An Informatics Architecture for an Exposome

Session II06 – Secondary Use of Data for Research (Interactive Learning)
AMIA 2016 Joint Summits on Translational Science
March 22nd, 2016

Dr. Ram Gouripeddi
Assistant Professor, Department of Biomedical Informatics
Chief Biomedical Informaticist, Biomedical Informatics Core,
Center for Clinical and Translational Science
University of Utah

Acknowledgements

- PRISMS PI(s)
  - Katherine Sward, RN, PhD
  - Julio C. Facelli, PhD
- PRISMS Team at University of Utah
- No disclosures

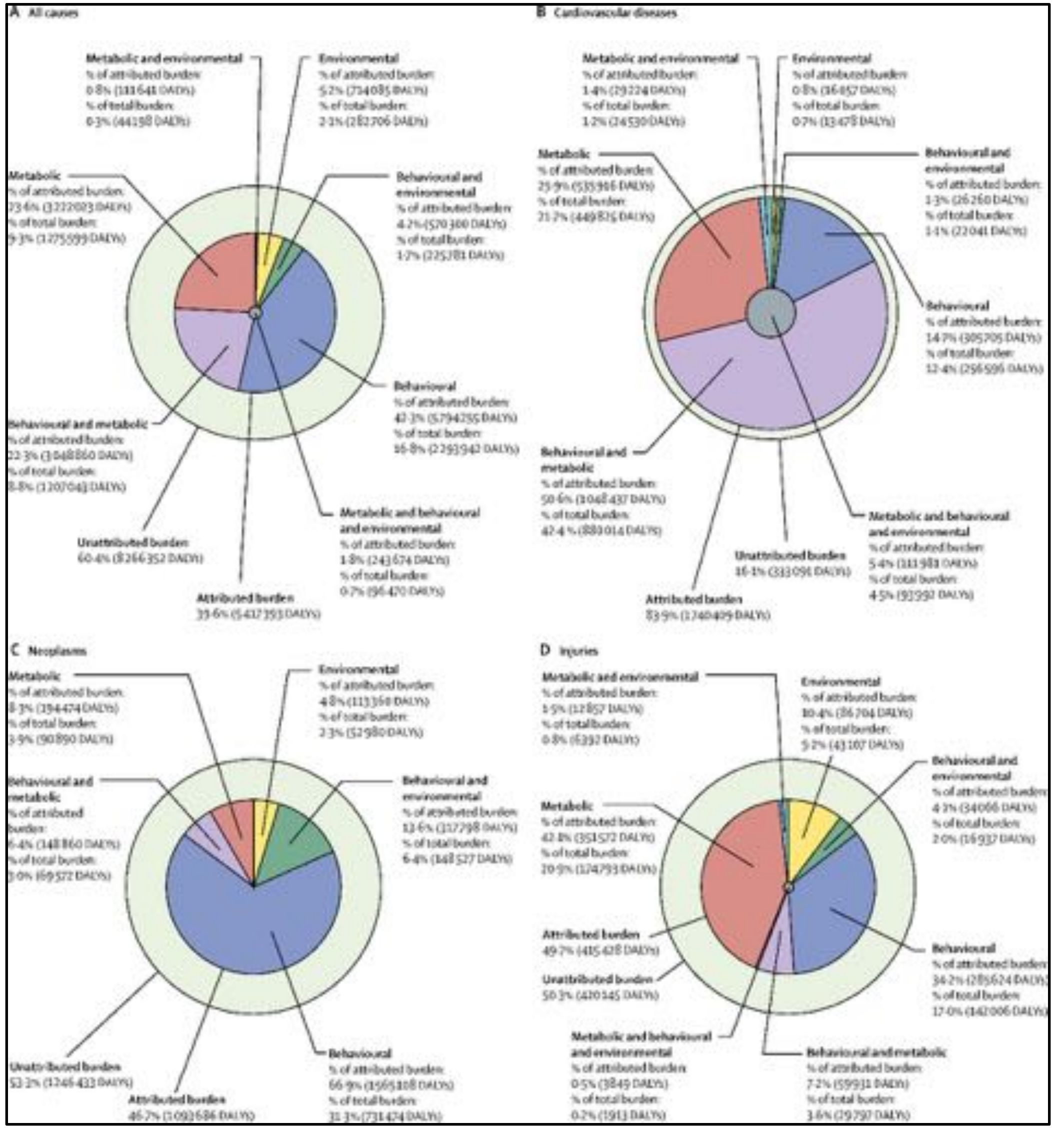
This effort was supported by U54EB021973, National Institute of Biomedical Imaging and Bioengineering, NIH and University of Utah Air Quality Seed Grant Program. OpenFurther supported by NCRR/ NCATS Grants UL1RR025764 and 3UL1RR025764-02S2, National Center for Clinical and Translational Science 1UL1TR001067, University of Utah Research Foundation, grant 1D1BRH20425 (DHHS), and R01 HS019862 from AHRQ, (DHHS).

Overview

- Effects of Environment on Health
- Key Concepts
- Initial Work – AMIA 2014
- Limitations
- Challenges and Informatics Methods and Solutions
- PRISMS
- Informatics Architecture

Effects of Environment on Health

- Phenotype: Result of interactions between genotype and environment.
- Environmental factors contribute significantly by themselves and their interaction[1] with behavioral, occupational and metabolic factors[1].



Disability-adjusted life-years attributable to behavioral, environmental, occupational, and metabolic risk factors[1].

