011 data set, sixteen new chemicals have been added to the inventory. This data started to become available late July 2012. Due to ongoing quality checks, the full data set is not expected to become available until December of 2011. The TRI program has also grown in size in the number of governmental leaders who may petition EPA to modify the chemical list or add additional facilities which fall within their legal jurisdiction. Tribal leaders now have some authorities which previously were only available to state governors.

How to access TRI Data?

Bulk Data

There are a number of ways to access the TRI data. For users who desire a automated method to create web mash-ups or load data directly into an application such as Microsoft Excel or Access, the Envirofacts system now offers a RESTful data service API. For additional information on this service refer to the documentation at <http://www.epa.gov/enviro/facts/services.html>.

For a more traditional access to the bulk data, two sets of data files are available, the Basic Data Files and the Basic Plus Data Files. The Basic Data files contain the most frequently used data elements, while Basic Plus contains all the data elements in the TRI reporting forms. Users of both the web service and the data files should consider these forms of the data as being as close to the raw data as possible. As the TRI regulations have changed over the years so have the data elements associated with the data. Successful use of the data files requires a solid understanding of the individual data elements, how they may relate to other data elements and how they may have changed over time.

Data Access Tools with Filtering Capabilities

The next level of data access is through a series of data access tools. The simplest and the latest tool is MyRTK which has been designed principally for viewing on small mobile devices. To maximize the number of compatible devices, MyRTK is a mobile web application as opposed to device application. Based on a given address, city or zip code, the application returns a map displaying the location of TRI facilities, along with some facilities which are required to report under the Clean Air Act (CAA), the Clean Water Act (CWA) or Resource Conservation and Recovery Act (RCRA). Only those facilities which are also likely to release large quantities of TRI chemicals and are regulated under one of the previously mentioned programs are displayed. As shown in Figure 1, the application also provides a brief snapshot of a facility's TRI releases and a bit of the facilities compliance history. This tool has been designed primarily for citizens interested in facilities in their communities.

Moving up the ladder of complexity is the TRI Explorer tool, shown in Figure 2, which has recently been migrated into the Envirofacts data system. TRI Explorer is web based and is one of only two tools which provides aggregated data at the state, county, tribal, chemical, industry levels along with a few other breakdowns. Although not all of the data elements contained on the reporting form are included, all of the release, transfer, and waste management data are included. TRI Explorer is also the most user friendly of the larger data tools. A large number of links are included for assistance in understanding the data filters and to understand terms used in the generated reports. Due to the fact that based on what data is selected and how it is filter a large number of unique conditions may need further

explanation. Every TRI Explorer report includes a custom generated set of footnotes to support understanding of the data returned in a report. A number of business rules are also embedded into the TRI Explorer application to both improve the usefulness of the data and minimize misuse of the data. One way TRI data can be misused is to give the appearance the release of a specific chemical or releases by a certain industry sector are increasing when in fact only a regulation change has occurred resulting in previously unreported releases now being reported. TRI Explorer trend reports break the resulting data tables apart at points where regulatory changes resulted in a change in the reported quantities. TRI Explorer also includes an algorithm to correct for over counting the releases and transfers when material from one TRI facility is shipped to a second TRI facility.

Figure 3, is a view of TRI.NET which is a tool designed specifically for power users. This is a downloadable tool built for speed and agility. TRI.NET includes the capability for a user to modify the SQL script, greatly increasing capabilities beyond the preformatted user interface options. Queries can easily be saved and sent to other users which help a community users to more easily work together. Mapping capabilities to Google Earth and Google Maps, radius search from any user specified location, along with some basic Environmental Justice data round out some of the advanced features of this application. One drawback of TRI.NET is since it was built for speed, most explanatory text and user aids have intentionally been omitted.

The last set of TRI data specific tools is the Envirofacts set of TRI query tools. Figure 4 illustrates the home page for Envirofacts with the TRI tools listed in the lower right corner, among other locations on the page. Excluding TRI Explorer, Envirofacts contains a series of 5 query tools ranging from basic to highly complex and customizable. One of the distinguishing concepts behind the Envirofacts tools is, with a few exceptions, they are facility based or form based. The level of detail at the facility level in the basic TRI Envirofacts queries is much greater than in the previously mentioned tools. In fact both TRI Explorer and TRI.NET link directly to specific Envirofacts reports for this additional detail data. TRI Envirofacts is also tightly integrated with FRS and RCRAInfo so that updates in one system immediately filter into the next system. This does result in a tradeoff between data stability versus availability of the latest data in the tools.

