Facilitating Chemical Risk Analysis by Organizing Toxicological Databases with a Flexible Open Platform

Conner Waybright, Brandon Ward, Ryan Brim, William Warner, Yok Lin Ong

ISS and the Department of Environment and Health

Istituto Superiore di Sanità (ISS), or the National Health Institute of Italy, is an organization whose mission is to guide public health policy based on scientific research. ISS conducts research in a wide variety of fields, such as food safety, epidemiology, and toxicology. ISS is divided into 6 Departments (shown below), 16 National Centers, 2 Reference Centers, 5 Technical-Scientific Services and a Notified Body. This project is concerned with the ISS Department of Environment and Health, or DAMSA. DAMSA conducts multidisciplinary research to identify, characterize, and analyze risk factors to environmental and social health.
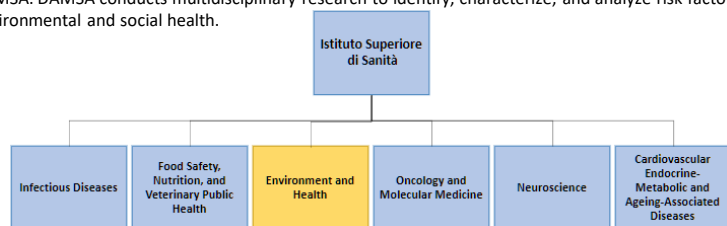
The Role of ISSTOX in Chemical Risk Analysis

ISS and DAMSA are interested in increasing the accessibility and interoperability of toxicological data. They began a project called ISSTOX to store quality toxicological data in five data sources available to the public. ISS asked our team to further develop this project by designing a system to store the ISSTOX data and facilitate public access to toxicity analysis.
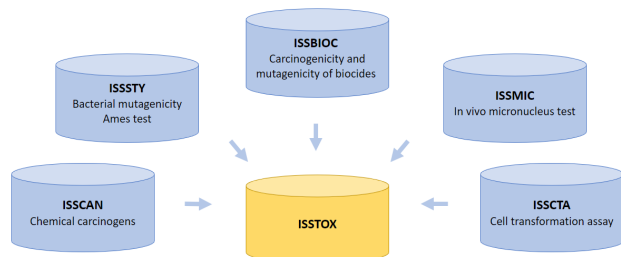
Our Role in ISSTOX Development

Our mission with ISSTOX is to design a user-interface and database system to increase public accessibility to toxicological data. Using stakeholder analysis, we identified all of the functions required in the minimum viable product. The system design required four subsystems: a database, a user-interface, a toxicity analysis component, and a chemical structure visualization component.

MVP Database Requirements	Weight
Store toxicity data in relational database	8.5%
Export data	7.5%
Standardize data headers	8.5%
Provide header definitions	2.5%

MVP Analysis Requirements	Weight
Calculate structural similarity	6.5%
Group chemicals	9.0%
Estimate chemical toxicity	5.0%
Export analysis reports	2.5%

MVP User Interface Requirements	Weight
Web accessible system	12.5%
User authentication	7.5%
Perform data functions without SQL	12.5%
View chemical data	7.5%

MVP Visualization Requirements	Weight
Display molecular structure	10%

Data Cleaning and Standardization

The original ISSTOX data had no standard vocabulary for data headers. Standardizing the data vocabulary increases interoperability by providing an internationally understood definition for each term. We standardized the data by identifying each attribute in the ISSTOX data and updating the attribute names to match the Organization for Economic Co-operation and Development's harmonized templates.

Many of the test result instances were stored as concatenated values in the ISSTOX data. This leads to an inefficient database design and makes querying on the data a difficult process. We addressed this problem by identifying and standardizing all existing attributes, and then by storing instances of test results in rows rather than columns.


Data Model Design

Addressing the database subsystem requirements necessitates the design of a single, chemical relational database to consolidate the five original ISSTOX data sources. The data design must be capable of storing all the original ISSTOX data with the ability to update and delete data in the future. The data modeling process included declaring data types, defining entities, and identifying cardinal relationships between entities.
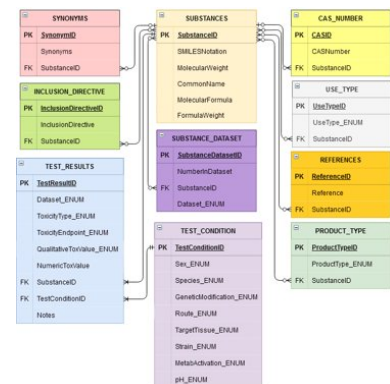

The User-Interface Design

The user-interface was designed in Adobe XD through an iterative feedback process with ISS. The Adobe XD design is an interactive mockup where every function of the minimum viable product is demonstrated, including the visualization and analysis subsystem requirements. This prototype was used to ensure all stakeholder requirements were met in the design phase of this project, and it will be used to guide the implementation phase in the future.
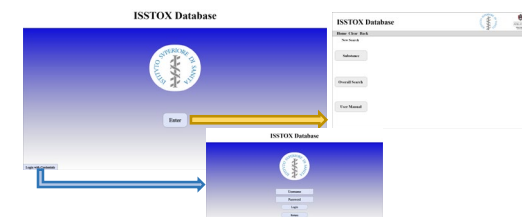
The focus of the user-interface design was to facilitate user-friendly, useful toxicity analysis on the ISSTOX data. The subsystem design allows users to create complex queries by using a series of buttons, search bars, and check boxes to select the information that should appear in the query results. This design allows public users with little to no coding or database experience to conduct useful toxicity analysis with relative ease.
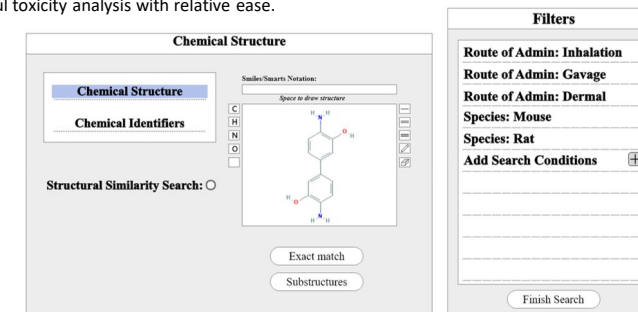
System Impact Analysis

Toxicity analysis efficiency of the new system design was tested against original analysis methods by conducting five different queries on each system. The metrics used to measure improvement were the number of clicks and the amount of time required to complete each search. The analysis revealed that the new system improves search efficiency by at least 65% and that new search features are now available that were not possible with the original ISSTOX system.