The Emory CIDMATH Data Hub: Centralizing Data for Infectious Disease Modeling and Analytics

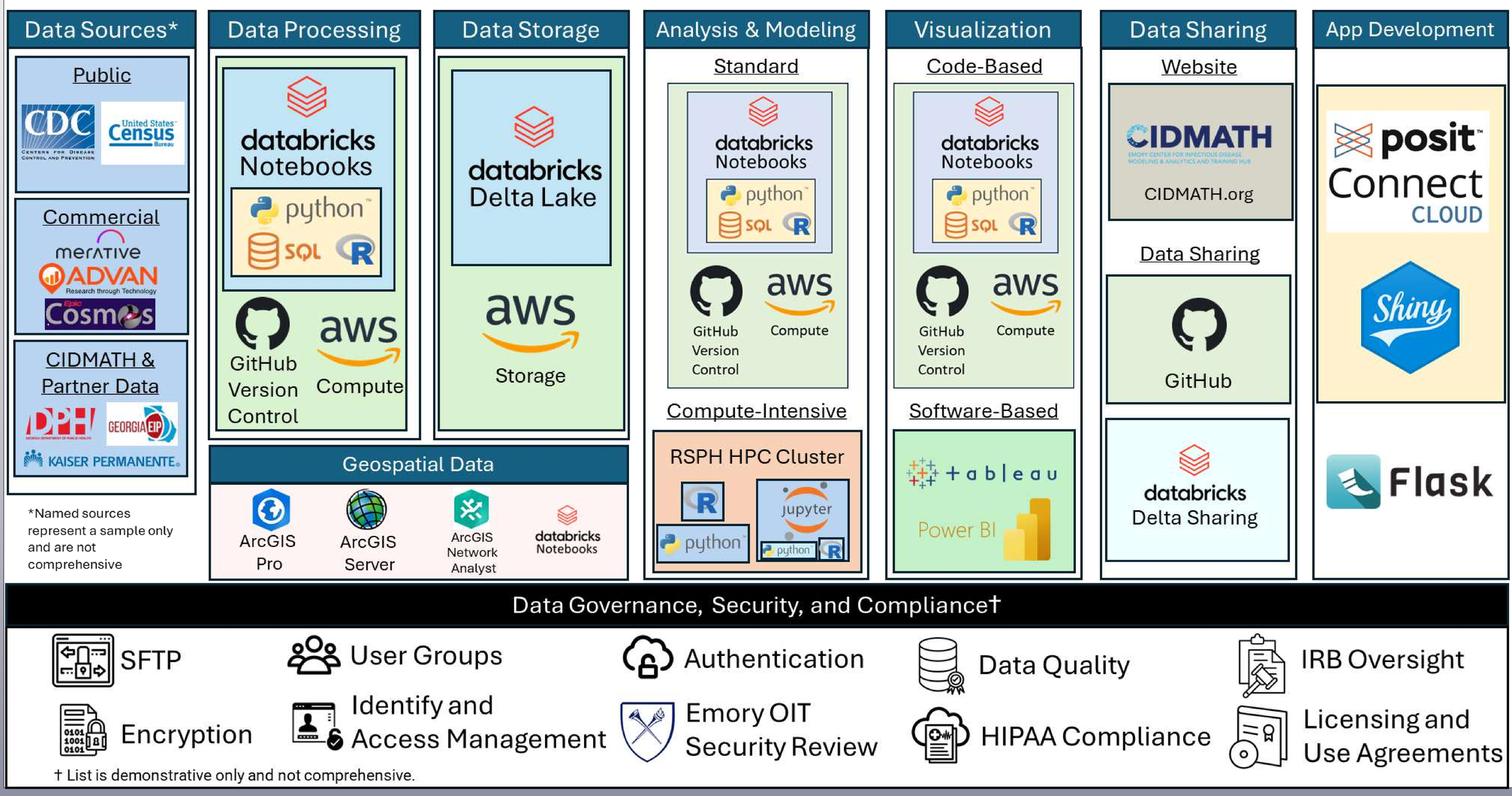
Connor Van Meter, MSPH[1]; Machi Shiiba, MPH[1]

[1]Center for Infectious Disease Modeling & Analytics and Training Hub, Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

ABSTRACT

Infectious disease modeling and analytics require timely access to diverse, heterogeneous data streams spanning environmental, clinical, behavioral, and programmatic domains. The Emory CIDMATH Data Hub is a centralized data platform designed to consolidate, augment, and integrate data across public, commercial, and partner sources into a unified environment supporting reproducible analytics, collaborative modeling, and scalable data pipelines – with appropriate data governance and security controls across data of varying sensitivity. Use cases span infectious disease forecasting, transmission model calibration, vaccine effectiveness analyses, wastewater epidemiology, and the development of public-facing tools and dashboards.

CIDMATH Data Infrastructure & Technology Overview



At its best, the data hub can be used to seamlessly integrate data across a variety of domains and streams to build a more complete/informative picture of the underlying system of disease transmission to better support preparedness and response activities.

1. **Pathogen & Disease**
   a. Disease Characteristics
   b. Clinical Outcomes & Severity
   c. Genomic & Molecular Epidemiology
2. **Surveillance & Detection**
   a. Traditional Disease Surveillance
   b. Lab & Testing
   c. Wastewater Surveillance
   d. Seroprevalence & Immunological Landscape
   e. Infodemic & Digital Surveillance
3. **Population & Behavior**
   a. Demographics
   b. Underlying Population Health & Comorbidities
   c. Behavioral & Risk Perception Data
   d. Social Contact Patterns
4. **Movement & Environment**
   a. Transportation & Mobility
   b. Built Environment & Land Use
   c. Environmental Factors
   d. Occupational & Congregate Setting Data
5. **One Health & Zoonotic**
   a. Wildlife Surveillance
   b. Domestic & Agricultural Animal Health
6. **Healthcare System**
   a. Healthcare Capacity & Infrastructure
   b. Healthcare Utilization & Access
   c. Public Health Department Operations
7. **Response & Economics**
   a. Policy & Non-Pharmaceutical Interventions
   b. Immunization & Pharmaceutical Interventions
   c. Supply Chain & Logistics
   d. Economic Impact & Healthcare Costs
8. **Global & Cross-Border**
   a. International Disease Surveillance & Reporting
   b. Cross-Border Mobility & Travel
   c. Global Variant & Pathogen Emergence
9. **Reference / Ontology Infrastructure**
   a. Geography Reference Data
   b. Temporal Reference Data
   c. Coding Systems & Ontologies
   d. Population Denominators & Crosswalks

ACKNOWLEDGEMENTS

This project was made possible by the Insight Net cooperative agreement CDC-RFA-FT-23-0069 from the CDC's Center for Forecasting and Outbreak Analytics.
Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.