Additional Tools closely associated with TRI data

Data for the sake of data has little social benefit by itself. Emissions data is just one key piece of the elusive puzzle in addressing public risk. Two tools are available to analysts and researchers to move beyond simply pounds of a chemical. First is the Risk-Screening Environmental Indicators (RESI) which is managed by EPA's Office of Pollution Prevention and Toxics (OPPT). RSEI is intended to be used as a screening tool to put the TRI release data into a chronic health context. When properly used RSEI has proven itself to be a very valuable aid in the looking at risk.

Often researchers are looking for a possible relationship between TRI emissions and specific health outcome. Looking across nearly 700 chemicals at one time, typically leads to very weak correlations if at all. Analysis of a single chemical at a time is both time consuming and again often leads to weak correlations due to limits in the size and distribution of data universe. Current health related research using TRI data attempts to look at groups of chemicals with one of the burdens on the researcher is in the selection of the right group of Tri chemicals. TRI-Chemical Hazard Information Profiles (TRI-CHIP) is a tool which includes the entire list of TRI chemicals and complete hazard information on each chemical. By working backwards, a user can identify the health end point of concern or a custom query of toxicity criteria and use TRI-CHIP to identify the associated TRI chemicals. This database contains links to all of the original documentation used to evaluate the health effects reported by entities such as EPA Office of Pesticide Programs (OPP), California Environmental Protection Agency (Cal/EPA), National Toxicology Program (NTP) and more.

Data Comparative Tools

Another set of tools closely associated with the TRI data, are comparative tools. These tools place TRI data alongside data from other EPA inventories. The multiple statutes which guide EPA regulations often result in some reporting overlap between programs. Programmatic differences often prevent a direct one to one match of data elements, but there is the expectation for some similarities. When differences are identified they could, be as previously stated, simply a difference in the reporting requirements, however the differences may also be due to data quality or compliance issues. TRI Comparative Analysis Tool was originally developed specifically to screen for data quality and compliance problems. The TRI Comparative tool was designed to be a partner site to EPA's Enforcement & Compliance History Online (ECHO) tool.

A similar feature to compare TRI data to other data sets has been added into the Discharge Monitoring Report (DMR) Pollutant Loading Tool. Although the DMR tools was not specifically design as a multi program comparative tool, both program offices involved saw the potential benefits of being able to place our data side by side. A recent outcome was where we saw a large TRI release to water yet a lack of data in the DMR for a specific facility. This ultimately led to the realization of an inconsistency in permitting requirements. Although neither a data quality nor compliance problem existed, EPA now has TRI data to support the need for improved permitting requirements in the future.

All of the data sets and tools mentioned above are easily accessible from the TRI home page at www.epa.gov/tri under the sidebar heading of TRI Data and Tools. A portion of this page is illustrated in Figure 5.

What is in the near future for TRI data users?

There are several activities underway to enlarge the TRI data universe and bring out more of the TRI data faster, easier, and with more hooks to combine TRI data with other data sources. As mentioned previously, 16 new chemicals have been added to the TRI chemical list and this data is already flowing into some of the public access tools. Additional data elements which we hope to role out in at least some of the TRI tools by the end of the year deal with Pollution Prevention (P2) data. The TRI reporting forms include a section number 8.11 which is an open text field where a facility can describe their P2 activities. Since this data element is often left blank and it is not a simple numeric value or a simple name descriptor, it is more difficult to use this data element for automatic analytical purposes and has not previously been pushed out into the TRI tools. There is also data from section 8.10 of the reporting form which includes codes for the types of P2 activities a facility may be implementing. The section 8.10 data is already in the Envirofacts tools and we are looking at the feasibility of pushing it out into more tools in the near future.

Internally to the program a new Extract Transform and Load (ETL) tool will be used to move the TRI data from our processing center into the Envirofacts tool. The ETL process has under gone extensive testing over the past year and is scheduled to be the principle method of moving TRI data into the Envirofacts database. This process will eliminate several manual steps currently required to move the data. Since every manual step is an opportunity for data quality bugs to enter the process additional layers of data quality checks are currently required. By automating the process, the data will be able to flow faster from receipt to dissemination points, at a lower cost, and with a higher confidence in data quality. The associated cost savings opens the opportunity to update the data more frequently. At the moment the ETL tool will strictly be used to feed the Envirofacts system. Potential streaming custom data feeds to other EPA and non-EPA data systems directly is on the horizon.

