A mechanistic use case to support data interoperability across the source to outcome continuum in Environmental Health Science

David N. Williams[1], David E. Hines[1], Shannon Bell[1], Carol M. Hamilton[1], Chris Mungall[2], Sierra Moxon[2], James Rineer[1]

[1]RTI International, Research Triangle Park, NC | [2]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Objectives

Coordinate standards across EHS study areas

Expand standards & terminologies

Improve data interoperability

Goals

- Environmental Health Science (EHS) encompasses multiple research subdomains along the source-to-outcome (S2O) continuum, each generating unique and nuanced data.
- Integration of data from across these research domains strengthens the EHS conclusions by combining multiple lines of evidence, but data may be incompatible or difficult to access. (Figure 1).
- This work bolsters data **interoperability,** a component of the FAIR (Findable, Accessible, **Interoperable**, Reusable) principles, by developing data and metadata standards and terminologies (S&T) to link data and models across the S2O continuum.
- We develop a pilot use case to facilitate standards testing and development.

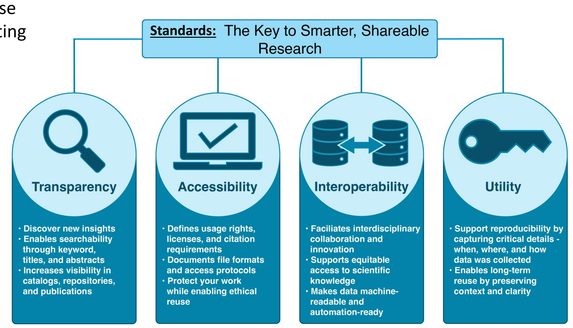

Figure 1: An overview of the ways that data and metadata standards can improve the quality of EHS research and support FAIR principles. We highlight how transparency (both visibility and documentation), accessibility (availability of data), interoperability (ability to appropriately facilitate data reuse), and utility (long term data preservation and value) are improved through the use of standards.

Approach

Source-to-Outcome Continuum

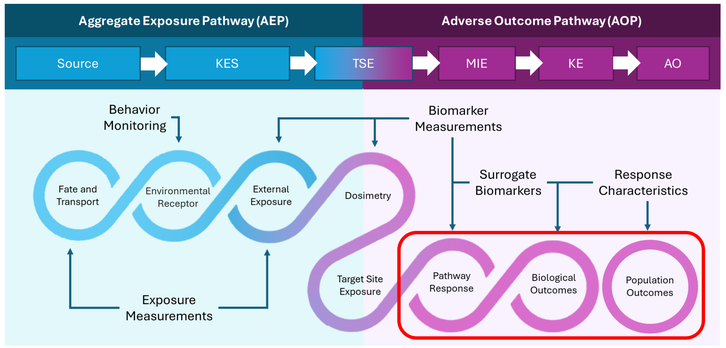

Figure 2: AEP-AOP, chain represents subdomains, while text and arrows show measurements that monitor them. Exposure subdomains (blue) and biological response subdomains (pink) are linked together through dosimetry (purple), red box highlights subdomains that are the focus of this poster; KES (Key Exposure State), TSE (Target Site Exposure), MIE (Molecular Initiating Event), KE (Key Event), AO (Adverse Outcome).

- We link the Aggregate Exposure Pathway[1] (AEP) and Adverse Outcome Pathway[2] (AOP) mechanistic frameworks to represent the S2O continuum (Figure 2).
- This work: 1) Engages domain experts to identify S&T needs
  2) Expands Biolink Model[3] to address needs
  3) Tests advancements with a **pilot use case**
- We approach the work in phases, focusing first on the biological outcomes in the AOP portion of the continuum.
- We use an Adverse Outcome Pathway Bayesian Network[4,5] (AOPBN) to represent the AOP for use in a pilot test model (Figure 3, yellow).

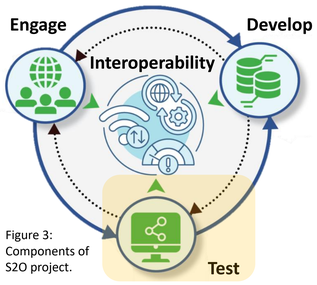

Figure 3: Components of S2O project.

Use Case

- We develop a **pilot use case** of $PM_{2.5}$ exposure leading to decreased lung function in humans as a quantitative test system for evaluating S&T.
- Here, we present the preliminary design and modeling approach for the Biological Outcomes component of the use case (Figure 4).

