Supporting source-to-outcome assessment in toxicology through data interoperability

David Hines[1], Shannon Bell[1], Patricia Ceger[1], Carol Hamilton[1], Nomi Harris[2], Sierra Moxon[2], Chris Mungall[2], Mark Nelms[1], James Rineer[1], David Williams[1]
[1]RTI International, Research Triangle Park, NC | [2]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract Number: 4384
Poster number: A137

Objectives

Identify EHS community needs

Expand standards & terminologies

Improve data interoperability

Background & Approach

- Environmental Health Science (EHS) encompasses multiple research domains along the source-to-outcome (S2O) continuum, each generating unique and nuanced data.
- We aim to: 1. Engage domain experts to **identify data standards and terminologies** (S&T) needs.
  2. **Improve data interoperability** by developing S&T to link data across domains.
  3. Leverage a pilot **use case** to test interoperability and highlight gaps in S&T.
- We link the Aggregate Exposure Pathway[1] (AEP) and Adverse Outcome Pathway[2] (AOP) mechanistic frameworks to represent the S2O continuum (Figure 1).

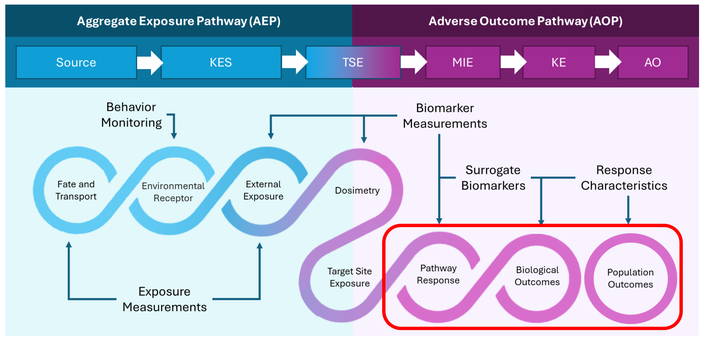


Figure 1: AEP-AOP, chain represents domains, text and arrows show measurements. Red box highlights domains that are the focus of this poster; KES (Key Exposure State), TSE (Target Site Exposure), MIE (Molecular Initiating Event), KE (Key Event), AO (Adverse Outcome).

Use Case

- We leverage a pilot use case from across the S2O continuum focused on PM2.5 inhalation associated with wildfire smoke resulting in decreased lung function (Figure 2) to:
  - Provide a focal study to guide expert discussion in **S&T prioritization** working groups
  - Facilitate quantitative data **integration** to test interoperability

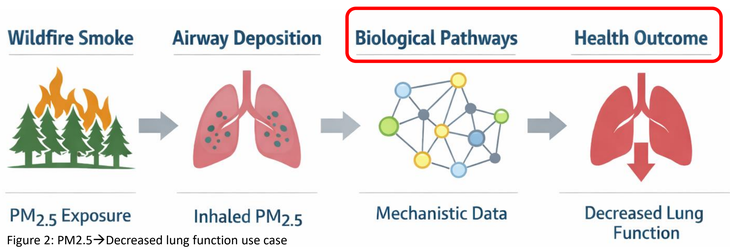


Figure 2: PM2.5→Decreased lung function use case

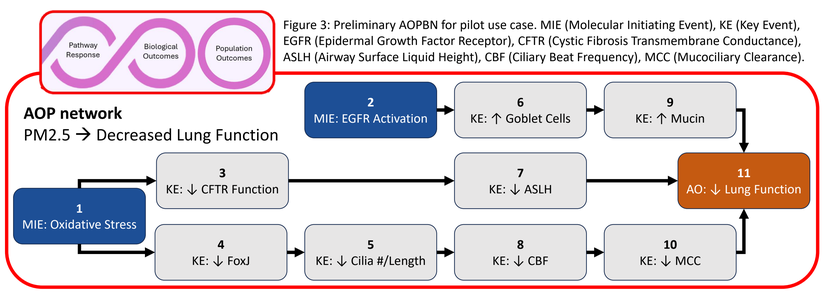


Figure 3: Preliminary AOPBN for pilot use case. MIE (Molecular Initiating Event), KE (Key Event), EGFR (Epidermal Growth Factor Receptor), CFTR (Cystic Fibrosis Transmembrane Conductance), ASLH (Airway Surface Liquid Height), CBF (Ciliary Beat Frequency), MCC (Mucociliary Clearance).

- We implemented an Adverse Outcome Pathway Bayesian Network[3,4] (AOPBN) model for the biological **pathway response/outcomes** portion of a use case (Figure 3).
- Within the AOPBN model:
  - Nodes are measurable KEs with defined states of activity.
  - Edges are conditional probability tables describing the relationships between KEs.

