Developing a Consolidated Toxicological Database in Microsoft Access to Perform Chemical Risk Analysis and Predictions

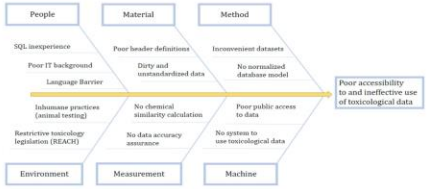
Quinn Salverson – Project Manager, Andrew Banks, Colin Connelly, Kaitlyn Frey, Isha Rajaram
Dr. Olga Tcheremenskaia - Principal Investigator, Dr. Chiara Battistelli - Principal Investigator,
Dr. Cecilia Bossa - Principal Investigator, Cristina Parenti – Sinergy Consulting, Professional Affiliate

ISS and the Department of Environment and Health

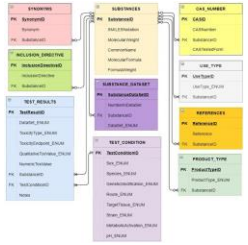
Istituto Superiore di Sanità (ISS) acts as the main center for research, control, and technical-scientific advice on public health in Italy. ISS is composed of six departments, but for this project we worked with the Department of Environment and Health. The Department of Environment and Health is focused on the prevention of diseases due to environmental factors such as air, soil, water quality, and human interaction with ecosystems.

2020-2021 ISS Capstone Team Achievements

The 2020-2021 ISS Capstone team created a fishbone diagram to determine the problems that led to the poor accessibility of the toxicological data.

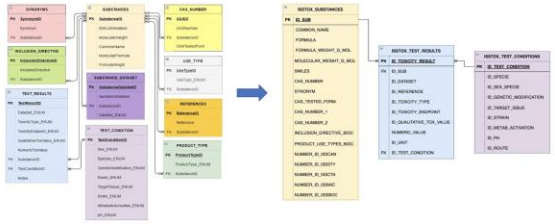


Last year's ISS Capstone team addressed some of these issues by designing a relational database and a corresponding user-interface system that would help to increase the accessibility to toxicological data.



Database Cleaning and Implementation

Our primary goal was to design and implement a streamlined version of Waybright et. al's ISSTOX relational database model. We designed a new database structure based on the recommendations of ISS.

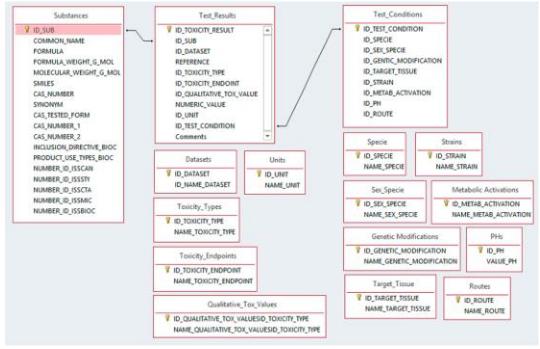


Our team was given five dirty and unstandardized datasets by ISS. The five datasets were cleaned using VBA code and Alteryx Software, and the OECD template was used to standardize the data.

TD50_Rat	TD50_Mouse	Canc	SAL	Rat_Male_Canc	Rat_Female_Canc	Mouse_Male_Canc	Mouse_Female_Canc
3.31	ND	3	3	3	3	3	3
NF	ND	1	3	ND	1	ND	ND

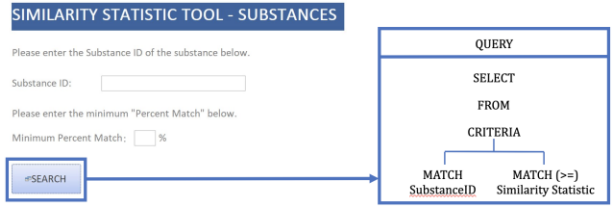
ID_TEST_RESULT	ID_SUB	ID_TOXICITY_TYPE	ID_TOXICITY_ENDPOINT	NUMERIC_VALUE	ID_UNIT	ID_TEST_CONDITION
1	1	CARCINOGENICITY	TD50	3.31	mg/kg body wt/da	1
2	1	CARCINOGENICITY	TD50	ND	mg/kg body wt/da	2
3	1	CARCINOGENICITY	CANS_OVERALL	3	mg/kg body wt/da	3
4	1	CARCINOGENICITY	SAL_OVERALL	3	mg/kg body wt/da	3
5	1	CARCINOGENICITY	CANS_SPECIE_SEX	3	mg/kg body wt/da	4
6	1	CARCINOGENICITY	CANS_SPECIE_SEX	3	mg/kg body wt/da	5
7	1	CARCINOGENICITY	CANS_SPECIE_SEX	3	mg/kg body wt/da	6
8	1	CARCINOGENICITY	CANS_SPECIE_SEX	3	mg/kg body wt/da	7

We consolidated all five datasets into one Excel File using OECD headers to transform the data into a relational database in Microsoft Access. We used VBA code to remove duplicates from the substance table when we were consolidating all the data from the datasets into a database.



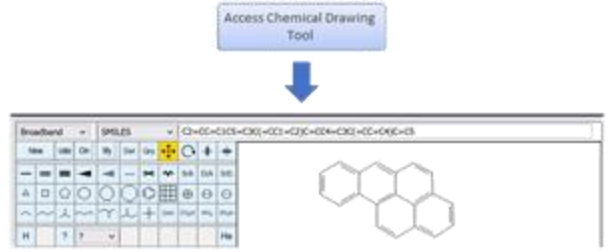
Analysis Tool

To allow the users to perform analysis with our database, we used the Alteryx Fuzzy Logic tool that leverages standard algorithms for pattern matching to calculate a similarity score between two substances. Once the similarity scores were calculated, we created a form and query to allow the user to specify the substance and percentage match in question.



Visualization Tool

We wanted to give users an ability to navigate the database by locating a chemical's specific SMILES code. We were able to connect an external search tool by programming a hyperlink to a button. The button takes users to the PubChem search tool.



We inserted pictures of chemical structures into the Access database for most of the chemical substances in the substances table. We converted CAS numbers into SMILES code. The SMILES codes were then converted to picture files that were uploaded to the access database.