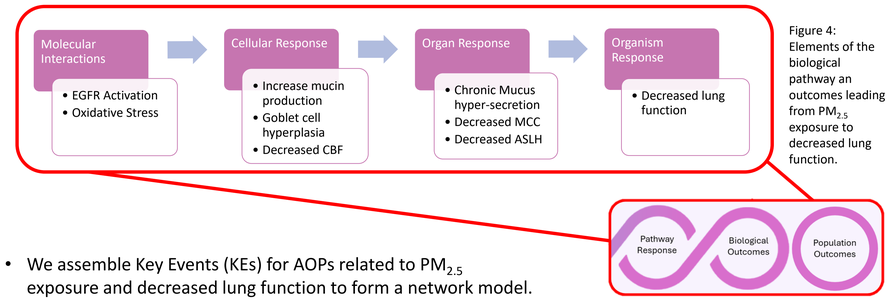

Figure 4: Elements of the biological pathway an outcomes leading from $PM_{2.5}$ exposure to decreased lung function.

- We assemble Key Events (KEs) for AOPs related to $PM_{2.5}$ exposure and decreased lung function to form a network model.
- A subject matter expert working group focused on guiding S&T development for Biological Outcomes is scheduled to begin meeting in Fall 2025.

AOP Bayesian Network (AOPBN)

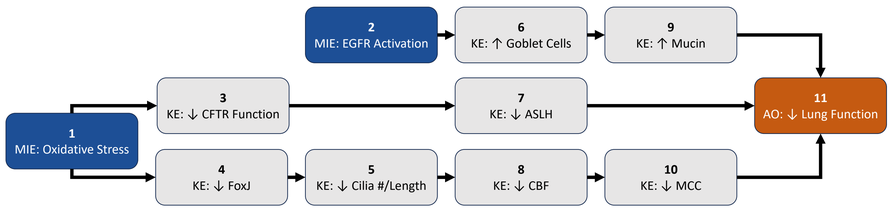

Figure 5: Preliminary AOPBN for decreased lung function used to test interoperability and evaluate the effects of developed S&T. Nodes are measurable KEs, edges are conditional probability tables describing the relationships between KEs. MIE (Molecular Initiating Event), KE (Key Event), EGFR (Epidermal Growth Factor Receptor), CFTR (Cystic Fibrosis Transmembrane Conductance), ASLH (Airway Surface Liquid Height), CBF (Ciliary Beat Frequency), MCC (Mucociliary Clearance).

- We assembled an AOPBN in the form of a Directed Acyclic Graph (DAG) where the relationships among nodes are described by conditional probability tables (Figure 5).
- The output of the AOPBN is a probabilistic prediction of the AO (decreased lung function), given the evidence for mechanistic KEs in the form of dose-response relationships.
- Conditional probability tables in the preliminary AOPBN are parameterized with hypothetical placeholder values, which will later be refined with literature data (Figure 6).
- While initial conditional probability tables use a binary "yes/no" response, higher resolution responses based on dose-response data and thresholds can be defined.

8 (↓CBF) \ 10 (↓MCC)	Y	N
Y	0.99	0.01
N	0.10	0.90

Figure 6: Example conditional probability table describing the relationship between node 8 (↓CBF) and 10 (↓MCC). Y (yes, active), N (no, inactive)

Model Needs

- Computational models can be used to build predictive relationships between nodes that can inform conditional probability tables.

Key Event	Computational Model	Model Inputs	Model Outputs
Oxidative Stress (ROS) [MIE]	Dose-Response, PM2.5 deposition to ROS	• Multiple Dose-response reference points • What are the measurements (direct vs proxy) • Impact of genetic variability or population	• Predicted ROS generation/ concentration
EGFR Activation [MIE]	Response-response, ROS concentration to EGFR expression levels	• ROS concentration • Genetic variability	• Predicted EGFR expression levels
Decreased FoxJ1	Response-response, ROS concentration to FoxJ1 levels	• ROS concentration • Genetic variability	• Predicted FoxJ1 protein levels
Decreased cilia number/length	Response-response, foxJ1 protein to ciliogenesis	• FoxJ1 protein levels	• Ciliogenesis gene expression levels • Ciliary length
Decreased Ciliary Beat Frequency (CBF)	Predictive, ciliary length to mucus and liquid layer	• Ciliary length	• Periciliary liquid layer • Mucus layer
Increased goblet cell production	Response-response, EGF to number of goblet cells	• EGF expression levels	• Predicted MUCs 5AC and 2 levels • Number of goblet cells
Decreased Mucociliary Clearance (MCC)	Response-response, CBF to Bronchial clearance	• CBF	• Predicted bronchial clearance
Increased Mucin production	Response-response, goblet cell expression levels to mucus amount	• MUCs 5AC and 2 levels	• Predicted mucus amount
Decreased lung function (AO)	Response-response, saccharin transit time to lung function	• Saccharin transit time • Mucin levels	• Individual/population forced expiratory volume estimates

Preliminary AOPBN Analyses

- Once conditional probability tables are established to describe relationships among nodes, the AOPBN can be queried to estimate the probability of other network events given the known data.
- We demonstrate this concept using three scenarios based on the preliminary hypothetical probability relationships.