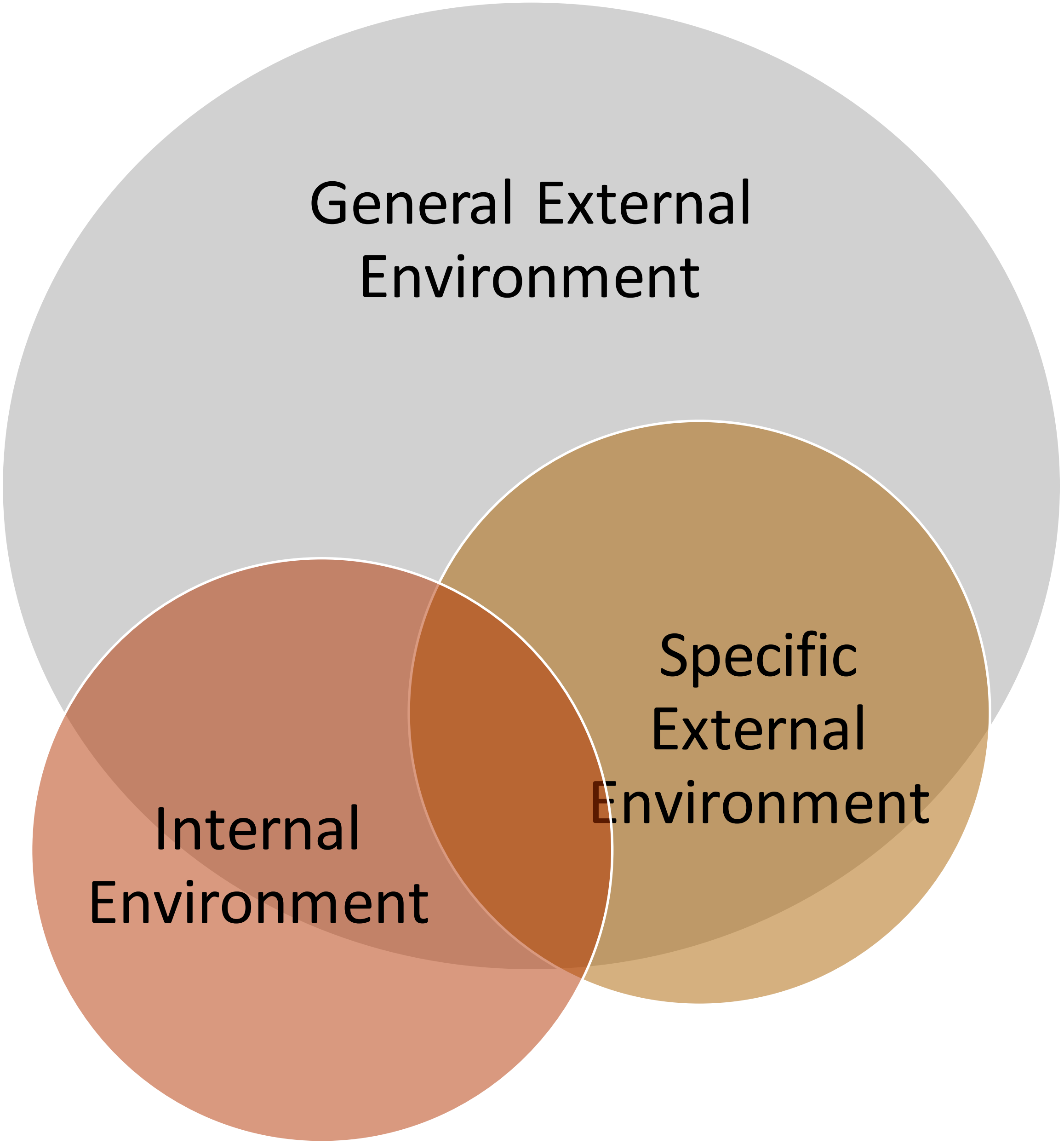
Flint Water Crisis

- Lead Poisoning in kids
  - Immune disorders
  - Criminal tendencies
  - Behavior and learning problems
  - Lower IQ and hyperactivity
  - Slowed growth
  - Hearing problems
  - Anemia
- No known safe level of lead in a child's blood.
- Lead Action Level: 10% of drinking water > 10 parts per billion.
- CDC's public health actions: when the level of lead in a child's blood ≥ 5 micrograms per deciliter[2].

http://electrochemistryresources.com/wp-content/uploads/2016/02/corrosion-water-pipe.jpg

Exposome[3-6]

- Encompasses life-course of environmental exposures (including lifestyle factors) from prenatal period onwards.
- Complements genome by providing a comprehensive description of lifelong exposure history.



Overlapping domains within exposome

Exposomics

- Study of defining, generating and utilizing exposomes in biomedical research.
- Ongoing efforts:
  - HELIX[7]: Early life exposome
  - EXPOsOMICS[8]: Assess exposures
  - HEALS[9]: Studies exposure to environmental stressors and health outcomes
  - NIH's Environmental influences on Child Health Outcomes (ECHO) Program[10]: Understanding the effects of environmental exposures on child health and development
- Requires a systems biology approach.
- *'Expotying':* Exposure of a biological entity usually with reference to a specific characteristic under consideration.
- Also called as Exposome Informatics, Exposure Information Science.
- Provides great opportunities to Biomedical Informatics[11].

Defining and Generating an Air Quality Exposome

Background

- Air Quality (AQ) has been associated with various adverse health effects
  - Asthma
  - Cardiovascular disease
  - Respiratory infections
  - Cancers
  - Impaired glucose tolerance during pregnancies[12-15].
- Researchers at the University of Utah are embarking on clinical studies to understand associations between the peculiar AQ patterns in Salt Lake City and clinical conditions:
  - Cerebral venous thrombosis
  - Exacerbations of idiopathic pulmonary fibrosis
  - Suicide
  - Reproductive outcomes
  - Cancers.

