Poster number:
PSII-09

Data Standards to Support Integrated Source to Outcome Modeling

David Hines[1], Shannon Bell[1], Carol Hamilton[1], Chris Mungall[2], James Rineer[1]
[1]Research Triangle Institute International; [2]Lawrence Berkeley National Laboratory

Objectives

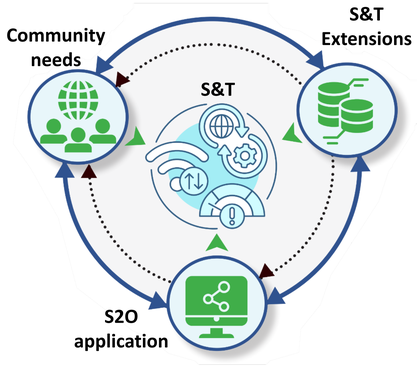
Coordinate standards across EHS study areas

Expand standards & terminologies

Improve data interoperability

Overview

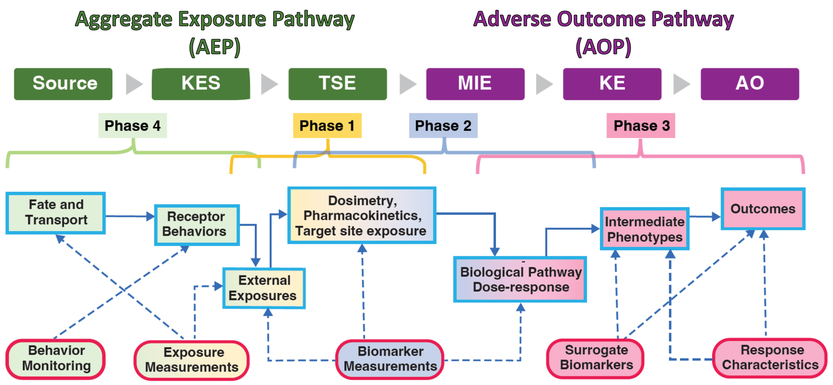
- Environmental Health Science (EHS) is extremely broad, spanning numerous subdomains.
- This project is a dedicated effort focused on coordination of standards and terminologies (S&T) across subdomains (e.g. chemical release, exposure, outcome) to support data **interoperability**.
- We aim to improve the precision of semantic descriptions to facilitate communication among humans and machines and strengthen predictive capabilities along the source-to-outcome (S2O) continuum.
- This work bolsters data interoperability, a component of the FAIR (Findable, Accessible, **Interoperable**, Reusable) principles, by:
  1. Engaging the expert and stakeholder communities to define standard and terminology (S&T).
  2. Expanding the Biolink Model to better describe chemical fate, exposure events, and biomarkers within environmental contexts.
  3. Establishing a functional workflow using a case study as a test system.



Introduction

- Understanding how exposures impact human health requires data to be interoperable, machine accessible, and understandable across a wide range of disciplines, but data are rarely reported for broader applications.
- Data and metadata S&T are essential for
  1. ensuring that data collected are interoperable across subdomains
  2. facilitating machine readability to promote broader applications
  3. Supporting integrated modeling of exposure and response.
- We will leverage the **aggregate exposure pathway** (AEP) and **adverse outcome pathway** (AOP) frameworks, which span multiple subdomains, to describe S2O continuum.
- This work will engage stakeholders from different subdomains to discuss interoperability challenges and needs, focusing on a few subdomains at a time in four phases.

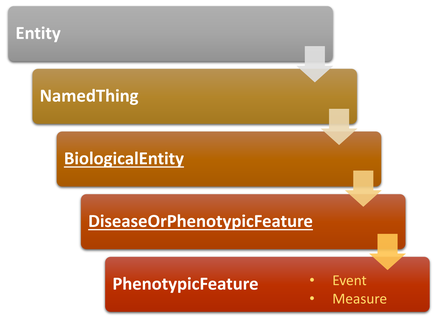
Phase 1 – Exposure and dosimetry
Phase 2 – Dose response & pathway perturbations
Phase 3 – Phenotypes and adverse outcomes
Phase 4 – Exposure pathways



Here you see the breakdown of subdomains across these frameworks. Solid boxes and arrows show subdomains and how they are related, while dotted boxes and arrows show measures that can be used to monitor those subdomains. The colors in this figure correspond to the phases of the work.