The GeoPlatform is a new agency wide capability which will unify how we overlay our emissions data with spatial data and offers new options for how we view our data. TRI likely will initially be using the platform to improve the tribal filters. The GeoPlatform's tribal data layer is improved and more current than the boundary set current used to filter the TRI data. Initially the

platform will only be available to users within the EPA firewall. However, results generated including maps generated may be electronically published to publically viewable web sites.

The last major tool activity currently underway is an analysis of all of the current TRI data dissemination tools within EPA, with a goal of integrating the spectrum of tool into a single unified tool or a smaller set of tools. At the heart of this integration effort are the analytical capabilities to automate a number of data quality steps which currently are performed using a combination of some of the existing tools, a few custom desktop tools and a large dose of manual intervention. Better integration of the TRI data access tools should also reduce frustration some users face in accessing TRI data. Currently users are not always aware of what options are available and often end not using the best tool available to meet their specific needs.

The final major set of activities underway we anticipate will improve the inventory in an indirect fashion. University engagement and community engagement projects along with a project to investigate how to better meet the needs of Tribal Environmental offices have all been launched during the past year to improve the understanding of the TRI data, the tools available, and refine EPA materials to be more relevant to the unique needs of these specific user groups. As we make the TRI data more understandable and usable to more individuals, the greater the opportunities there is for other to alert us to questionable data or more importantly identify facilities which intentionally or unintentionally are not reporting to the inventory and have slipped under everyone's radar.

The current TRI data and support tools has undergone significant improvements to enhance the value offered both to technical communities and to the citizens the data is meant to serve. Some limitations with use of the data have recently been addressed. In many ways, the Toxics Release Inventory today is not the same creature many users interacted with or heard about years ago. For many, it may be time to take TRI out for another spin around the block.

For additional information on any of the topics discussed in the paper, your best single point to visit is the TRI program home page at www.epa.gov. If the information available does not meet your needs, please use the contact us link near the top of the page.

List of Figures

Figure 1. Screen shot of MyRTK application of a facility report

Figure 2. Typical user interface for TRI Explorer

Figure 3. View of TRI.NET with grouping variables displayed

Figure 4. View of Envirofacts home page

Figure 5. List of TRI data access and analysis tools

Figure 1. Screen shot of MyRTK application of a facility report.