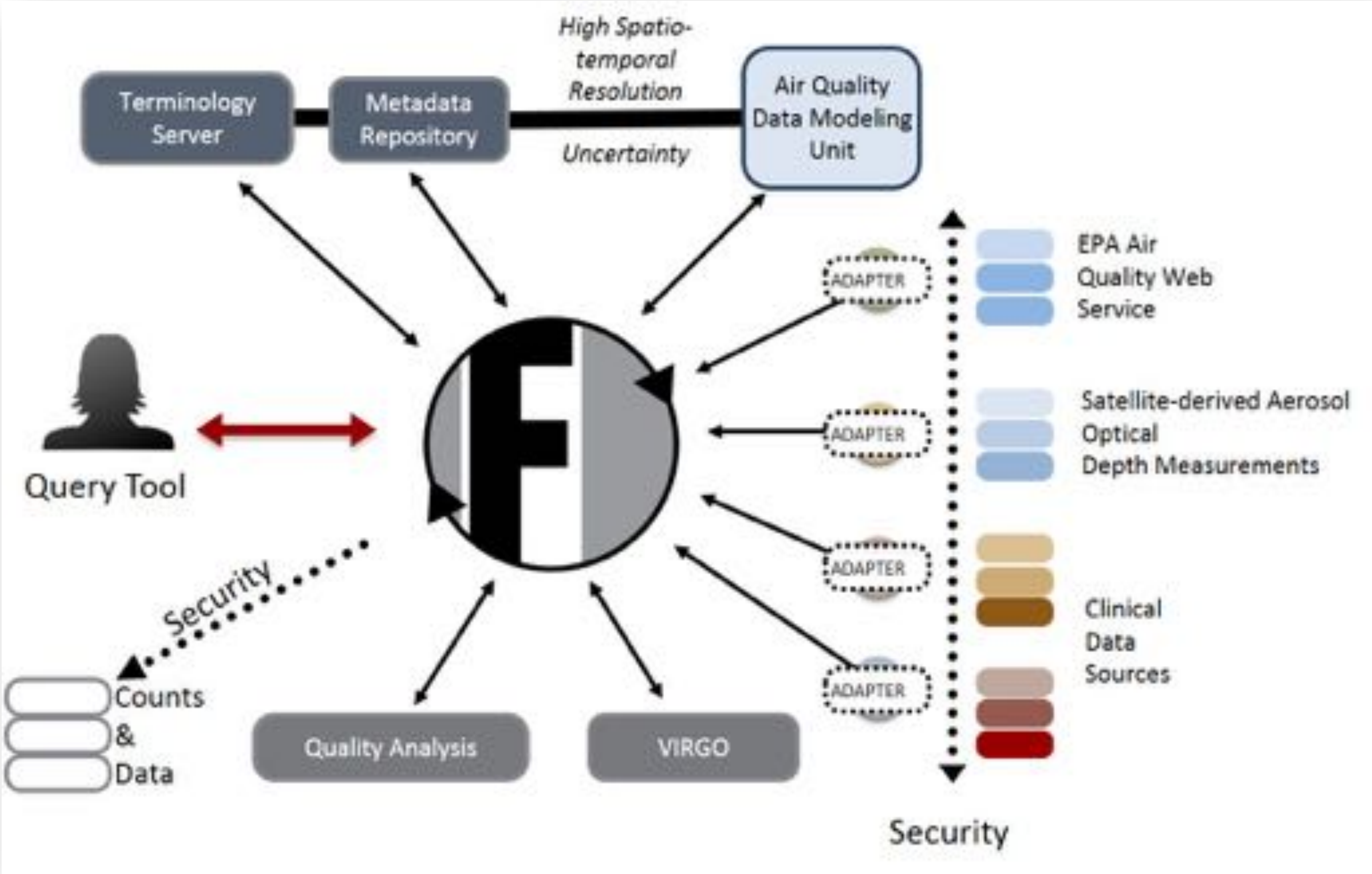
Salt Lake City Air Quality

Courtesy: Dr. K. Kelly

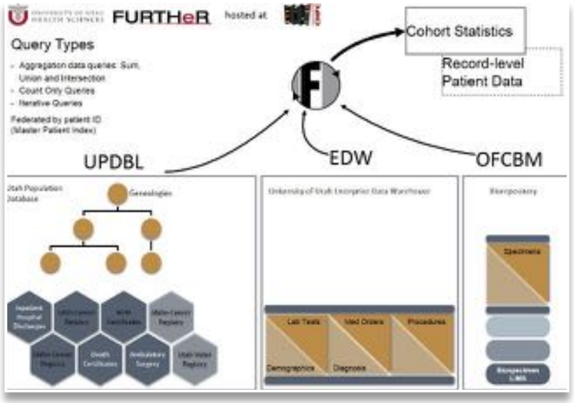
- Prone to winter inversions where colder surface temperatures trap fine particulate matter ($PM_{2.5}$) which poses serious health concerns.
- Summer months in the valley have increased ozone (O^3) levels[16].
- Natural/Quasi-experimental conditions.

OpenFurther[17-18]

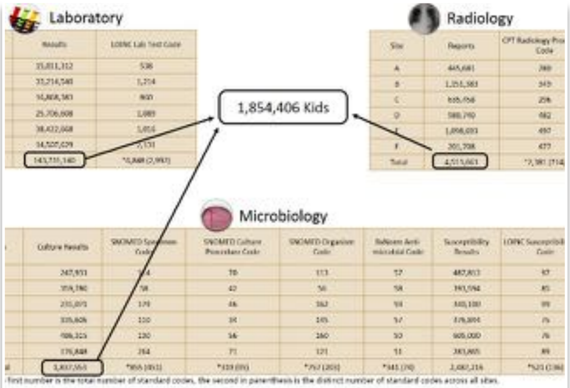
- Query Tool
- Federated Query Engine
- Data Source Adapters
- Admin & Security Components
- Virtual Identity Resolution on the GO (VIRGO)
- Quality & Analytics Framework
- Metadata Repository
- Terminology/Ontology Server
- Air Quality Modelling Unit



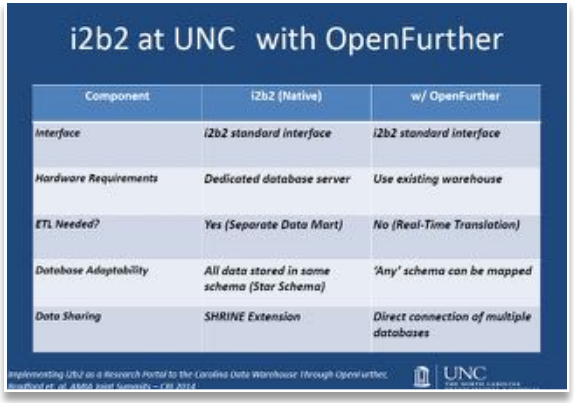
OpenFurther Deployments and Uses



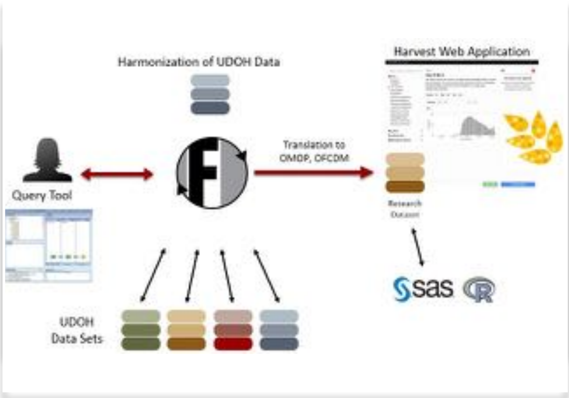
Cohort Selection, University of Utah



Comparative Effectiveness Research, PHIS+



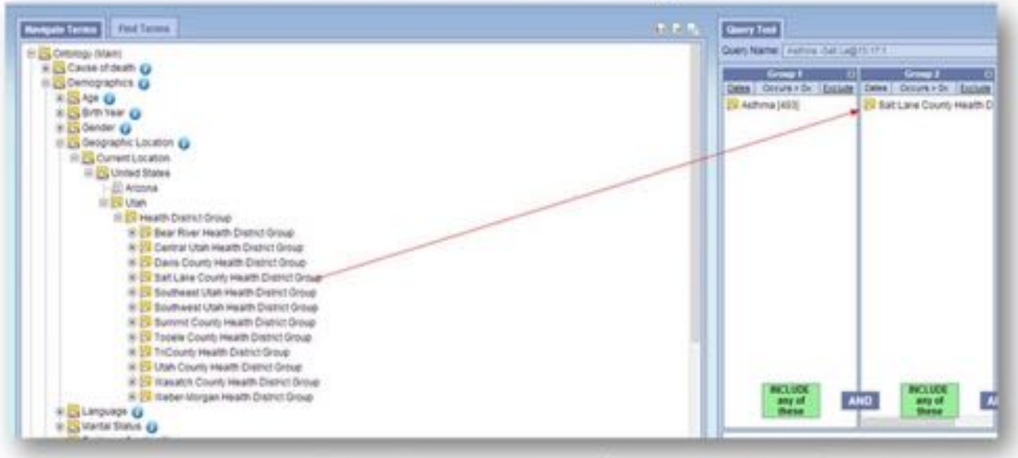
Cohort Selection, University of North Carolina



