



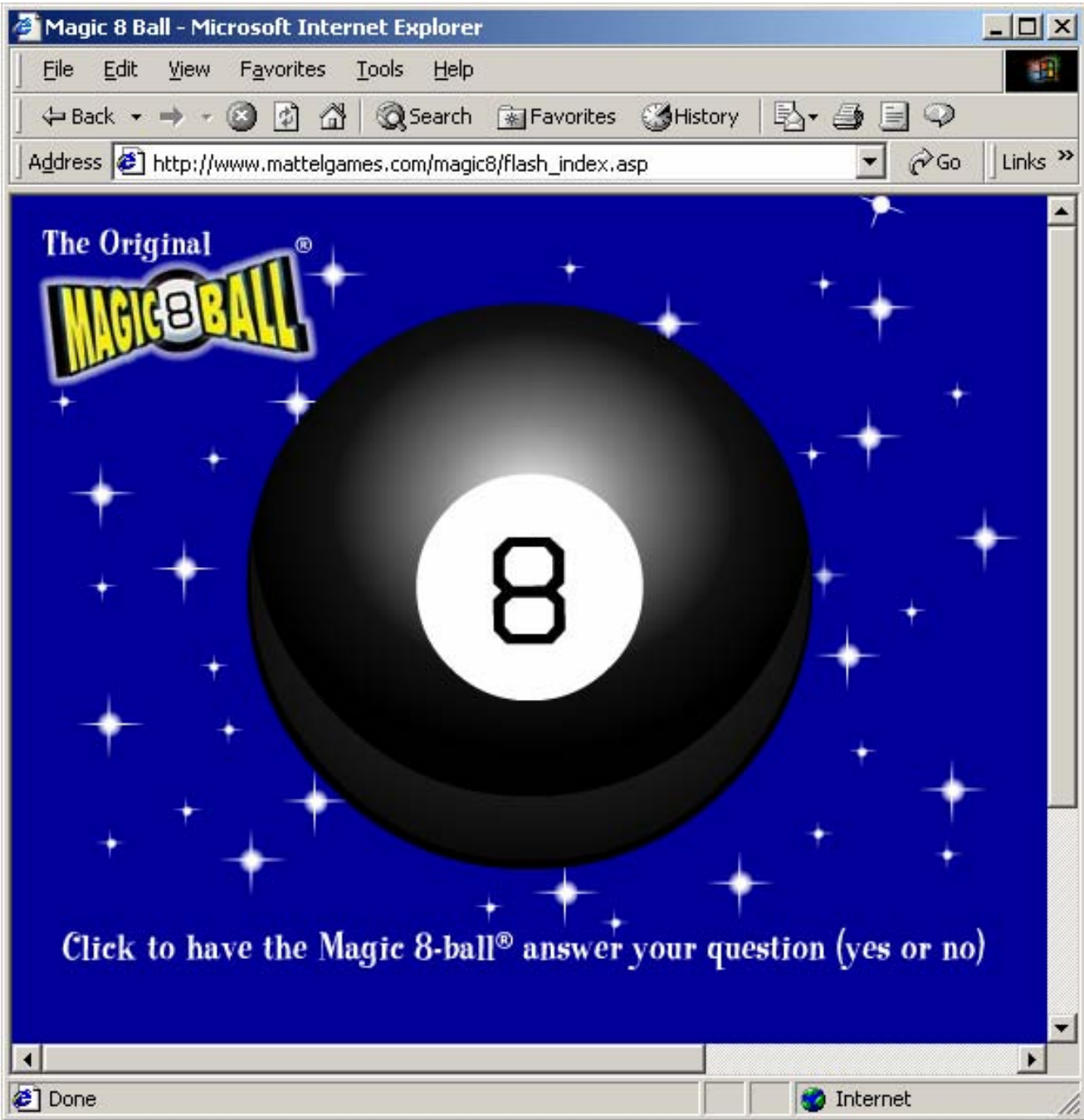
Center  
for  
Bioinformatics

# What Approaches are Needed to Advance Proteomic Analysis?

Ken Buetow

NCICB/NCI/NIH/DHHS





The Original  
**MAGIC 8 BALL**



Click to have the Magic 8-ball® answer your question (yes or no)



