Standards Development Toolkit: A Standards Theory Overview

Dave Mann, Ph.D.

JoAnn Brooks, Ph.D.

MITRE

Background

- Success of the CVE (Common Vulnerabilities and Exposures) information standard led to push for more
  - "We need a CVE for [some new area of cyber-security]"
- But efforts to establish new standards led to unexpected conflicts … raising some key questions
  - Why does computation work well in some fields but not others?
  - Why do some fields or disciplines stabilize and others don't?
  - What does this imply for cyber security?
  - What are we doing when we create schemas for cyber?
- MITRE invested internal funds to improve our understanding of
  - The range of design choices available
  - How standards get used
  - Why some standards efforts fail and others succeed

  by applying insights from non-engineering fields
  - Sociology of shared meaning
  - Library & Information Science, Knowledge Organization Science
  - Cognitive Psychology

Outline

- Background
- General Theory
- Theory Applied to Different Genres of Human-Oriented Standards & Recommendations
- Theory Applied to Potential Standards Efforts & Recommendations
- Summary: Recommendations to Make Standards Efforts More Effective

Classifications

Categorizations

Dictionaries

Controlled Vocabularies

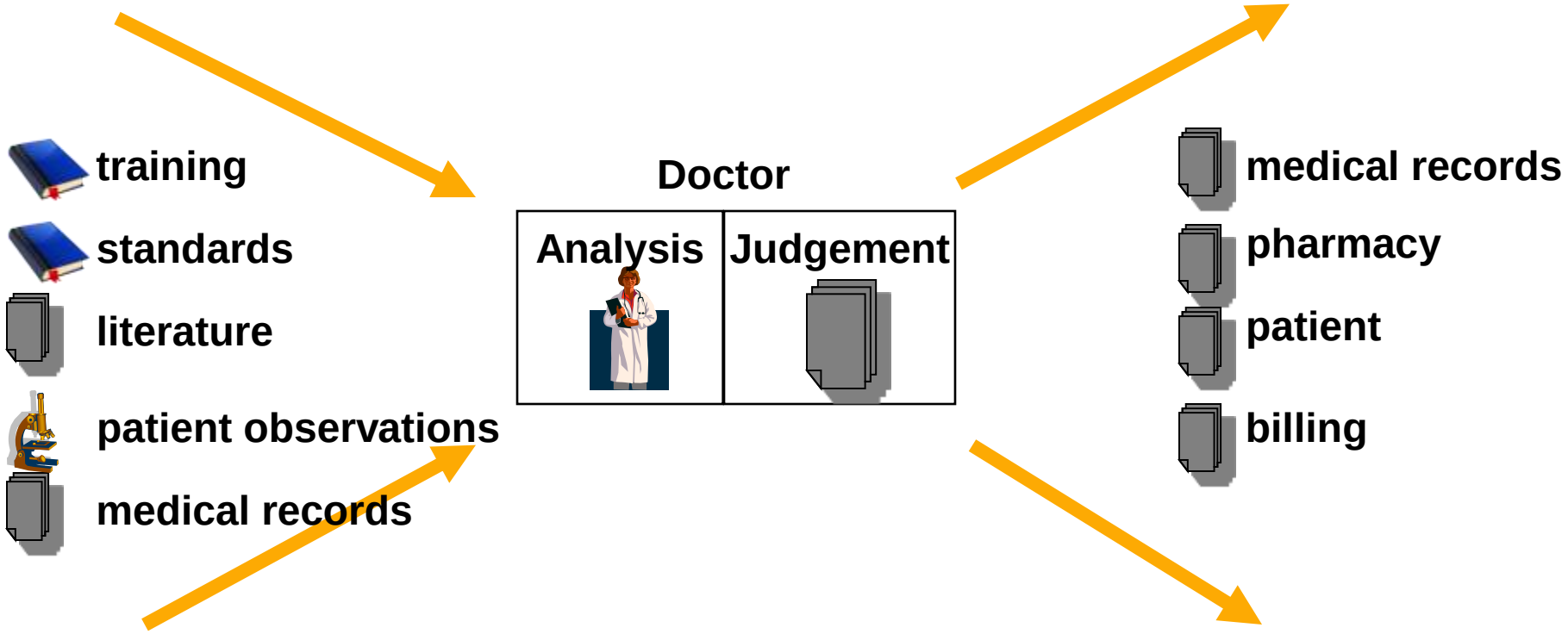
Taxonomies

Classified Identifiers

Ordered Identifiers

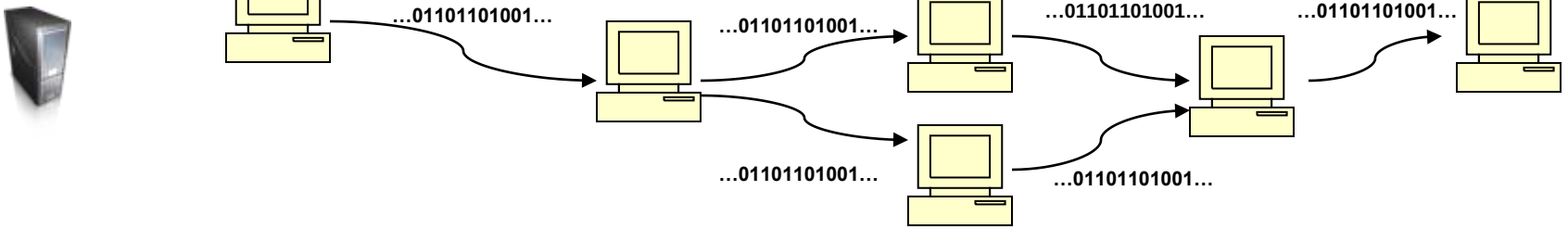
Nominal Identifiers

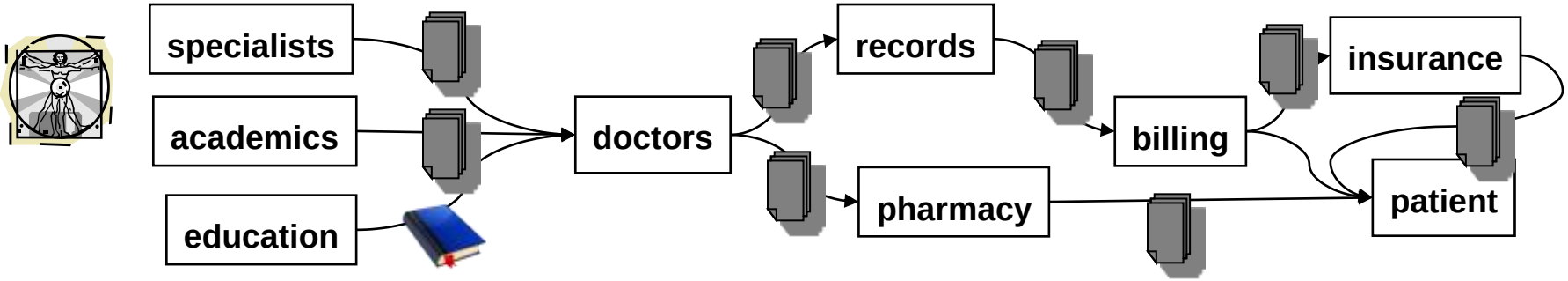
GENERAL THEORY
The Basic Analysis Process



We'll call these "information products", where (in general) "information" does NOT imply digital format