Data Integration & Analytics Pipeline, Utah Department of Health

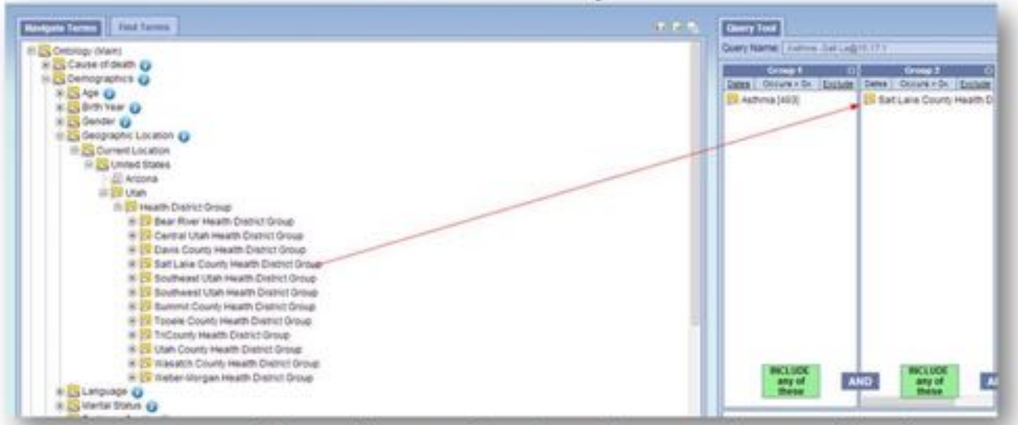
Air Quality - Clinical Data Federation

Asthma in January 2014



615 patients with a diagnosis of asthma in Salt Lake County and average $PM_{2.5}$ 28 micrograms

Asthma in January 20th 2014



25 patients with a diagnosis of asthma who reside in Salt Lake County and average $PM_{2.5}$ 50 micrograms

Worst Inversion Day

- Demonstrated feasibility of federating air quality data from Environmental Protection Agency (EPA) with clinical data from University of Utah using OpenFurther[19-20].
- Ability to select different cohorts of patients living in SLC county and having clinical conditions (e.g. asthma) occurrences that were related to temporal variations of air pollutant concentration.

Air Quality Monitoring in Salt Lake County

- Three monitoring stations in Salt Lake County.
- AQ species concentration variations due to topography, altitude and meteorology[21-22]
- What is the air quality at any other location?
- Need for cross-linking patient locations and condition occurrences: <u>High Resolution Spatio-temporal Air Quality Grid</u>

Limitations

Air Quality Exposome

$\int$

Pollutant & Quantity

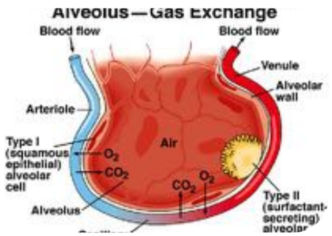
Travel

Home

Ventilation

Outdoors

Clinical Conditions



Biological Membranes

Others

Air Quality **Socio-economic** **Behavioral** **Clinical/Physiological** **Genomic** **Proteomic**

Biomedical Research Air Quality Requirements

- Primary need: understand risks associated with being exposed with various air pollutants.
- Manifestations following exposure could occur
  - Immediately
  - After a lag phase
  - Could persist over long durations.
- Need for understanding pathophysiology and mechanisms of these manifestations.
- Current research mainly associates single pollutant and clinical conditions, future areas of research could include exposures to multiple pollutants.

Utilizing Air Quality Data in Biomedical Research

- Integrating AQ and biomedical data needs to support
  - Spatio-temporal variations of air pollutant species.
  - Heterogeneous data.
  - Location of individuals.
  - Timing of the occurrence of events.
- AQ data and research requirement granularities vary from instantaneous to longer duration averages depending.
- Simplification of understanding and integrating AQ data with biomedical data.
- Support bench, translational, clinical and population research.

Challenges and Informatics Methods and Solutions

- Data Sources
- Mathematical Modeling
- Uncertainty Characterization
- Data Integration
  - Semantics
  - Metadata
  - Time & event modeling
  - Infrastructure for multi-scale, multi-omics integration
- Presentation/Visualization
  - Salient feature extraction

University of Utah’s PRISMS Informatics Infrastructure

Pediatric Research using Integrated Sensor Monitoring Systems (PRISMS)

- Sensor-based, integrated health monitoring systems for measuring environmental, physiological, and behavioral factors in pediatric epidemiological studies of asthma, and eventually other chronic diseases[23].
  - Sensor Development Projects
  - Informatics Platform Technologies
  - Data and Software Coordination and Integration Center
- Utah Team: Electric Engineering, Chemical Engineering, Computer Science, Atmospheric Sciences, Industrial Engineering, Informatics, Software Developers, Nursing, Pediatrics.

Air Quality Data Sources

Different air quality species

- Particular Matter: $PM_{2.5}$, PM_{10}, UPF
- Ozone
- Carbon Monoxide
- NO_x (nitric oxide and nitrogen dioxide)
- Sulphur Dioxide
- Lead
- Water Vapor
- Carbon Dioxide
- Volatile Organic Compounds

Choice of selectable sources for each species

High resolution spatio-temporal AQ grid

- Personalization

Types of Air Quality Sources

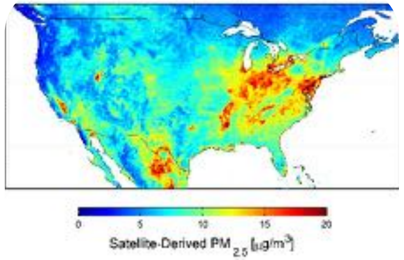
Personal Sensors

Laser Ceilometers

Novel Sensors

Mobile Sensors

Balloon Sensors



Satellite-derived aerosol optical depth measurements

State Environmental Department Networks

Environmental Protection Agency

Air Quality Mathematical Models

- Fill gaps in measured data with mathematical models.
- A library of AQ data models to provide high spatio-temporal resolution with a framework validate the model output.