Community Engagement

- We assembled a pathway **response/outcomes working group** of researchers and stakeholders with subject mater expertise in respiratory distress endpoints along the use case AOP network.
- Over the course of six meetings, the working group:
  1. Reviewed relevant literature and extracted methodological text for mechanistic endpoints along the AOP network.
  2. Worked with our team members to identify crucial elements of mechanistic data that are necessary for secondary data use.
  3. Prioritized the identified elements using a score of 1 to 3 (Table 1).
  4. Iteratively advised our team on development of a data model schema.

Select examples of data elements / prioritization for ASLH key event	
Depth of periciliary layer in micrometers	1
Detailed cell culture conditions	1
Method used to prevent evaporation	2
Fluorescent label or tracer used	3
Ionic composition (Cl^-, Na^+, K^+)	3

Table 1: Example of select data elements from studies measuring ASLH and associated rank scores established from working group discussions; 1- critical for data interpretation and comparison, 2- important contextual information, 3- specialized or supplementary.

Semantic Development

- We developed a LinkML-based data model schema through an iterative process informed by working group feedback and domain expert consultation

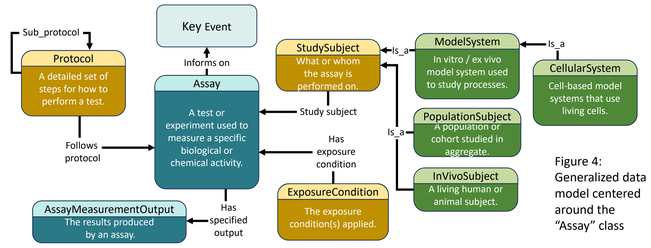


Figure 4: Generalized data model centered around the "Assay" class

- The schema design includes a generic, reusable framework centered around the **Assay** class (mapped to OBI:0000070)[5] that can be applied to domain-specific instantiations (Figure 4).
- The Assay class provides five core slots that capture the context of an experiment (Table 2):

Core slot	Description
informs_on_key_event	Links an assay to a specific **Key Event** in an AOP. • Connects measurement data to the mechanistic framework.
study_subject	Defines **what the assay is performed on**; accepts: • ModelSystem (in vitro) • InVivoSubject (human/animal with metadata) • PopulationSubject (cohort-level data)
has_exposure_condition	Describes the **treatment or stressor** applied. • Records exposure agent, concentration, duration, and timing • Supports multiple exposures (e.g., dose–response or co-exposure)
follows_protocols	References protocols **describing methodology**. • Accepts Multiple protocol types (e.g., imaging, staining) • May include sub-protocols
has_specified_output	References **measurement results** (AssayOutputMeasurement). • Separates results (output) from assay context (method and subject)

Table 2: components of the Assay class

- We created a companion class, AssayOutputMeasurement (mapped to IAO:0000109)[6] to provide the scaffolding for measurement results.
- Each domain-specific assay has a paired output class that extends AssayOutputMeasurement with typed measurement slots (Figure 5).

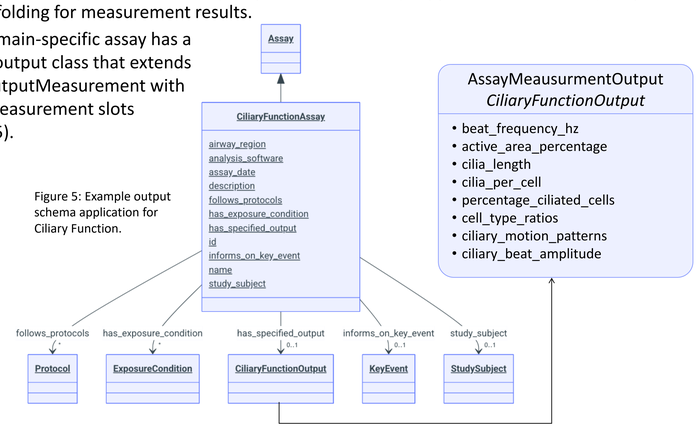
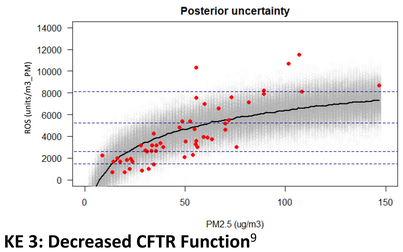


Figure 5: Example output schema application for Ciliary Function.

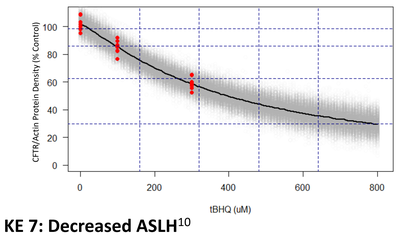
Use Case

- We constructed an AOPBN using literature data to build Bayesian regression models[7], then simulated data variability to generate conditional probability tables (Figure 6).