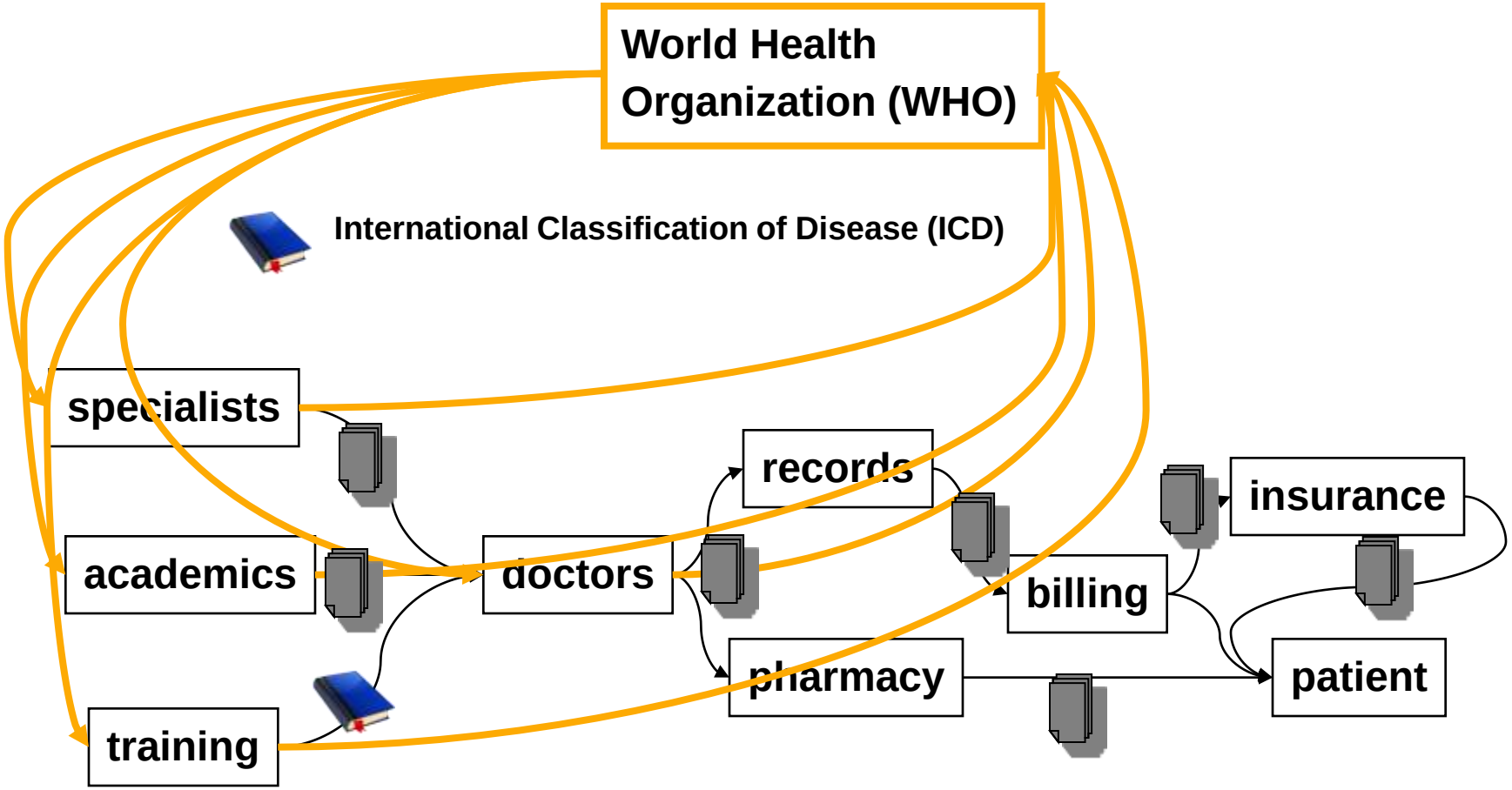
GENERAL THEORY
Human Analysis Ecosystem & the Network





- **Machine Standards – Allow computers systems to function and communicate**
- **Human Standards – Organize fields, disciplines, enterprises**
  - **International Classification of Disease (ICD), CVE, CCE, CWE…**

GENERAL THEORY
The Standards "Feedback Loop"



GENERAL THEORY
Types of Machine Oriented Standards

- Installable programs
- Algorithms
- Programming languages
- File formats
- Protocols
- Shared libraries