Environmental Protection Agency – Center for Disease Control Model[24]

- Validated on the east coast
- Doesn't consider Altitude
- 12 kilometer resolution
- Hierarchical Bayesian model

Generalized Additive Mixed Models[25]

- Describe regional and small-scale spatial and temporal gradients
- Uses measured PM concentrations, monitoring site location, GIS-based location-specific characteristics and location-and month-specific meteorological data, and spatial smoothing of monthly and long-term averages

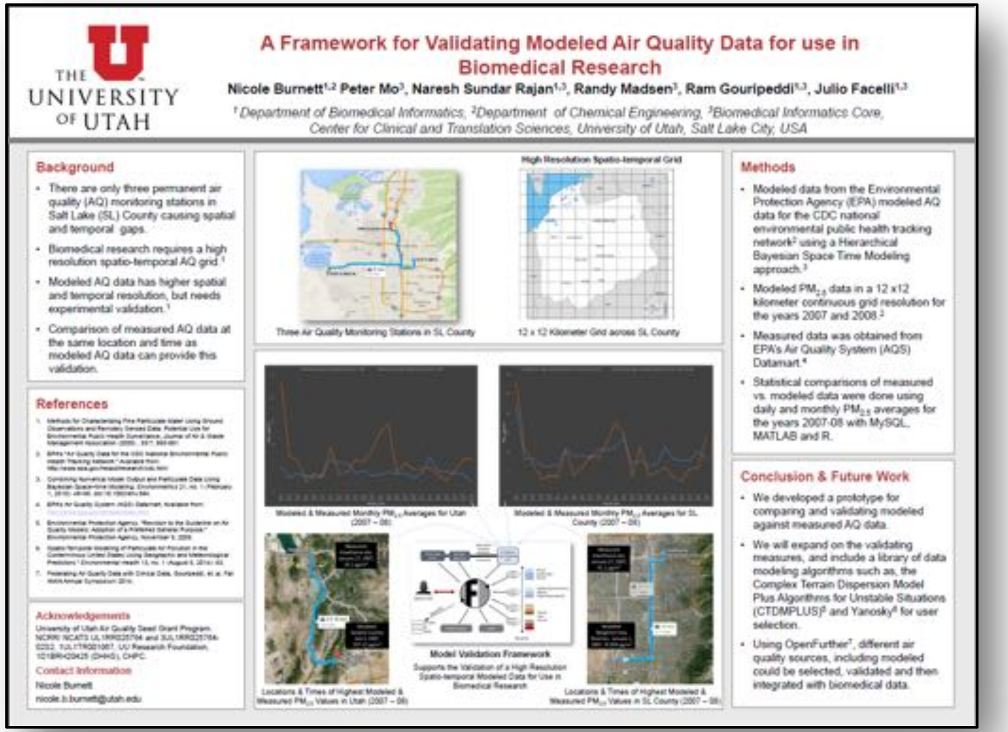
Uncertainty Characterization[26]

Reducible Uncertainties

- Uncertainties in input values of known conditions (e.g. emission characteristics and meteorological data)
- Errors in measured concentrations which are used to compute concentration residuals.
- Inadequate model physics and formulation

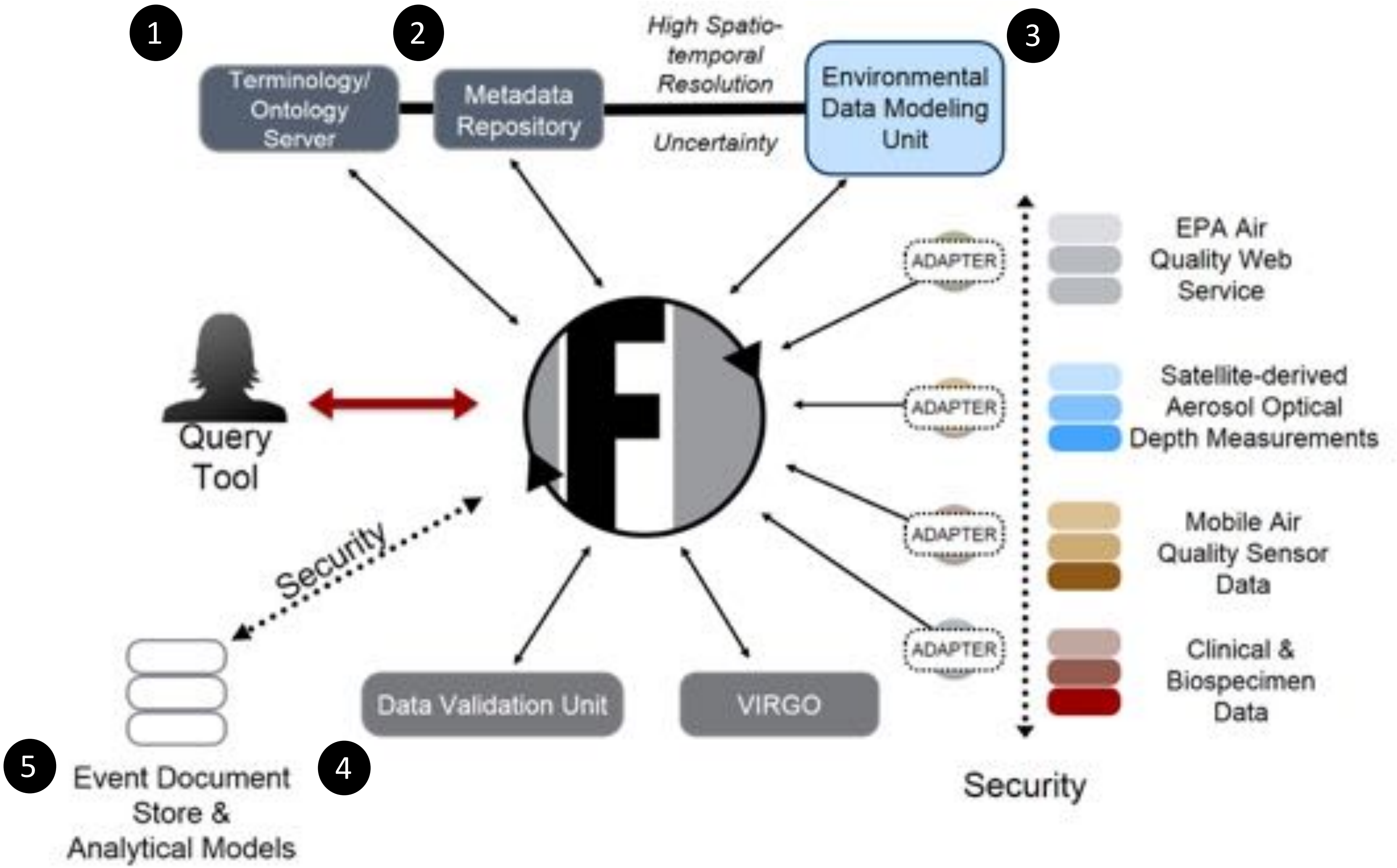
Exposure Uncertainty

- Difference between personal exposures and community average exposure.
- Difference between daily community average exposure and true ambient concentration.
- Difference between measured and true ambient concentration (measurement error).



