

# ***Knowledge Environments for Biomedical Research***

**Michael F. Huerta, Ph.D.**

Associate Director for Scientific Technology Research  
National Institute of Mental Health, NIH



# Digital Biomedicine

- Characteristics of biomedical research data:
  - ◆ Vast
  - ◆ Complex
  - ◆ Diverse
- Digital approaches are:
  - ◆ Increasing – as more data are acquired digitally, and computation becomes more powerful
  - ◆ Necessary – to make sense and use of these data
  - ◆ Key – to advancing biomedical research



# Digital Biomedicine

- Characteristics of biomedical research data:
  - ◆ Vast
  - ◆ Complex
  - ◆ Diverse
- Digital approaches are:
  - ◆ Increasing – as more data are acquired digitally, and computation becomes more powerful
  - ◆ Necessary – to make sense and use of these data
  - ◆ Key – to advancing biomedical research



# Foundational Challenge for Digital Biomedicine

*Developing and evolving common knowledge environments for biomedical research*



# What are Knowledge Environments (KEs)?

- Digital tools and resources depend upon explicit “understandings” of concepts related to the data upon which they compute
- Examples of data-related concepts:
  - ◆ What the data represent (e.g., nucleotide or gene)
  - ◆ Semantic relationships among different data types (nucleotides comprise genes)
  - ◆ Ontologic relationships among different data types (A, T, C, and G are all in the nucleotide class)
  - ◆ Relationship of terms used to definitions in real world (is gene defined as a particular sequence or a heritable trait?)
- The entire set of understandings represents the knowledge environment (KE)
- Each digital tool or resource has its own KE – its own “internal logic”



## What are Knowledge Environments (KEs)?

- Digital tools and resources depend upon explicit “understandings” of concepts related to the data upon which they compute
- Examples of data-related concepts:
  - ◆ What the data represent (e.g., nucleotide or gene)
  - ◆ Semantic relationships among different data types (nucleotides comprise genes)
  - ◆ Ontologic relationships among different data types (A, T, C, and G are all in the nucleotide class)
  - ◆ Relationship of terms used to definitions in real world (is gene defined as a particular sequence or a heritable trait?)
- The entire set of understandings represents the knowledge environment (KE)
- Each digital tool or resource has its own KE – its own “internal logic”



## What are Knowledge Environments (KEs)?

- Digital tools and resources depend upon explicit “understandings” of concepts related to the data upon which they compute
- Examples of data-related concepts:
  - ◆ What the data represent (e.g., nucleotide or gene)
  - ◆ Semantic relationships among different data types (nucleotides comprise genes)
  - ◆ Ontologic relationships among different data types (A, T, C, and G are all in the nucleotide class)
  - ◆ Relationship of terms used to definitions in real world (is gene defined as a particular sequence or a heritable trait?)
- The entire set of understandings represents the knowledge environment (KE)
- Each digital tool or resource has its own KE – its own “internal logic”



## What are Knowledge Environments (KEs)?

- Digital tools and resources depend upon explicit “understandings” of concepts related to the data upon which they compute
- Examples of data-related concepts:
  - ◆ What the data represent (e.g., nucleotide or gene)
  - ◆ Semantic relationships among different data types (nucleotides comprise genes)
  - ◆ Ontologic relationships among different data types (A, T, C, and G are all in the nucleotide class)
  - ◆ Relationship of terms used to definitions in real world (is gene defined as a particular sequence or a heritable trait?)
- The entire set of understandings represents the knowledge environment (KE)
- Each digital tool or resource has its own KE – its own “internal logic”



## Why are Common KEs Important?

- Unless planned, different digital tools and resources have different KEs
- In order for digital data, tools and resources to work together, they must either share the same KE, or have the ability to have their KEs to map to each other.
- Without a common KE (or the ability to map across different KEs):
  - ◆ Different data, tools, and resources cannot work together and cannot be shared in meaningful ways – this diminishes the value of the NIH investment in them
  - ◆ Different data cannot be compared rigorously with each other – this diminishes their scientific value



## Why are Common KEs Important?

- Unless planned, different digital tools and resources have different KEs
- In order for digital data, tools and resources to work together, they must either share the same KE, or have the ability to have their KEs to map to each other.
- Without a common KE (or the ability to map across different KEs):
  - ◆ Different data, tools, and resources cannot work together and cannot be shared in meaningful ways – this diminishes the value of the NIH investment in them
  - ◆ Different data cannot be compared rigorously with each other – this diminishes their scientific value



## Why are Common KEs Important?

- Unless planned, different digital tools and resources have different KEs
- In order for digital data, tools and resources to work together, they must either share the same KE, or have the ability to have their KEs to map to each other.
- Without a common KE (or the ability to map across different KEs):
  - ◆ Different data, tools, and resources cannot work together and cannot be shared in meaningful ways – this diminishes the value of the NIH investment in them
  - ◆ Different data cannot be compared rigorously with each other – this diminishes their scientific value



***Common KEs are essential for optimally comparing, sharing, and understanding digital data, tools, and resources***



# Strategies for Developing Common KEs

## ■ Top-down

- ◆ Pretty straightforward
- ◆ Best for very well-understood areas
- ◆ Compliance and acceptance can be difficult to achieve

## ■ Bottom-up

- ◆ Takes a very long time
- ◆ It may not happen at all

## ■ Something in-between

- ◆ Guidance and support from “above”
- ◆ Leadership and participation from “below”
- ◆ Requires motivated players



# Strategies for Developing Common KEs

- Top-down
  - ◆ Pretty straightforward
  - ◆ Best for very well-understood areas
  - ◆ Compliance and acceptance can be difficult to achieve
  
- Bottom-up
  - ◆ Takes a very long time
  - ◆ It may not happen at all
  
- Something in-between
  - ◆ Guidance and support from “above”
  - ◆ Leadership and participation from “below”
  - ◆ Requires motivated players



# Strategies for Developing Common KEs

- Top-down
  - ◆ Pretty straightforward
  - ◆ Best for very well-understood areas
  - ◆ Compliance and acceptance can be difficult to achieve
- Bottom-up
  - ◆ Takes a very long time
  - ◆ It may not happen at all
- **Something in-between**
  - ◆ Guidance and support from “above”
  - ◆ Leadership and participation from “below”
  - ◆ Requires motivated players



# Toward Common Knowledge Environments

- NIH science comprises numerous communities
- Communities described by data-related concepts, e.g.:
  - ◆ Genetic data
  - ◆ Behavioral data
  - ◆ Epidemiologic data
- Not all communities are equally:
  - ◆ Motivated or able to use digital approaches
  - ◆ Motivated or able to participate in developing KEs



# Toward Common Knowledge Environments

## First Steps: KEBR Breakout Groups

- Address crucial issues to form strategy/ies
- Each group to address same issues – in different order
- Report back will demonstrate points of convergence and divergence
- Output of KEBR – Elucidation of strategies for developing, sustaining, and enhancing KEBRs
- Strategies → Leadership of major biomedical informatics initiatives at NIH → Implementation plans