Pure Machine Oriented Standards

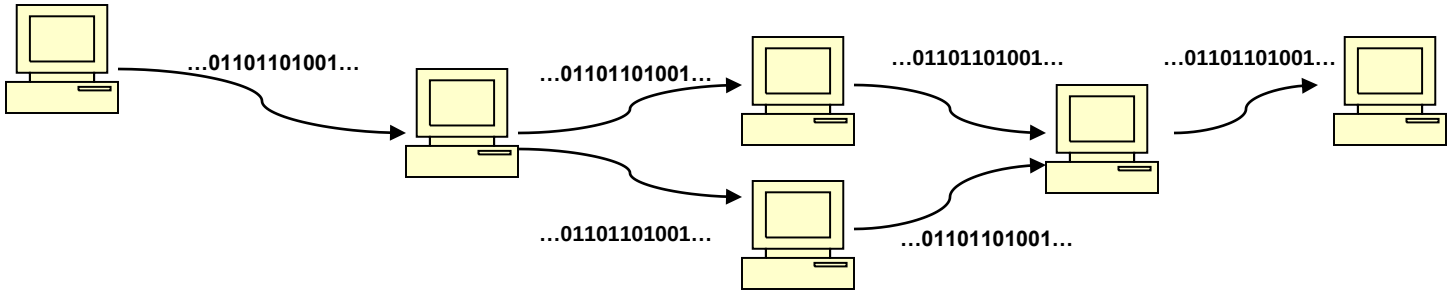
- Formal taxonomies
- Schema (aka Formal Thesauri)
- Ontologies
- Bibliographic Indexing Systems
  - Top down
  - Faceted

Human/Machine InterfaceStandards

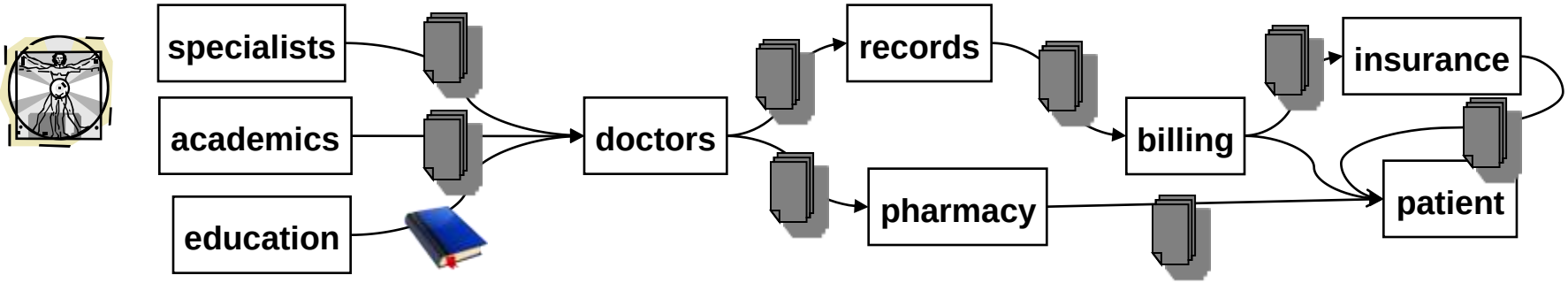
GENERAL THEORY
Types of Human Oriented Standards

- Classifications (non-overlapping classes)
- Categorizations (overlapping categories)
- Dictionaries, Glossaries, Thesauri
- Controlled Vocabularies
- Taxonomies and Hierarchies
- Classified Indexes
- Ordered Identifiers
- Nominal Identifier Systems

RECOMMENDATION
Think in Social and Organizational Terms



- Engineers focus on computer systems but…
- … they often overlook (or can't see) the social and organizational systems of use



THEORY
Types of Human Oriented Standards

- Assume you're familiar with most machine-oriented standards
- Consider types of human-oriented standards:
  - Classifications (non-overlapping classes)
  - Categorizations (overlapping categories)
  - Dictionaries, Glossaries, Thesauri
  - Controlled Vocabularies
  - Taxonomies and Hierarchies
  - Classified Indexes
  - Ordered Identifiers
  - Nominal Identifier Systems

THEORY
Classifications

Definition

Built on the premise of classical category theory.