- Selection of appropriate AQ sources and models
- Inherent: Variations in unknown conditions
- Reducible: Associated with the model and input conditions.
- Exposure Uncertainty: Arising due to differences in person's exposure and true ambient AQ levels.

OpenFurther Modifications



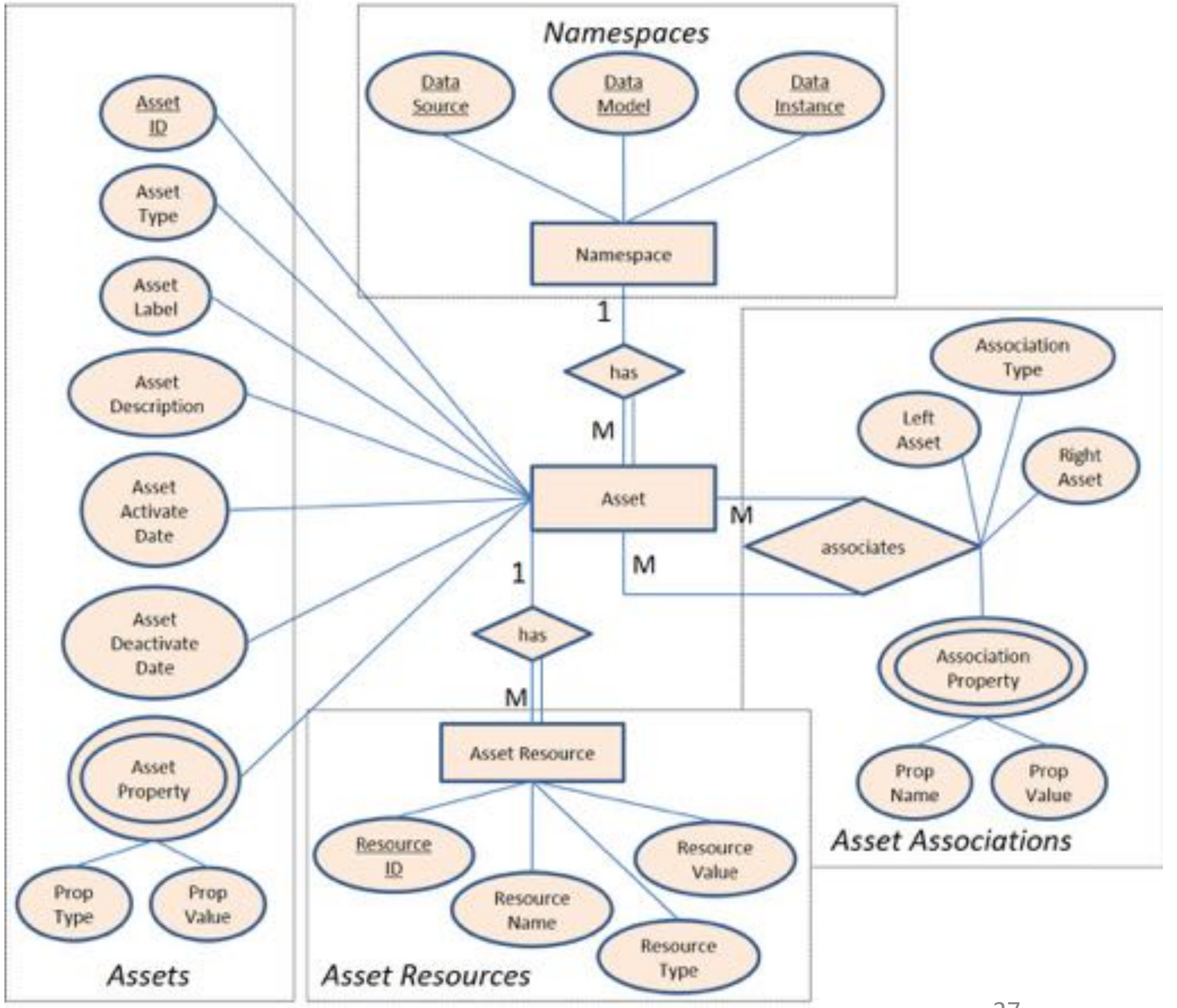
Semantics for Data Integration

- Semantic interoperability for Internet of Things (IoT)[27]
- Stored in Terminology/Ontology Server
- Examples

Semantic Sensor Network Ontology[28]	• Describes sensors and observations, and related concepts.
Sensor Model Language (SensorML)[29]	• Standard models and XML schema for describing sensors systems and processes associated with sensor observations.
PhenX Phenotypic Terms[30]	• Standard measures related to complex diseases, phenotypic traits and environmental exposures.
Exposure ontology (ExO)[31]	• Facilitate centralization and integration of exposure data to inform understanding of environmental health.
Standard biomedical ontologies and terminologies	• Gene Ontology, UniProt, SNOMED etc.