Semantic development

- The **Biolink Model** is an open-source standard for harmonization across biomedical knowledge graphs that includes terms and relationships useful for computationally representing mechanistic pathways.
- The **Biolink Model** requires additional model classes to better describe the components and interactions of external biological responses following exposure.
- This work will improve S&T across S2O subdomains in this model by
  1. Extending coverage, including for exposure concepts
  2. Better distinguishing biomarker measurements from the biological events they represent.
- We will make use existing ontologies to expand the **Biolink Model** by mapping them to standards and/or additional classes, with emphasis on extending capabilities for fate and transport and exposure events; candidate ontologies are show in the table to the right.
- Ontologies from the Open Biological and Biomedical Ontology (OBO) Foundry are preferred because they are logically consistent and already benefit from a high degree of coordination among the ontology developers
- This project will also leverage the ontologies, tools, and resources from the Environmental Health Language Collaborative (EHLC).



Examples of ontologies for extending Biolink Model. Ontologies are listed down the left column. The top row represents types of entities involved in an AEP/AOP (Exhibit 1). An X in a cell indicates coverage of the entity type by an ontology.

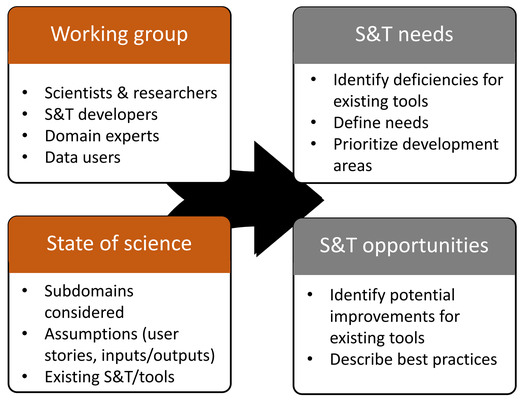
Entity Type	EM	EX	IX	TX	ME	CE	TE	OE	AO	PO
Physical & Biological Information Ontologies										
ENVO[1]	X	X								
EXO[2]		X	X	X						
ECTO[3]	X	X	X	X						
CHEBI[4]				X	X					
PRO[5]					X					
GO[6]					X	X				
CL[7]						X				
UBERON[8]							X	X		
MP[9]							X	X	X	
MonDO[10]									X	
PCO[11]										X
Measurement Information Ontologies										
HHEAR[12]		X							X	X
MI[13]					X					
BAO[14]					X	X	X			
OBI[15]					X	X	X	X		
MedDRA[16]									X	
NCIT[17]		X	X	X	X	X	X	X	X	X

EM – Exposure Medium; EX – External exposure; IX – Internal exposure; TX – Target site exposure; ME – Molecular initiating event; CE – Cellular event; TE – Tissue event; OE – Organ event; AO – Individual adverse outcome; PO – Population adverse outcome

- We will clarify distinctions between biomarkers measurements and biological events to better facilitate the use of multiple orthogonal measurements to monitor a single event.
- Example: in the *PhenotypicFeature* class, we will distinguish between decreased liver function, which is a system change, as well as alanine aminotransferase level, which is a specific measurement of decreased liver function.

Understanding needs

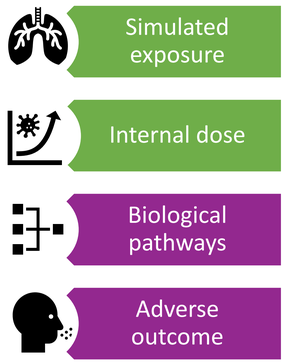
- Increased adoption and harmonization of data and metadata S&T across subdomains can address barriers linking EHS data along the S2O continuum.
- We will aim to identify
  1. Key needs to enable cross-domain data sharing
  2. Challenges and opportunities related to data sharing
  3. Domain-specific priorities for S&T development.
- **Working groups** of experts and stakeholders across EHS subdomains will be assembled at each phase of the project.
- **Working groups** will convene for a series of mini workshops to build consensus around a set of common S&T connecting their subdomains.