Scenario 1: The oxidative stress node (1) is activated

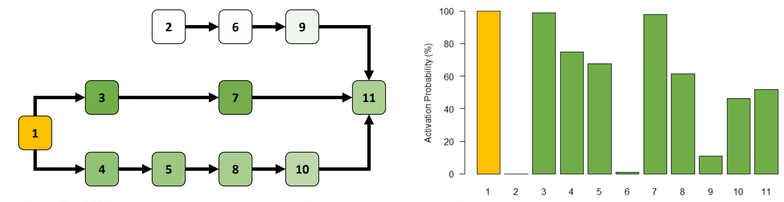

Figure 7: AOPBN response at the node level from a single input source of oxidative stress

Scenario 2: The EGFR Activation node (2) is activated

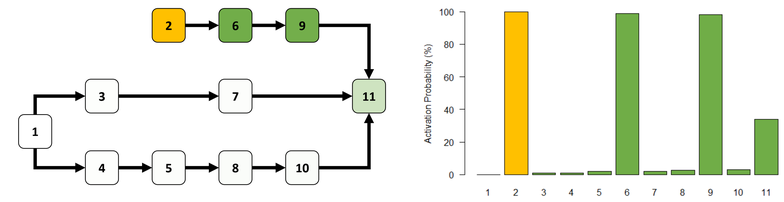

Figure 8: AOPBN response at the node level from a single input source of EGFR Activation

Scenario 3: Data show both decreased MCC (10) and increased Mucin Production (9)

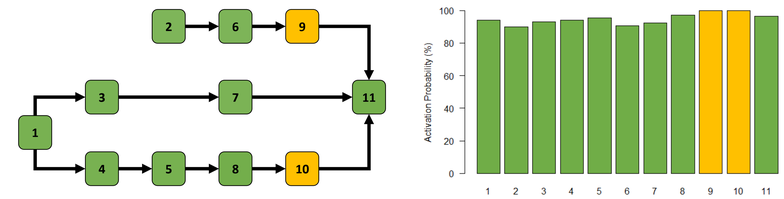

Figure 9: AOPBN response at the node level from a data that show both decreased mucociliary clearance and increased mucin production

- Model structure allows for incorporation of multiple data streams into analysis to test data integration advancements from S&T development.

Next Steps

- Leverage dose-response and response-response data to drive models to inform conditional probability relationships.
- Conduct sensitivity analyses to understand the role of each network component, identify critical data needs, and provide metrics of model performance.
- Identify S&T needs and data gaps, then provide guidance to the working group for the Biological Outcomes component of the S2O continuum.
- Document and test effects of S&T developments made in Biolink Model on AOPBN performance.

Future Integration

- The Biological Outcomes model will be linked to computational models representing environment fate and transport, exposure, and dosimetry to create a single pilot use case to integrate data spanning the entire S2O continuum (Figure 10).

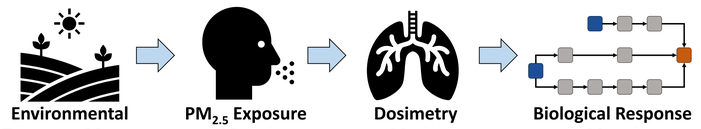

Figure 10: Multiple models representing the S2O continuum will be linked to form the pilot use case. The AOPBN developed in this work will be linked to dosimetry models, which in turn are linked to other models, facilitating data integration across the use case.

References:

1. Teeguarden, J.G., Tan, Y.M., Edwards, S.W., Leonard, J.A., 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. Unni, D.R., Moxon, S.A., Bada, M., et al., 2022. Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. Clinical and translational science, 15(8), pp.1848-1855.
4. 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), pp.512-523.
5. Kitson, N.K., Constantinou, A.C., Guo, Z., et al., 2023. A survey of Bayesian Network structure learning. Artificial Intelligence Review, 56(8), pp.8721-8814.

Project Site

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