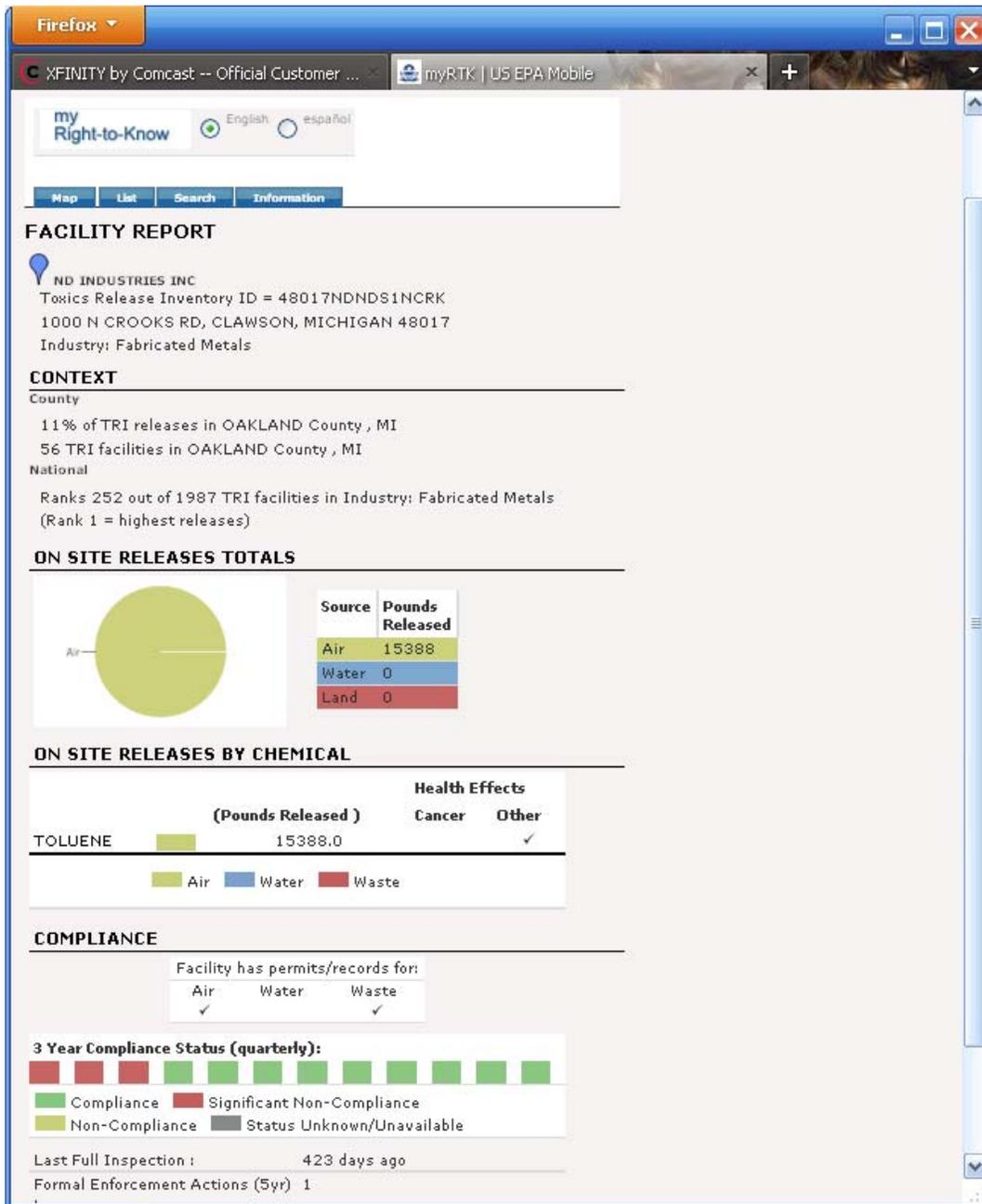


Figure 2. Typical user interface for TRI Explorer.