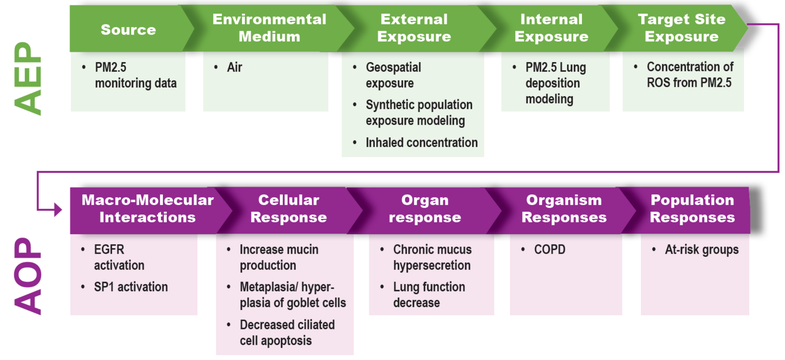


- Input from **working groups** will be essential ensure S&T meet the needs of the community and to promote their adoption and use.

Test system

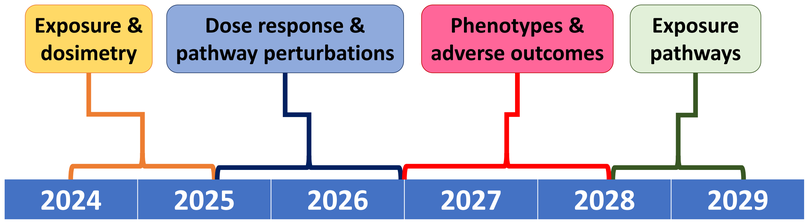
- We will construct a quantitative **test system** centered around a use case for $PM_{2.5}$ and chronic obstructive pulmonary disease (COPD) to
  1. Evaluate functionality of data standards
  2. Identify gaps that prevent interoperability
  3. Provide guidance to continued S&T development.
- This use case was selected because there are well-defined exposure mechanisms, biological pathways, and outcomes.
- $PM_{2.5}$ exposure results in Reactive Oxygen species (ROS), which can activate the epidermal growth factor receptor (EGFR) leading to mucus hypersecretion and decreased lung function.
- We will build the test system by assembling smaller interoperable model units, guided by an AEP-AOP construct.
- Synthetic population modeling will act as a "scrambled census" to drive test system exposures with realistic heterogeneity while protecting private information.





Community engagement

- Work is targeted across defined phases



- The utility of EHS S&T depends on how well they match the terminology and conventions of the subdomains they cover.
- We will work with **subject matter experts** (SMEs) within each subdomain to develop S&T that meet the current needs of the field.
- You can get involved! Participate in a working group or contribute domain expertise!

Contact us at : DataStandards@rti.org

References

1. OBO Foundry. Environment Ontology. https://obofoundry.org/ontology/envo.html
2. OBO Foundry. Exposure ontology. https://obofoundry.org/ontology/exo.html
3. OBO Foundry. Environmental conditions, treatments and exposures ontology. https://obofoundry.org/ontology/envo.html
4. OBO Foundry. Chemical Entities of Biological Interest. https://obofoundry.org/ontology/chebi.html
5. OBO Foundry. PRotein Ontology (PRO). https://obofoundry.org/ontology/pr.html
6. OBO Foundry. Gene Ontology. https://obofoundry.org/ontology/go.html
7. OBO Foundry. Cell Ontology. https://obofoundry.org/ontology/envo.html
8. OBO Foundry. Uberon multi-species anatomy ontology. https://obofoundry.org/ontology/uberon.html
9. OBO Foundry. Mammalian Phenotype Ontology. https://obofoundry.org/ontology/mp.html
10. OBO Foundry. Mondo Disease Ontology. https://obofoundry.org/ontology/mondo.html
11. OBO Foundry. Population and Community Ontology. https://obofoundry.org/ontology/pco.html
12. HHEAR. Publications. https://hheardatacenter.mssm.edu/Resource/Get
13. OBO Foundry. Molecular Interactions Controlled Vocabulary. https://obofoundry.org/ontology/mi.html
14. BioAssayOntology. http://bioassayontology.org/
15. OBO Foundry. Ontology for Biomedical Investigations. https://obofoundry.org/ontology/obi.html
16. MeDRA. Medical dictionary for regulatory activities. https://www.meddra.org/
17. OBO Foundry. NCI Thesaurus OBO Edition. https://obofoundry.org/ontology/cl.html

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**