Metadata

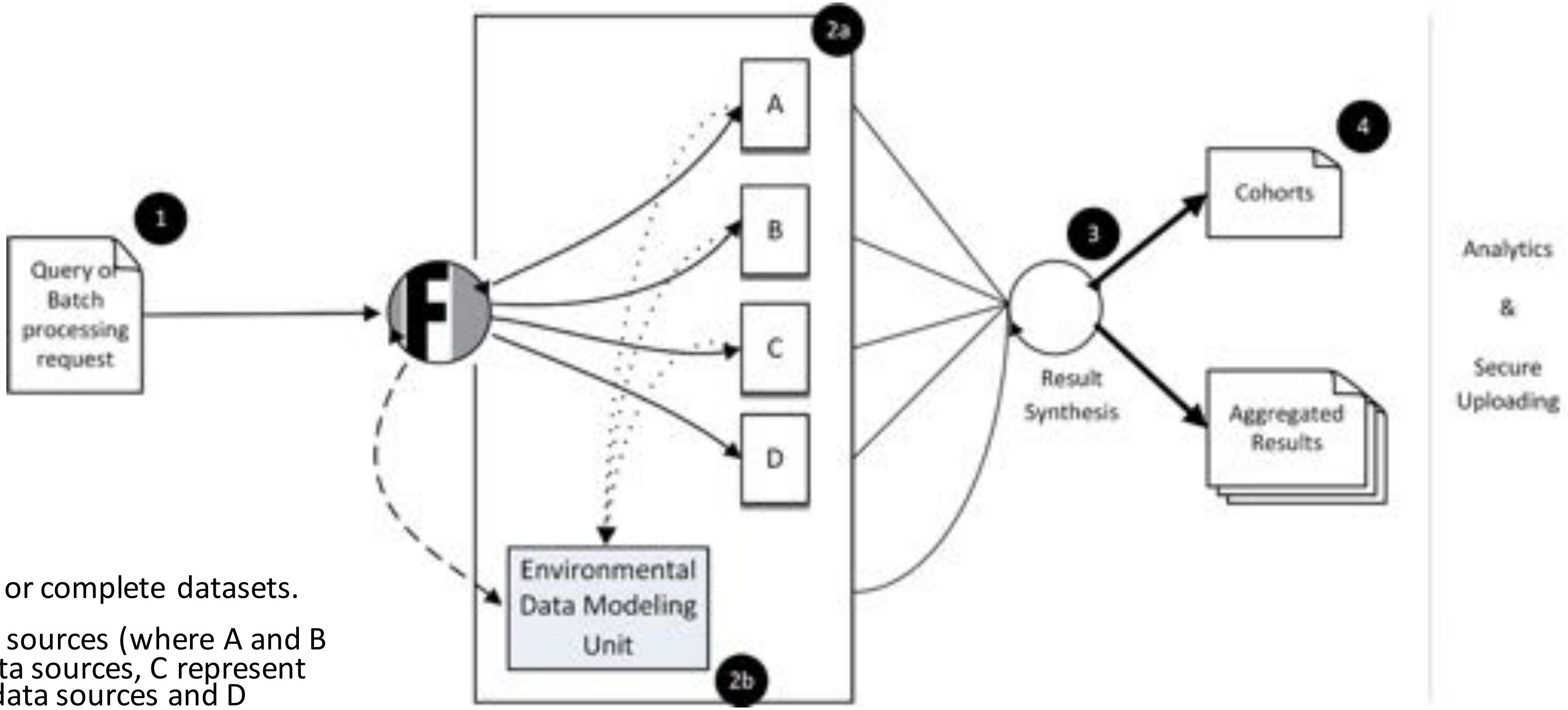
- Stored in Metadata Repository[18]
- Relational or graph stores
- Stores
  - Source and Central Data Models
    - Harmonized sensor data model
  - Data provenance and associated uncertainty
  - Inter-model transformative functions



Time & Events

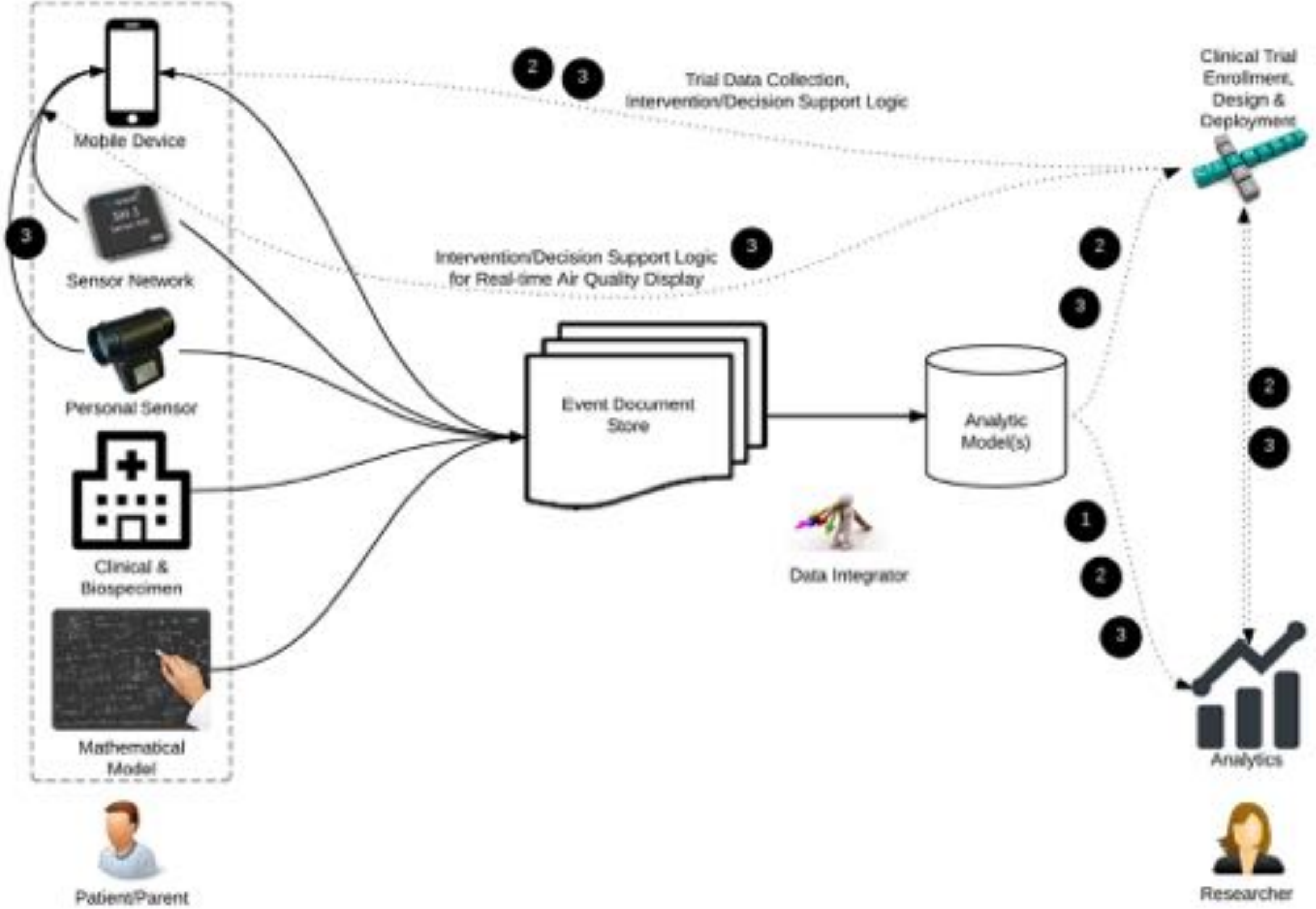
- Data modeled and stored in primitive form on a timeline as events.
- Transformed to higher/analytical models based on use-cases.
- Time modeled as[32]:
  - Unbounded: Contains upper and/or lower bounds with respect to its order relationship.
  - Dense: an infinite set of smaller units.
  - Discrete: every element has both an immediate successor and an immediate predecessor, if unbounded, and within the bounds if bounded.
  - Instants & Intervals (upper and lower time points).
  - Finest granularity available with the source.

Data Integration Workflow



1. User can query for a cohort or complete datasets.
2. (a & b) Heterogeneous data sources (where A and B represent mobile sensor data sources, C represent environmental monitoring data sources and D represent biomedical data sources), and (if needed) mathematical models using EDMU are selected.
   - Environmental data (A, B & C) harmonized to the central models stored in MDR. Selection of mathematical models managed in the EDMU.
3. OF synthesize results in different analytical models.
4. Presents them as cohorts and/or aggregated results.