# Loosely integrated perspective from:

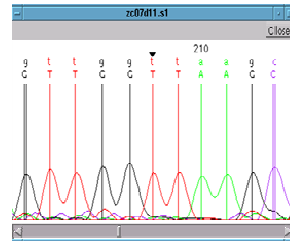
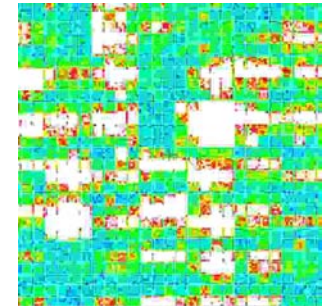
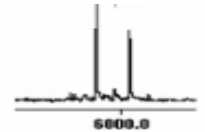
- Fallen population geneticist/biostatistician
- Interloping laboratory scientist
- Non-credentialed bioinformatician



# Scarred veteran of the molecular technology revolution ...

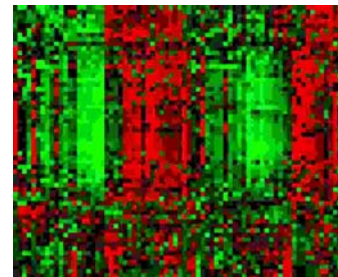
## ■ Genetics

- Southern blots
  - RFLPs
  - VNTRs
- PCR-based
  - STRPs
  - Oligo-based assays
  - Mass spec.
- Mapping
  - Linkage
  - association



## ■ Genomics

- Sequencing
  - Silver stained gels
  - Four color sequencing – gel-based
  - Four color sequencing – capillary
- Transcript profiling
  - SAGE
  - cDNA array
  - Oligo nucleotide





# Lessons learned...

- Initial promise always overstated
  - Innovator's dramatic finding required to break through
  - Insufficient experience with platform
    - known versus unknown demons
  - Disconnect with independent biologic/mechanistic validation



# Lessons learned...

- Open is good!
  - Data sharing
  - Open source code
  - Analytic jamborees





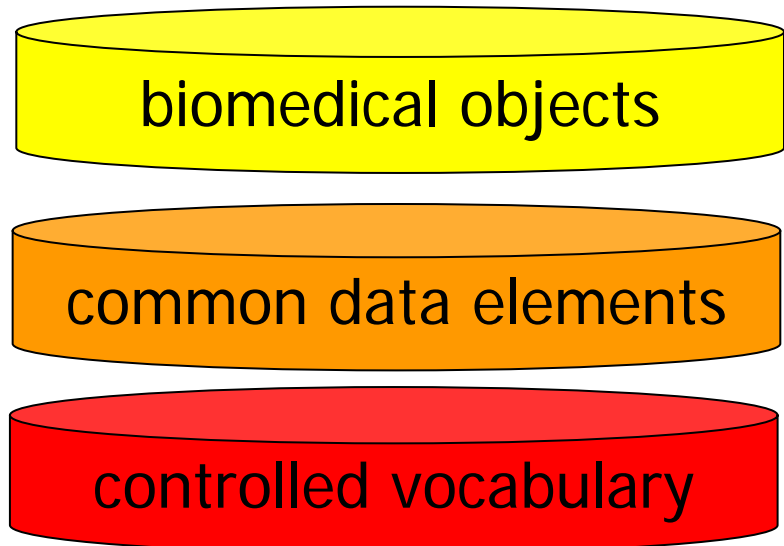
# Lessons Learned...

- Standards versus standardization
  - Quality control
    - Experimental protocols
    - Standard external reference
  - Internal controls
  - Standard methodology
  - Data standards
    - Use established standards where they exist
    - Modify/extend existing standards where ever possible
    - Develop new standards “just in time”
  - Standards can NOT be proprietary



# caCORE – common ontologic representation environment

- Information integration
- Cross-discipline reasoning







# Enterprise Vocabulary

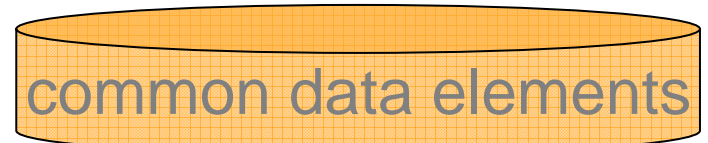
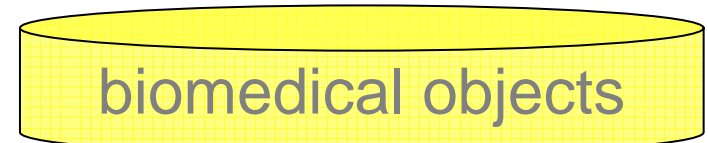
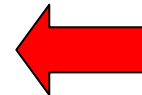
## ■ NCI Meta-Thesaurus