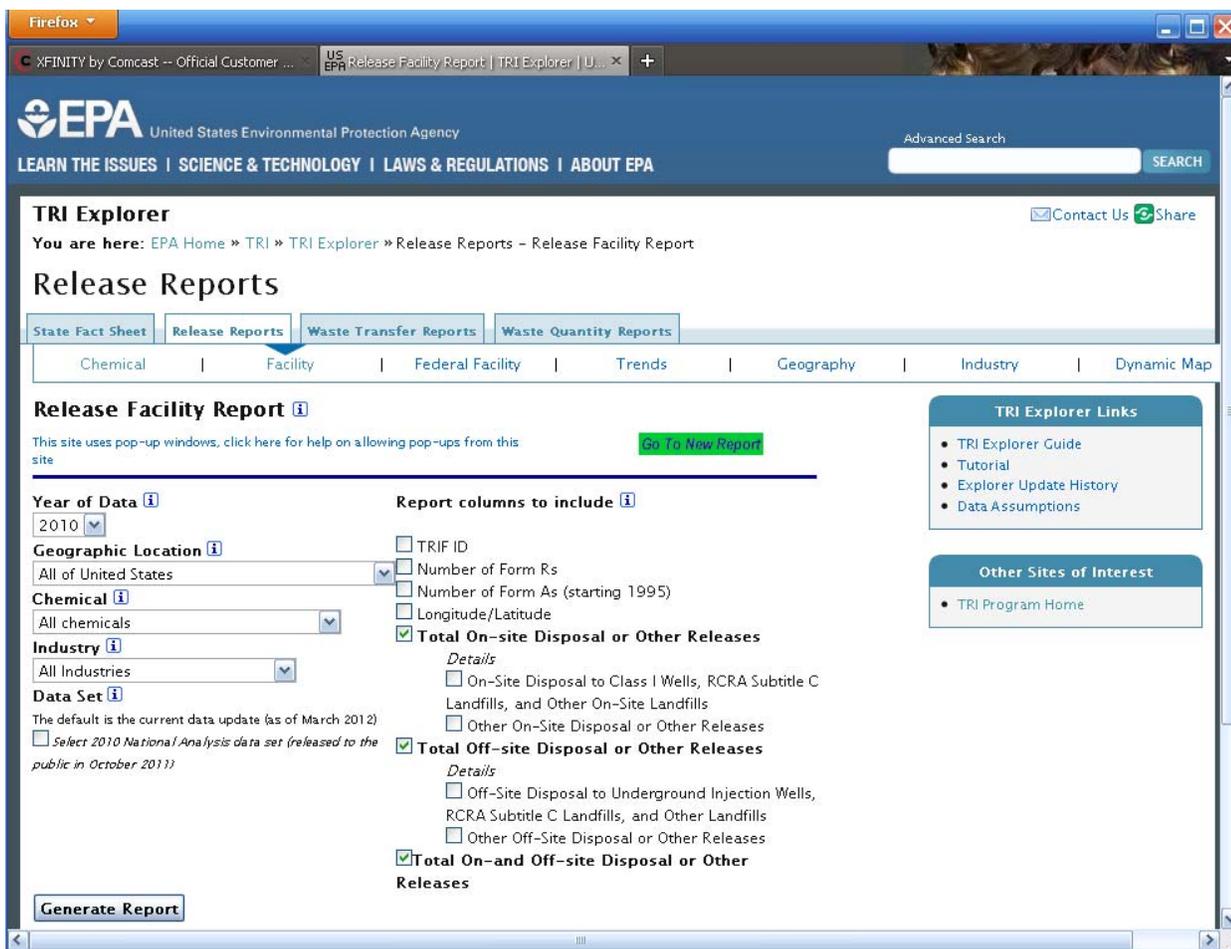


Figure 3. View of TRI.NET with grouping variables displayed.