Research Use-Cases



Retrospective & Population Studies

Conclusion

- Scalable informatics architecture that is generalizable beyond air quality and pediatric asthma.
- Integrates multi-scale and multi-omics data.
  - Genome-phenome-exposome
- *Big Data* integration: volume, velocity, variety, veracity for research value.
- Robust pipeline for research data delivery with decision support.
- Support different types of research.

Thank You

ram.gouripeddi@utah.edu

OpenFurther

OpenFurther.org

References

1. J. N. Newton, A. D. Briggs, C. J. Murray, D. Dicker, K. J. Foreman, H. Wang, M. Naghavi, M. H. Forouzanfar, S. L. Ohno, R. M. Barber, and others, "Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 386, no. 10010, pp. 2257–2274, 2015.
2. US EPA, "Basic Information about Lead in Drinking Water." Available: https://www.epa.gov/your-drinking-water/basic-information-about-lead-drinking-water.
3. C. P. Wild, "Complementing the Genome with an 'Exposome': The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology," *Cancer Epidemiol Biomarkers Prev*, vol. 14, no. 8, pp. 1847–1850, Aug. 2005.
4. C. P. Wild, "The exposome: from concept to utility," *Int. J. Epidemiol.*, vol. 41, no. 1, pp. 24–32, Feb. 2012.
5. G. W. Miller and D. P. Jones, "The Nature of Nurture: Refining the Definition of the Exposome," *Toxicol. Sci.*, vol. 137, no. 1, pp. 1–2, Jan. 2014.
6. G. W. Miller, *The Exposome: A Primer*, 1 edition. Amsterdam ; Boston: Academic Press, 2013.
7. "HEALS," *HEALS*. Available: http://www.heals-eu.eu/.
8. EXPOsOMICS. http://www.exposomicsproject.eu/
9. "Home - HELIX | Building the early life exposome." Available: http://www.projecthelix.eu/.
10. NIH's Environmental influences on Child Health Outcomes (ECHO) Program. https://www.nih.gov/echo
11. F. Martin Sanchez, K. Gray, R. Bellazzi, and G. Lopez-Campos, "Exposome informatics: considerations for the design of future biomedical research information systems," *Journal of the American Medical Informatics Association*, vol. 21, no. 3, pp. 386–390, May 2014.
12. Kesten S, Szalai J, Dzyngel B. Air quality and the frequency of emergency room visits for asthma. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 1995 Mar;74(3):269–73.
13. Weisel CP, Zhang J, Turpin BJ, et al. Relationships of Indoor, Outdoor, and Personal Air (RIOPA). Part I. Collection methods and descriptive analyses. Res Rep Health Eff Inst. 2005 Nov;(130 Pt 1):1–107; discussion 109–127.
14. Fleisch AF, Gold DR, Rifas-Shiman SL, et al. Air Pollution Exposure and Abnormal Glucose Tolerance during Pregnancy: The Project Viva Cohort. Environ Health Perspect. 2014 Feb 7 [cited 2014 Mar 7]; http://ehp.niehs.nih.gov/1307065/
15. Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos Environ. 2011 Nov;45(35):6267–75.
16. Utah Concludes Winter Inversion Season, Residents Proactively Engaged. http://www.deq.utah.gov/News/docs/2014/03Mar/DAQ_NewRelease_AirQualityStats_draftv2.pdf

References

17. Openfurther.org
18. Gouripeddi R, Schultz ND, Bradshaw RL, et al. FURTHeR: An Infrastructure for Clinical, Translational and Comparative Effectiveness Research. American Medical Informatics Association, 2013 Annual Symposium; 2013 Nov 16; Washington, D.C. http://knowledge.amia.org/amia-55142-a2013e-1.580047/t-10-1.581994/f-010-1.581995/a-184-1.582011/ap-247-1.582014
19. Gouripeddi, R., and Julio C Facelli. "Programmatically Linking Air Quality Indicators with Clinical Data." presented at the Air Quality, People and Health, 2nd Annual Retreat, University of Utah Guest House, University of Utah, Salt Lake City, April 14, 2014. http://www.airquality.utah.edu/files/2014/04/Ram_Programmatically-Linking-Air-Quality-Indicator-with-Clinical-Data.pdf.
20. Gouripeddi, R., Rajan, N.S., Madsen, R. Warner, P.B., Facelli, J.C., Federating Air Quality Data with Clinical Data, Annual Symposium of the American Medical Informatics Association, 2014
21. Whiteman CD, Hoch SW, Horel JD, Charland A. Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. Atmos Environ. 2014 Sep;94:742–53.
22. G. D. Silcox, K. E. Kelly, E. T. Crosman, C. D. Whiteman, and B. L. Allen, "Wintertime PM2.5 concentrations during persistent, multi-day cold-air pools in a mountain valley," Atmospheric Environment, vol. 46, pp. 17–24, Jan. 2012.
23. Pediatric Research Using Integrated Sensor Monitoring Systems. https://www.nibib.nih.gov/research-funding/prisms
24. McMillan, Nancy J., David M. Holland, Michele Morara, and Jingyu Feng. "Combining Numerical Model Output and Particulate Data Using Bayesian Space–time Modeling." *Environmetrics* 21, no. 1 (February 1, 2010): 48–65. doi:10.1002/env.984.
25. Yanosky, Jeff D., Christopher J. Paciorek, Francine Laden, Jaime E. Hart, Robin C. Puett, Duanping Liao, and Helen H. Suh. "Spatio-Temporal Modeling of Particulate Air Pollution in the Conterminous United States Using Geographic and Meteorological Predictors." Environmental Health 13, no. 1 (August 5, 2014): 63. doi:10.1186/1476-069X-13-63.
26. Burnett, N, Mo, P, Rajan, NS, Madsen, R, Gouripeddi, R. Facelli, JC, A Framework for Validating Modeled Air Quality Data for use in Biomedical Research, Environment and Sustainability Research Symposium, February 17, 2015, Union Ballroom, University of Utah, Salt Lake City. Utah.
27. ALLIANCE FOR INTERNET OF THINGS INNOVATION, "Semantic Interoperability," 2015.
28. Semantic Sensor Network Ontology: https://www.w3.org/2005/Incubator/ssn/ssnx/ssn
29. Sensor Model Language (SensorML): http://www.opengeospatial.org/standards/sensorml
30. PhenX Phenotypic Terms: https://bioportal.bioontology.org/ontologies/PHENX
31. Exposure ontology (ExO): https://bioportal.bioontology.org/ontologies/EXO
32. C. Combi, E. Keravnou-Papailiou, and Y. Shahar, *Temporal Information Systems in Medicine*, 2010 edition. New York: Springer, 2010.