(Cross-map standard vocabularies/ontologies, e.g. SNOMED, MEDRA, ICD)

- Semantic integration, inter-vocabulary mapping
- UMLS Metathesaurus extended with cancer-oriented vocabularies
  - 800,000 Concepts, 2,000,000 terms and phrases
  - Mappings among over 50 vocabularies

## ■ NCI Thesaurus

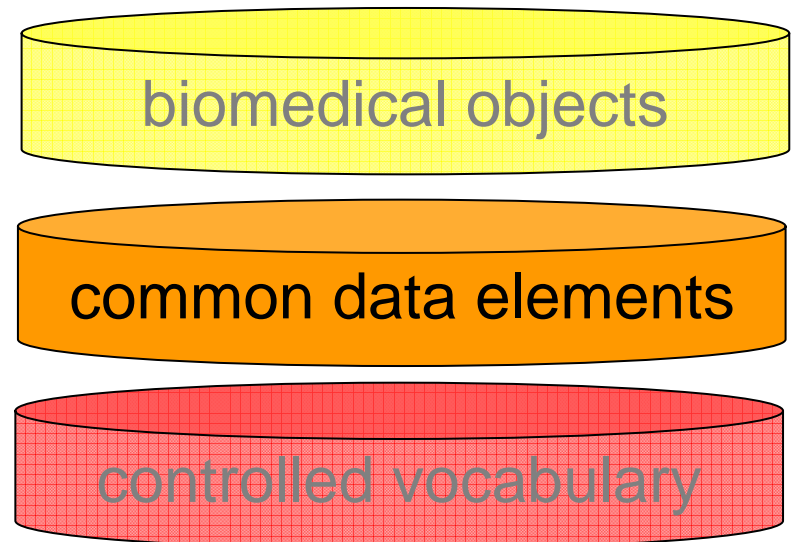
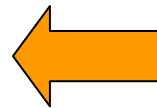
- Description logic-based
- 18,000 "Concepts"
  - Concept is the semantic unit
  - One or more terms describe a Concept – synonymy
  - Semantic relationships between Concepts





# Common Data Elements

- Structured data reporting elements (e.g. LOINC)
- ISO11179 compliant





# Lessons Learned...

- Quality measures are transforming
  - Qualitative and quantitative
  - Objective measures critical



# Lessons Learned...

- The devil is in the details
  - Experimental inputs can be as critical as important as outputs
  - Laboratory information management systems (LIMS)



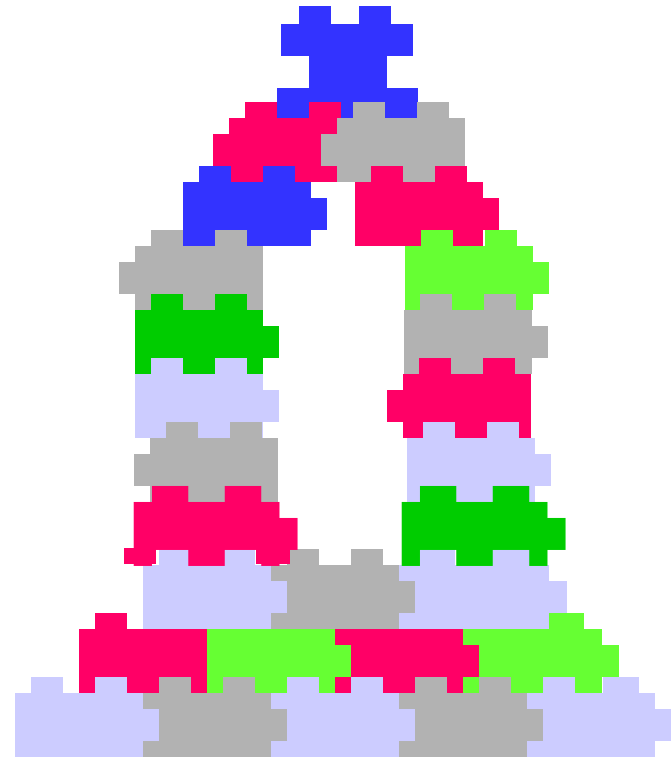
# Lessons Learned...

