Opportunity Title: CompTox Data Science at EPA
Opportunity Reference Code: EPA-NSSC-0009-34-12-7-21

Organization	U.S. Environmental Protection Agency (EPA)
Reference Code	EPA-NSSC-0009-34-12-7-21
How to Apply	Click HERE to Apply
Description	The EPA National Student Services Contract has an immediate opening for a full time ComTox Data Science position with the Office of Research and Development at the EPA facility in Research Triangle Park, NC.

The Office of Research and Development at the EPA supports high-quality research to improve the scientific basis for decisions on national environmental issues and help EPA achieve its environmental goals. Research is conducted in a broad range of environmental areas by scientists in EPA laboratories and at universities across the country.

What the EPA project is about

The Center for Computational Toxicology and Exposure (CCTE) supports ORD by providing solutions-driven research to rapidly evaluate the potential human health and environmental risks due to exposures to environmental stressors and ensure the integrity of the freshwater environment and its capacity to support human well-being. CCTE researchers are developing and applying cutting edge innovations in methods to rapidly evaluate chemical toxicity, transport, and exposure to people and environments. Within CCTE, the Chemical Characterization and Exposure Division (CCED) performs research to develop and advance analytical chemistry, computational chemistry, and cheminformatic approaches that are critical to the rapid characterization of the presence, structural characteristics, and properties of chemicals that underlie chemical exposure, environmental fate, toxicokinetics and toxicity

What experience and skills will you gain?

As a team member, you will support research under the Chemical Safety for Sustainability (CSS) research program on enhancing the Generalized Read-across (GenRA) workflow to address other similarity contexts beyond structural similarity.

Read-across is a data gap filling technique used within analogue and category approaches where information on a (source) substance is used to infer the same property for a similar (target) substance. Read-across approaches are traditionally expert driven subjective assessments and as such it is difficult to quantify their performance. The Generalized Read-across (GenRA) approach developed and published in 2016 (Shah et al., 2016) aimed to investigate new ways of performing objective read-across predictions from which a baseline in performance could be quantified. GenRA has since been implemented as a webtool GenRA within the EPA Chemicals CompTox Dashboard (Helman et al., 2019). This work has been an ongoing collaboration between 2 divisions within CCTE leveraging the advances in computational toxicological data such as high throughput screening, high throughput transcriptomic and high throughput phenotypic

profiling (so-called New Approach Methodologies (NAM) data) in conjunction with chemistry knowledge to make *in vivo* toxicity predictions. The current research focus on GenRA is to evaluate the contribution of NAM data to characterize different similarity contexts and their role in informing *in vivo* toxicity predictions.

The GenRA approach involves using a fingerprint representation of a substance (either chemical or NAM based) and using it to identify 'similar' substances with associated toxicity data to make a read-across prediction. You will further develop the GenRA approach to evaluate other NAM data and develop prototypes to implement the insights derived.

How you will apply your skills

> **Research-related responsibilities**, but are not limited to:

- Development of modelling data sets using both experimental and *in silico* data sources.

- This will require compiling new data extracted from the literature (e.g. following literature reviews), generating new data by running different *in silico* prediction tools and querying data stored in existing MySQL/Mongo databases. Note: Some MySQL databases are best queried using custom made R packages.

- Developing GenRA and other models to predict toxicity outcomes from chemical structure and NAM data.

  - Read-across analyses will be performed in Python using Jupyter notebooks

  - Other modelling approaches may include random forest (RF), support vector machines (SVM), k-nearest neighbour (kNN), artificial neural networks (ANN)

  - Read-across and other models will be written using existing GenRA code and other machine learning libraries in Python

- Applying other *in silico* tools e.g. OECD Toolbox which encodes (Q)SAR and structural alert information to generate relevant *in silico* predictions and molecular descriptor information.

- Respond to data requests from colleagues as needed (e.g. retrieve data according to specific criteria, or generation of *in silico* predictions) through database queries or running specific software.

- Development of prototype tools for data visualization and predictive models.
  Communications-related responsibilities:

  - Participate as a member of a multi-disciplinary research team;

  - Interact with other members of the development team as well as EPA scientists;

- Thoroughly document all work as directed by EPA mentor to comply with EPA quality assurance procedures for transparency and reproducibility of work; and

- Summarize work in internal reports/memos/presentations to be used by EPA scientists.

Required Knowledge, Skills, Work Experience, and Education

- Master's coursework in one or more of statistics, data science, machine learning, user experience and design.

- Experience with quantitative techniques, basic statistics, and use of spreadsheets.

- Proven proficiency in the Python language

- Experience with using the Linux shell.

- Experience in working with the distributed version control system git.

- Experience with MySQL and MongoDB

- Strong reading comprehension skills and experience logically interpreting pieces of data.

- Experience with computational or mathematical modeling and/or data science techniques in any discipline.

Desired Knowledge, Skills, Work Experience, and Education

- Basic experience with the R language is highly desirable.

- Experience in developing graphical user interfaces using JavaScript frameworks and using RESTful webservices

Location: This job will be located EPA's facility in Research Triangle Park, NC.

Salary: Selected applicant will become a temporary employee of ORAU and will receive an hourly wage of $30.16 for hours worked.

Hours: Full-time.

Travel: Travel related to the position is not anticipated.

Expected start date: The position is full time and expected to begin December 2021. The selected applicant will become a temporary employee of ORAU working as a contractor to EPA. The initial project is through May 14, 2022, with up to 3 additional option periods.

For more information, contact EPAjobs@orau.org. *Do not contact EPA directly.*

Qualifications
- Be at least 18 years of age **and**

- Have earned at least a Masters' degree in physics, chemistry, biology, engineering, applied sciences, environmental health, public health, exposure science, computer sciences, information technology, data science, or a related field of study from an accredited university or college within the last 24 months **and**

- Be a citizen of the United States of America or a Legal Permanent Resident.

EPA ORD employees, their spouses, and children are not eligible to participate in this program.

Eligibility Requirements

- **Citizenship:** LPR or U.S. Citizen
- **Degree:** Master's Degree received within the last 24 month(s).
- **Discipline(s):**
  - **Chemistry and Materials Sciences** (12 👁)
  - **Engineering** (27 👁)
  - **Life Health and Medical Sciences** (46 👁)
  - **Physics** (16 👁)

Affirmation

I certify that I am at least 18 years of age; a recent graduate with at least a Masters' degree in physics, chemistry, biology, engineering, applied sciences, environmental health, public health, exposure science, computer sciences, information technology, data science, or a related of study from an accredited university or college within the last 24 months; a citizen or a Legal Permanent Resident of the United States of America; and not a current employee of EPA ORD or the spouse or child of an EPA ORD employee.

Click HERE to Apply

ORAU is an Equal Opportunity Employer **(EOE AA M/F/Vet/Disability)**; visit the ORAU website for required employment notices.