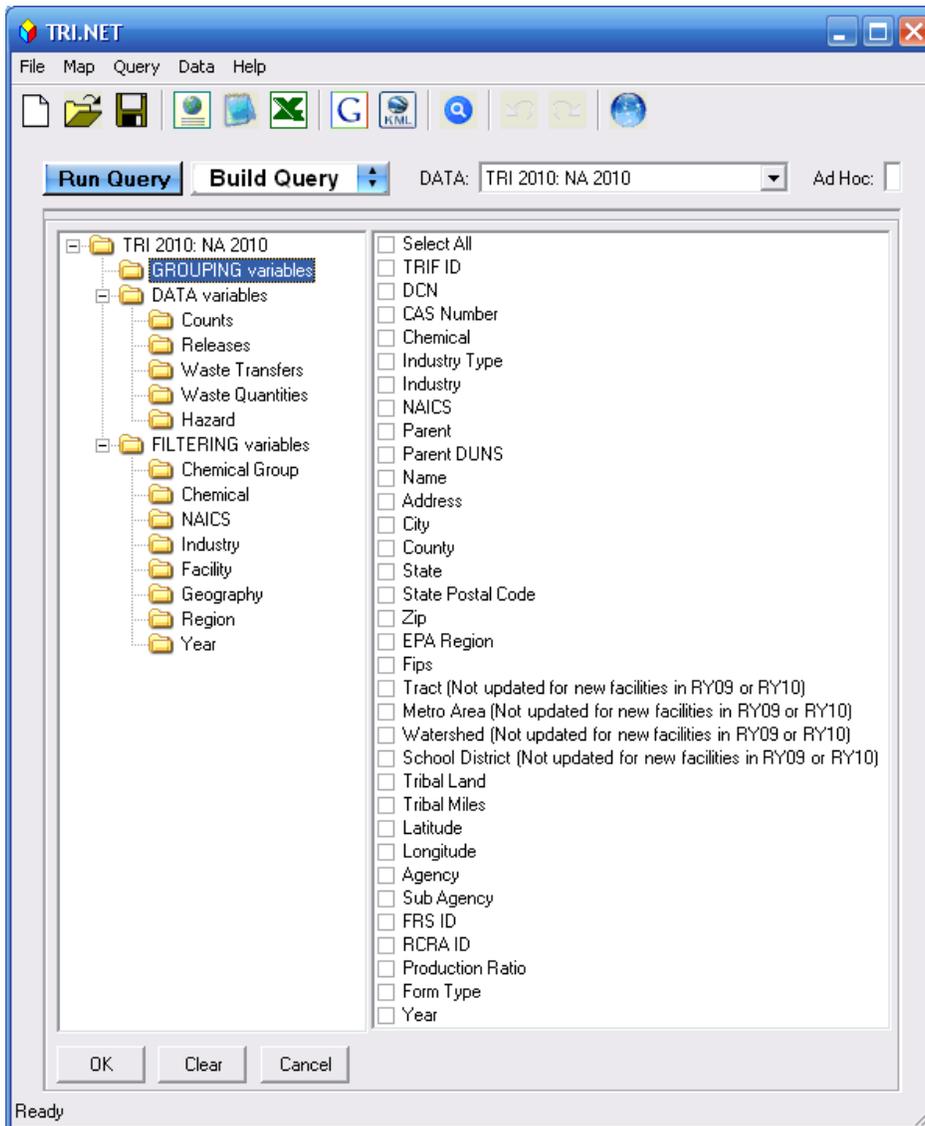
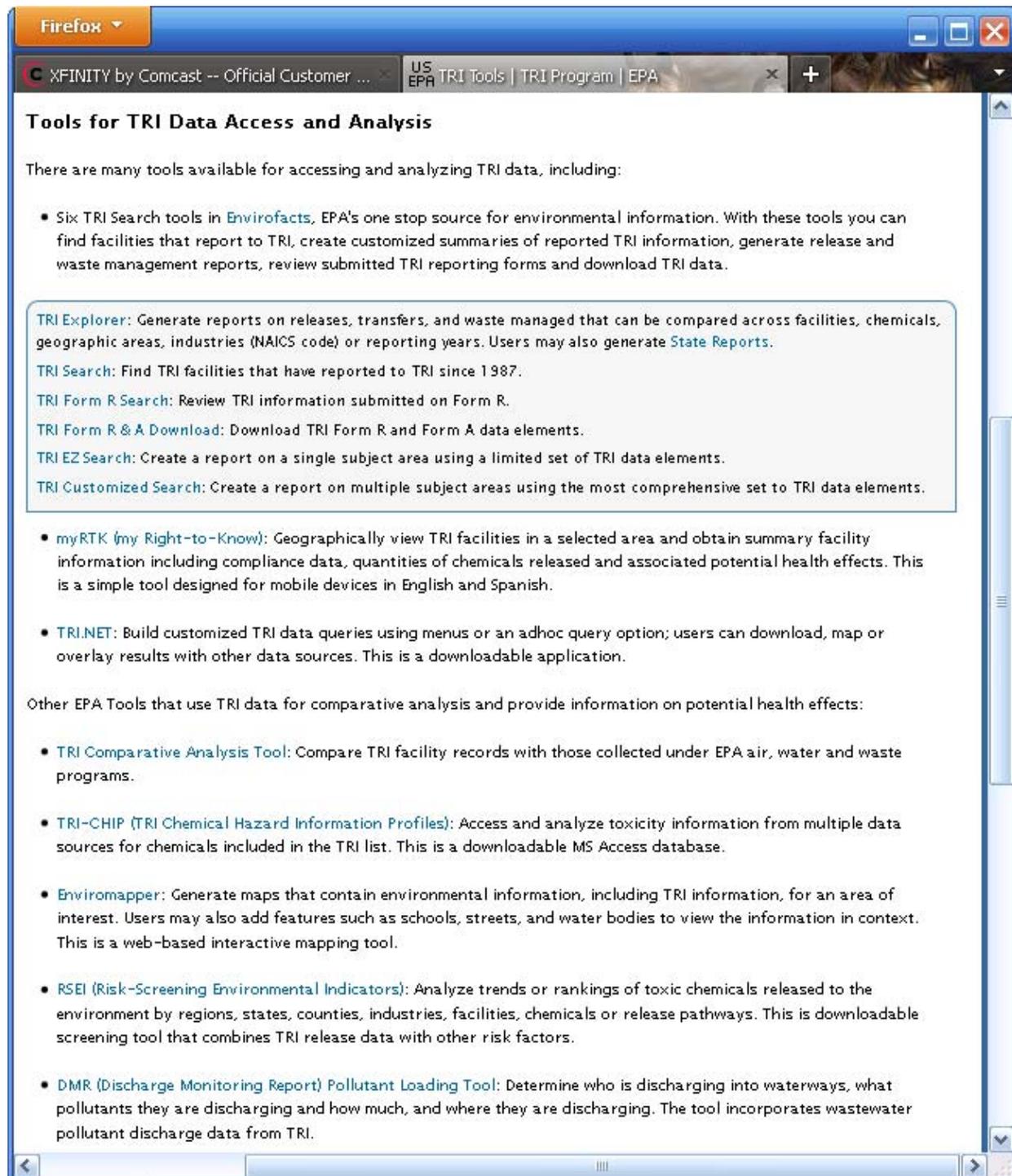


Figure 4. View of Envirofacts home page.



Figure 5. List of TRI data access and analysis tools.

View is portion of the webpage available from www.epa.gov/tri under the “TRI Data and Tools” link.



KEY WORDS

Toxics Release Inventory

TRI

Emergency Planning and Community Right-to-Know Act (EPCRA)

Pollution Prevention Act

TRI-ME

Discharge Monitoring Report (DMR)

MyRTK

TRI.NET

Envirofacts

TRI-CHIP

TRIBAL

Indian Land

TRI Comparative Analysis Tool

University engagement

Community engagement