Data Management and Analysis Solutions for Meta-Analysis of Multi-Domain Data

ClarLynda Williams-DeVane, Ph.D.
Assistant Professor of Bioinformatics/Biostatistics

Systems Toxicology: Research Problem

Issues with the Typical Workflow for Human and Animal studies
- No centralized location for all experimental information
- Early release data are subject to change during project life cycle
- Data accessibility is difficult
- Redundant data processing task
- Lack of enterprise-wide definitions for information types & values
- Collaborative research (multi-discipline, multi-location) presents data management challenge
- No access to important metadata in scientists notebooks
Solution: WikiLIMS

WikiLIMS Feature

The EPA system provides scientists with the following features:

- Automate repetitive data processing tasks
  - Sample Tracking
  - Repetitive Calculations/Plots
- Allows data to be reused for subsequent meta-analysis
  - No need to search for, assemble, or reformat data
- Methods/data from different Scientists and between labs are organized and available when the paper is written
- Ease of Publication
  - Methods and Results are easily accessible
- Generate Work Request
- Store data in a version control system
Cardiopulmonary Meta-Analysis

• Multiple Strains of Rats
  – Hypertensive
  – Diabetic
• Three Air Pollutants
  – Zinc
  – Diesel Exhaust
  – Ozone
• 5 Experiments
• Two PI’s
• Common Normalization Pipeline
• Common Microarray Platform
• Common Clinical Measures
• Representative Demographics
• Modeling
  – Commonalities

Conclusions

Results
• Develop Models Multi-Domain Data
• Increased Power
  – Smaller Gene Set
  – Larger Sample Size
  – Future Work
• Prototype supporting human data
• Further development of analysis algorithms

Acknowledgements
• USEPA
  – Stephen Edwards, Ph.D.
  – Archana Radadia
  – Nina Rountree
• Bioteam
  – Bill Van-Etten
  – Jiesheng Zhang

National Cancer Institute grant 552986