MIE 1: Oxidative stress[8]



KE 3: Decreased CFTR Function[9]



KE 7: Decreased ASLH[10]

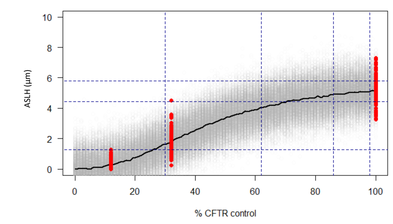
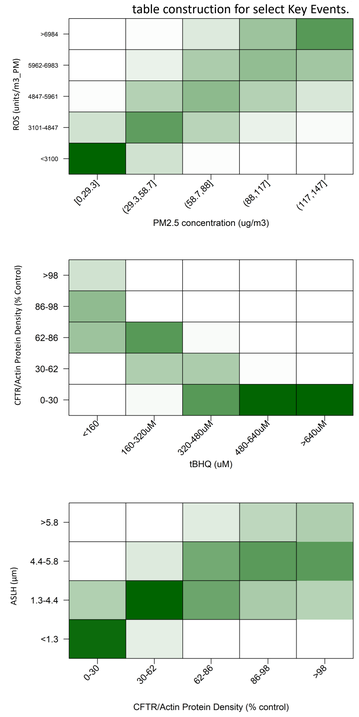
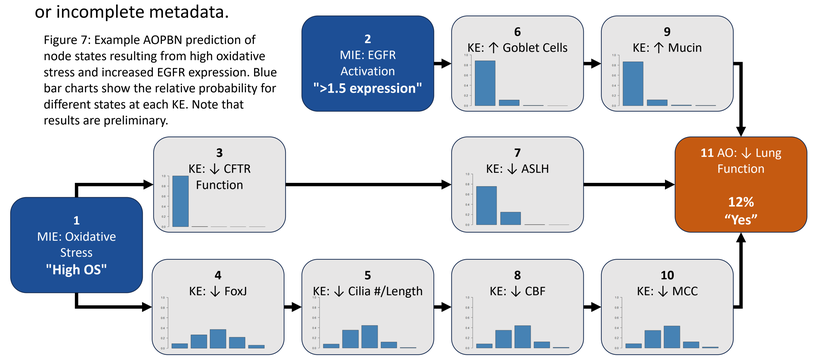


Figure 6: Example conditional probability table construction for select Key Events.



- The assembled AOPBN enables **probabilistic prediction** of KEs based on input states; in the severe example below, an individual would have a 12% chance of decreased lung function (Figure 7).
- Where necessary, we made and documented scaling **assumptions to align data sources** with missing or incomplete metadata.

Figure 7: Example AOPBN prediction of node states resulting from high oxidative stress and increased EGFR expression. Blue bar charts show the relative probability for different states at each KE. Note that results are preliminary.



Conclusions & Next Steps

- Our use case demonstrates how integration of disparate data sources across multiple domains is critical to next-generation, **quantitative applications in EHS**
- Approachable S&T promotes **FAIR** (Findable, Accessible, Interoperable, Reusable) principles, enhancing data **integration** and facilitating the use of emerging tools like **Artificial Intelligence**.
- We developed an **adaptable model schema**, available here: https://github.com/EHS-Data-Standards
- Next steps in this work will continue to expand standard schemas across the S2O continuum, focusing on **dosimetry** and **exposure**.

Get involved! Contact us at : DataStandards@rti.org
GitHub site : https://s2o-datastandards.github.io/

References:
1. Teeguarden, J.G., et al., 2016. Completing the link between exposure science and toxicology for improved environmental health decision making: the aggregate exposure pathway framework.
2. Villeneuve, D.L., Crump, D., Garcia-Reyero, N., et al., 2014. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicological sciences, 142(2), pp.312-320.
3. Burgoon, L.D., Angrish, M., Garcia-Reyero, N., et al., 2020. Predicting the probability that a chemical causes steatosis using adverse outcome pathway Bayesian networks (AOPBNs). Risk Analysis, 40(3), 512-523.
4. Kitson, N.K., Constantinou, A.C., Guo, Z., et al., 2023. A survey of Bayesian Network structure learning. Artificial Intelligence Review, 56(8), 8721-8814.
5. Ontology for Biomedical Investigations (OBI) https://obi-ontology.org/
6. Information Artifact Ontology (IAO) https://obofoundry.org/ontology/iao.html
7. Moe S.J., et al., 2021. Quantification of an Adverse Outcome Pathway Network by Bayesian Regression and Bayesian Network Modeling, Integrated Environmental Assessment and Management, 17(1) 147–164
8. Park, et al.. 2018. Reactive oxygen species (ROS) activity of ambient fine particles (PM2. 5) measured in Seoul, Korea. Environment international, 117, 276-283.
9. Cantin, A., et al. 2006. Oxidant stress suppresses CFTR expression. American Journal of Physiology-Cell Physiology 2006 290:1
10. Zhang L., et al. 2009. CFTR Delivery to 25% of Surface Epithelial Cells Restores Normal Rates of Mucus Transport to Human Cystic Fibrosis Airway Epithelium. PLOS Biology July 21, 2009.

This work is funded by the National Institute of Environmental Health Sciences under award number 1R24ES036127

For more Information, contact David Hines at dhines@rti.org