- Today's tools are not likely to be tomorrow's
  - Killer app's
    - Accessible, useful, user friendly apps critical to adoption
    - Not always the best approach (Eisen's cluster analysis)
  - Everything old is new again
    - On the shoulder's of giants...
    - Simpler methods are better
  - Design infrastructure that facilitates rapid exploration of new methods
    - Open source
    - Isolate data from applications
    - Component architecture



# Components: software parts

- Small parts are better for building flexible shapes
- Have a uniform interface medium
- Snap-together connectivity
- Internals can be made from widely varying technologies



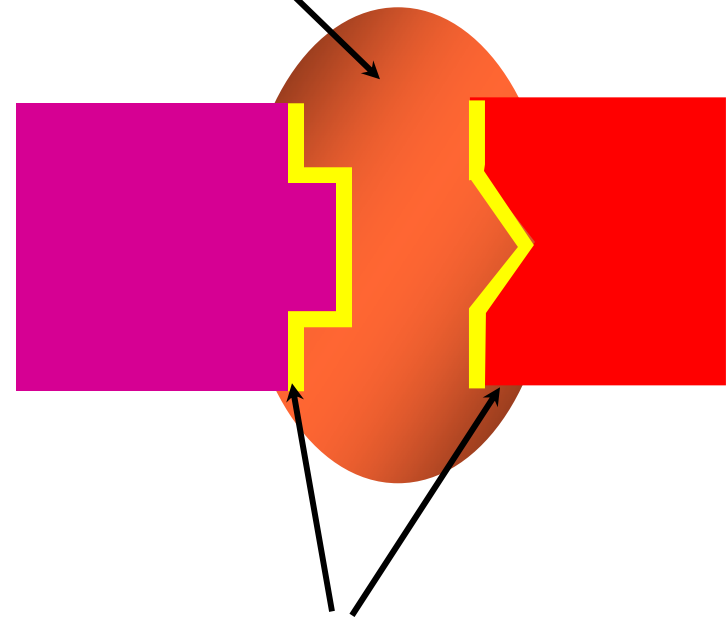




# Boundaries and Interfaces

- focus on boundaries, interfaces, how things fit together,
- not on the internal details of how they're built: assume that will be diverse & changing

The glue that binds parts together is metadata infrastructure

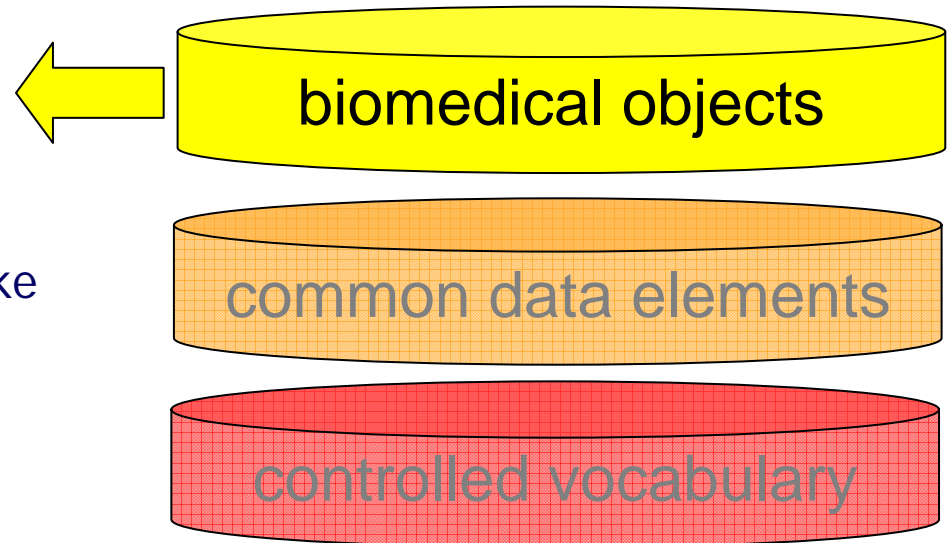


Shape of boundary is defined in APIs



# Biomedical Information Objects

- Computer model of a biomedical object – “Plato’s Forms”
  - capture properties of object
  - can be joined together to make complex systems
  - isolate data from data source
  - isolate applications from data
- Examples:
  - HL7-RIM
  - MAGE-OM