- **Defined by essential features**
- **Membership determined by features**
- **No typicality**
- **Classifier independence**
- **Inheritance**
- **Non-overlapping boundaries**

Examples

Vehicle Identification Number (definition)

Dewey Decimal Classification

Library of Congress Classification

Focus

Define classes, but not terms

Structure

Strict hierarchy

THEORY
Categorizations

Definition

Categories are related to each other in arbitrary and non-hierarchical ways.

- **Ambiguous definitions**
- **Flexible/contingent membership**
- **More/less typical members**
- **Non-transitive inheritance**
- **Overlapping boundaries**

Examples

Musical genres

Encyclopedias

Most white papers

This list of “types”

Focus

Emphasize categories, but not terms

Structure

General graph

THEORY
Dictionaries

Definition

A categorization with a clear focus on labels

- **Typically rendered hierarchically at presentation level in print (shelving problem)**
- **Typically rendered as graph on web (e.g., Webster's online, Princeton Wordnet)**

Examples

Standard dictionaries (word, definition)

Glossaries (word, definition)

Language translation dictionaries (relationships)

Thesauri (relationships)

Focus

Labels

Structure

General graph

THEORY
Controlled Vocabularies

Definition

Dictionary that enforces a single term for every unique definition

- Deep history in Library Sciences
- Typically used to create tags for search and retrieval
- Subject Headings – Used by catalogers to describe works of art. Typically use broad terms
- Bibliographic Thesauri – Used by indexers to apply search terms to documents
- Similar to ontological CV

Examples

ACM or AMS subject headings

Card catalog subject headings

Focus

Term and definition

Structure

General graph

THEORY
Taxonomies

Definition

A classification that assigns an (agreed upon) label or term to each class

- Hierarchy = Relationships form rooted tree
- Classification = Class definitions related hierarchically
- Assigns names to all nodes at all levels

Examples

Section headings in a paper (the shelving problem)

International Code of Zoological Nomenclature

(Taxonomist operate with "taxonomic freedom" and are free to assign different names to the same species)

Focus

Term

Structure

Hierarchical

THEORY
Classified Indexes

Definition

Utilizes an established classification to classify individual items and to assign identifiers to classified items

- **IDs are unique alphanumeric strings**
- **IDs only assigned to classified items, not to classes in classification (not a taxonomy)**

Examples

Library of Congress Classification (LCC) numbers

Dewey Decimal Classification (DDC) numbers

Vehicle Identification Numbers (VINs)

International Standard Book Numbers (ISBNs)

Focus

Labels

Structure

Hierarchical

THEORY
Ordered Identifiers

Definition

Assigns ordered alpha-numeric identifiers to objects in a manner such that y correspond to a recognizable ordering of the labeled objects

Examples

Addresses

Room numbers

Serial numbers (sometimes)

Exit numbers

Focus

Label

Structure

Ordered list

THEORY
Nominal Identifiers

Definition

Assigns unique alpha-numeric identifiers to objects within a single set or category in a manner such that no descriptive information about the individual object is encoded in the identifier

Examples

License plates on cars (within a state)

Drivers license numbers (within a state)

Inventory tracking numbers

Employee (student) IDs

Social Security Numbers (mostly)

Focus

Label

Structure

Flat set

THEORY
Machine-Oriented vs. Human-Oriented

Machine-oriented standards must have these to support compilation

MACHINE-ORIENTED	HUMAN-ORIENTED
Attribute defined classes	Ambiguous categories
Fixed semantics	Indexicality
Formal syntax	Preference orders

Humans don't think and communicate like computers (generally speaking)

RECOMMENDATION
Be clear on intended use

- **THREE DIFFERENT KINDS OF STANDARDS EFFORTS**

1. **Classic Machine-Oriented Standards**
   - **Programs, languages, protocols**

2. **Human/Machine Interface Standards**
   - **Ontologies and bibliographic indexing systems**

3. **Human Oriented Standards**
   - **Categorizations, Dictionaries, Taxonomies, Identifier systems**

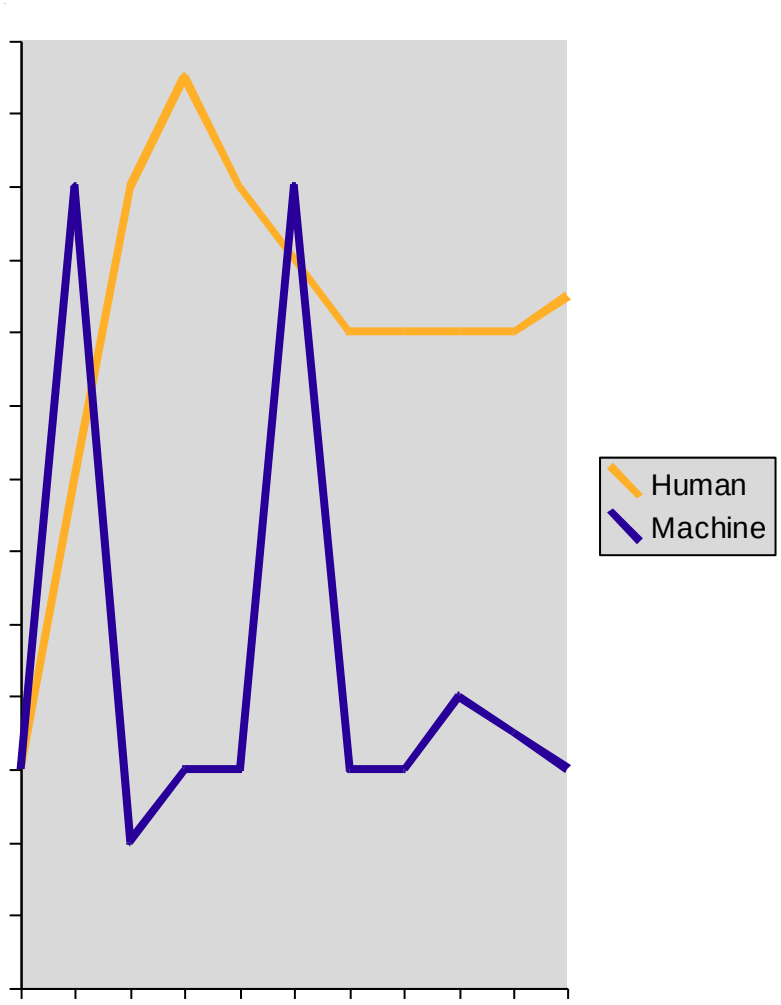
Confuse these goals at your peril

RECOMMENDATION
Delay design choices

- All three major groups have a range of structural types
  - Understand differences among the types
  - Leverage types as design patterns
- Match design patterns against requirements
  - Don’t rush to judgment
  - When in doubt, simplify
- Avoid premature statements like:
  - We need a taxonomy for [fill in blank]
  - We need a CVE for [fill in blank]
  - We need an ontology for [fill in blank]

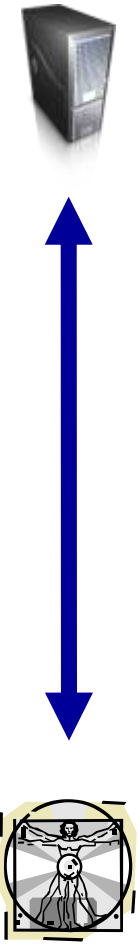
RECOMMENDATION
Accept programmatic differences

- **Machine Oriented**
  - Punctuated releases
  - Build and deploy
  - C, HTML, XML, Java
- **Human Oriented**
  - Ongoing revisions
  - Maintenance tail
  - Dictionaries, SSN, Species names