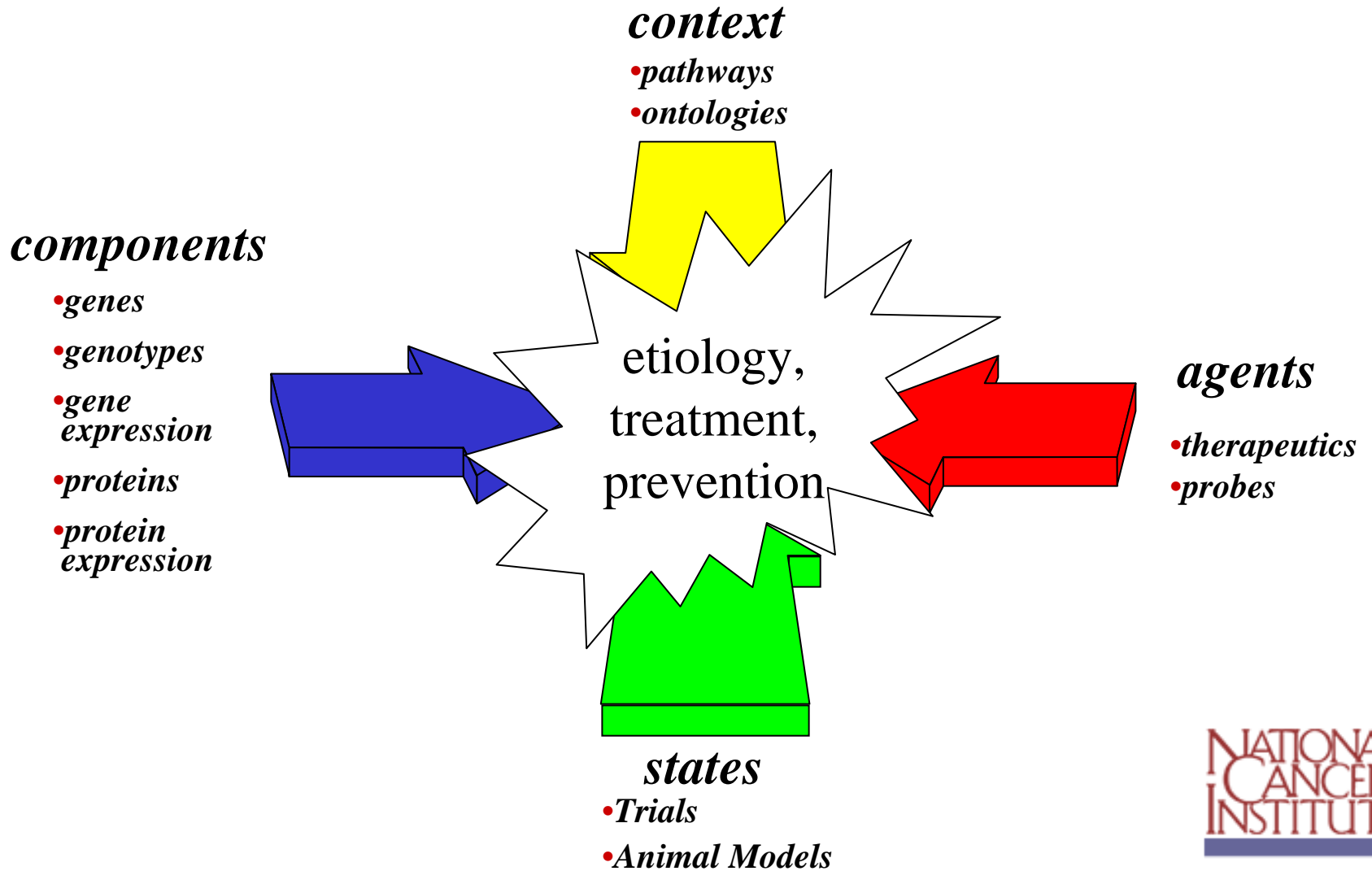
# Lessons Learned...

- Study design really does matter
  - New technologies ALWAYS initially appear to reduce the need for rigor
  - Animal models
    - Critical to validation
  - Clinical research considerations
    - Training/testing sets
    - validation sets
    - Over fitting



# Lessons Learned...

- You really are going to want to connect these results to other outcomes!
  - Other data types
  - Clinical outcomes





## **Pathway Database**

- Enhance value of imperfect, but available, pathway knowledge
- Make biological assumptions explicit
- Merge data from separate pathways
- Build a causal framework to support (future) quantitative simulation/analysis