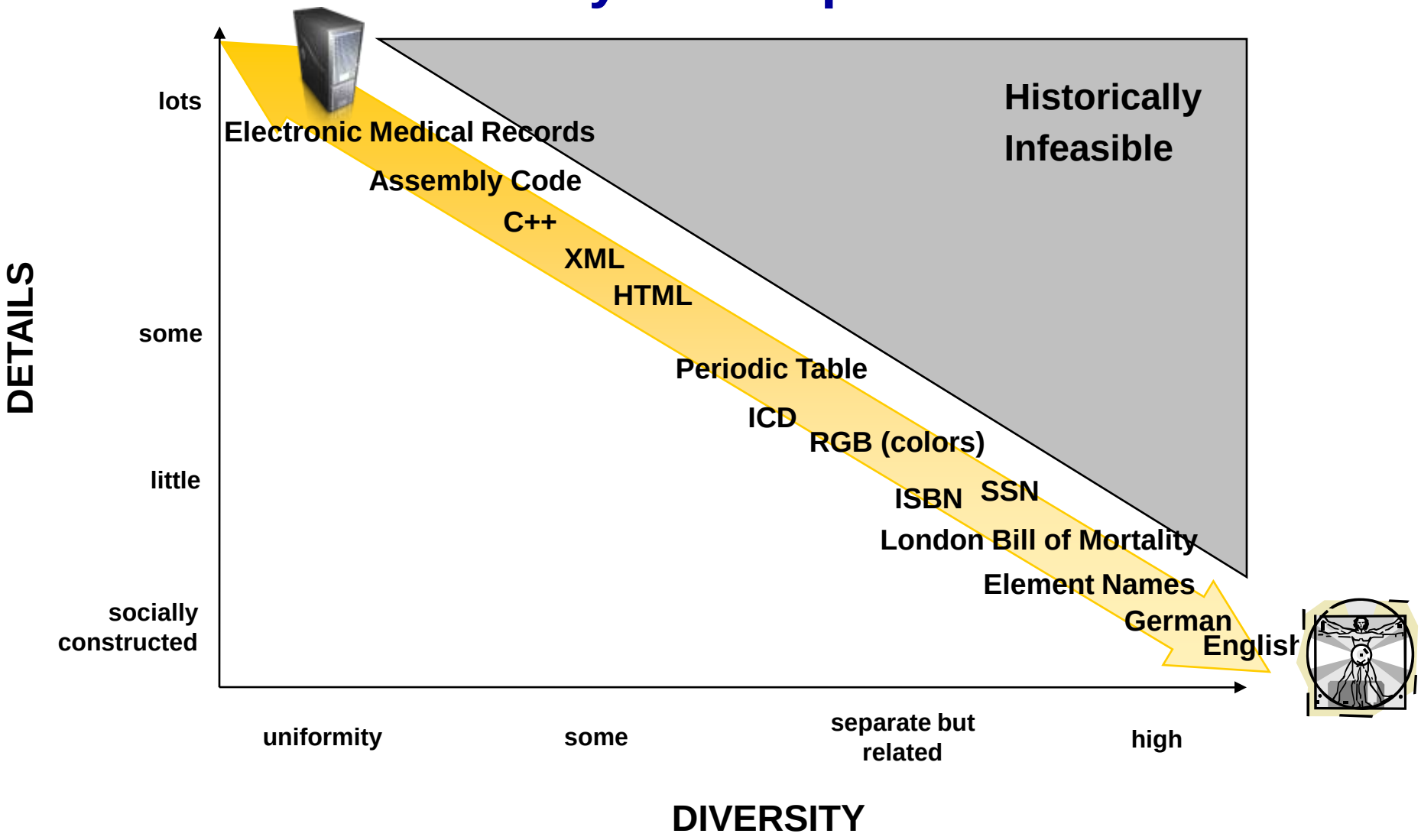


THEORY
Standards Viewed as "The Stack"

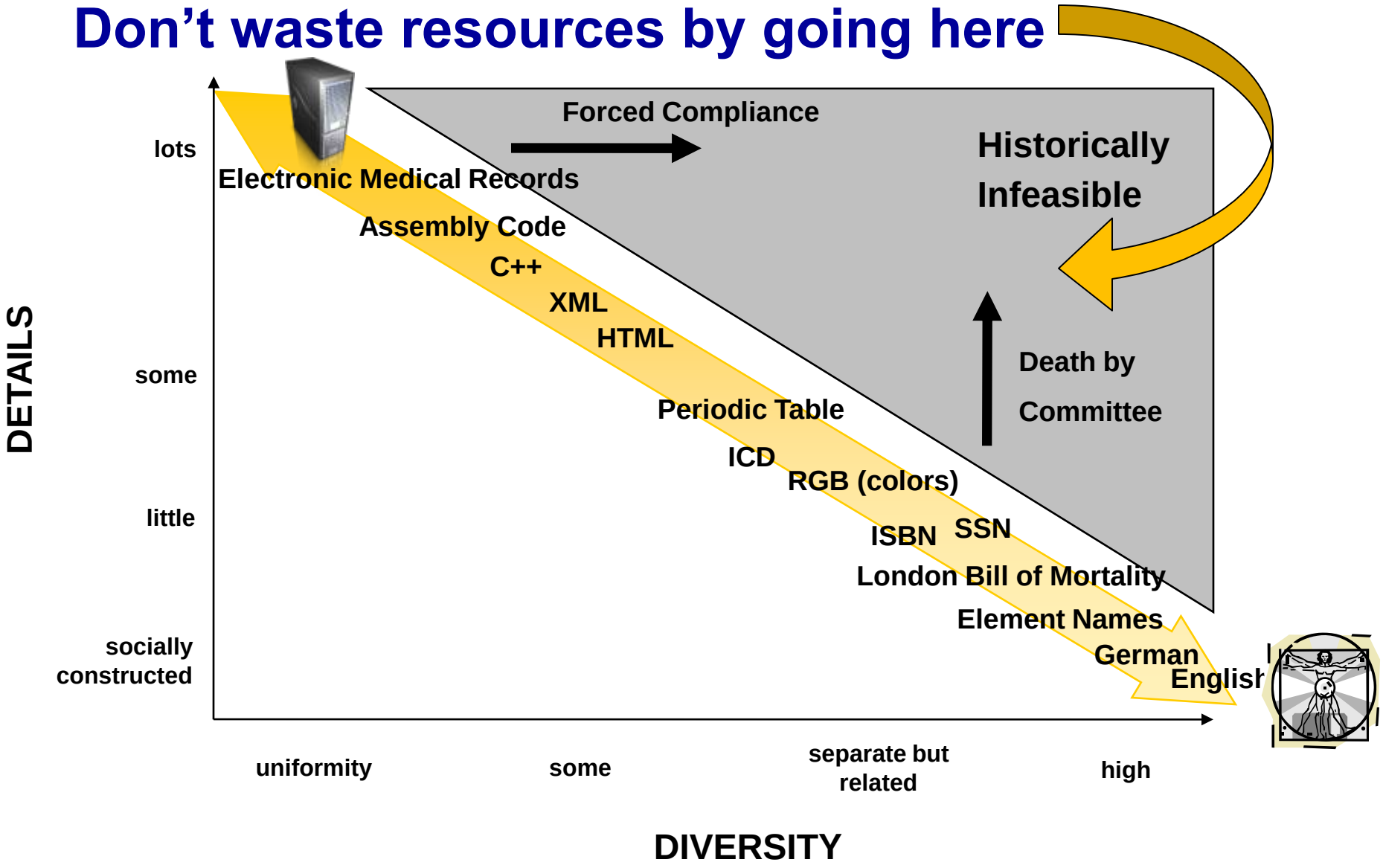
- Data Processing Systems
  - Databases, Applications
- Programming Languages
  - C, C++, Java
- Protocols
  - FTP, HTTP
- Knowledge Structures
  - Schemas, Ontologies
- Classification Systems
  - Species Names
- Controlled Vocabularies
  - Element Names
- Identifier Systems
  - VIN, SSN, ISBN



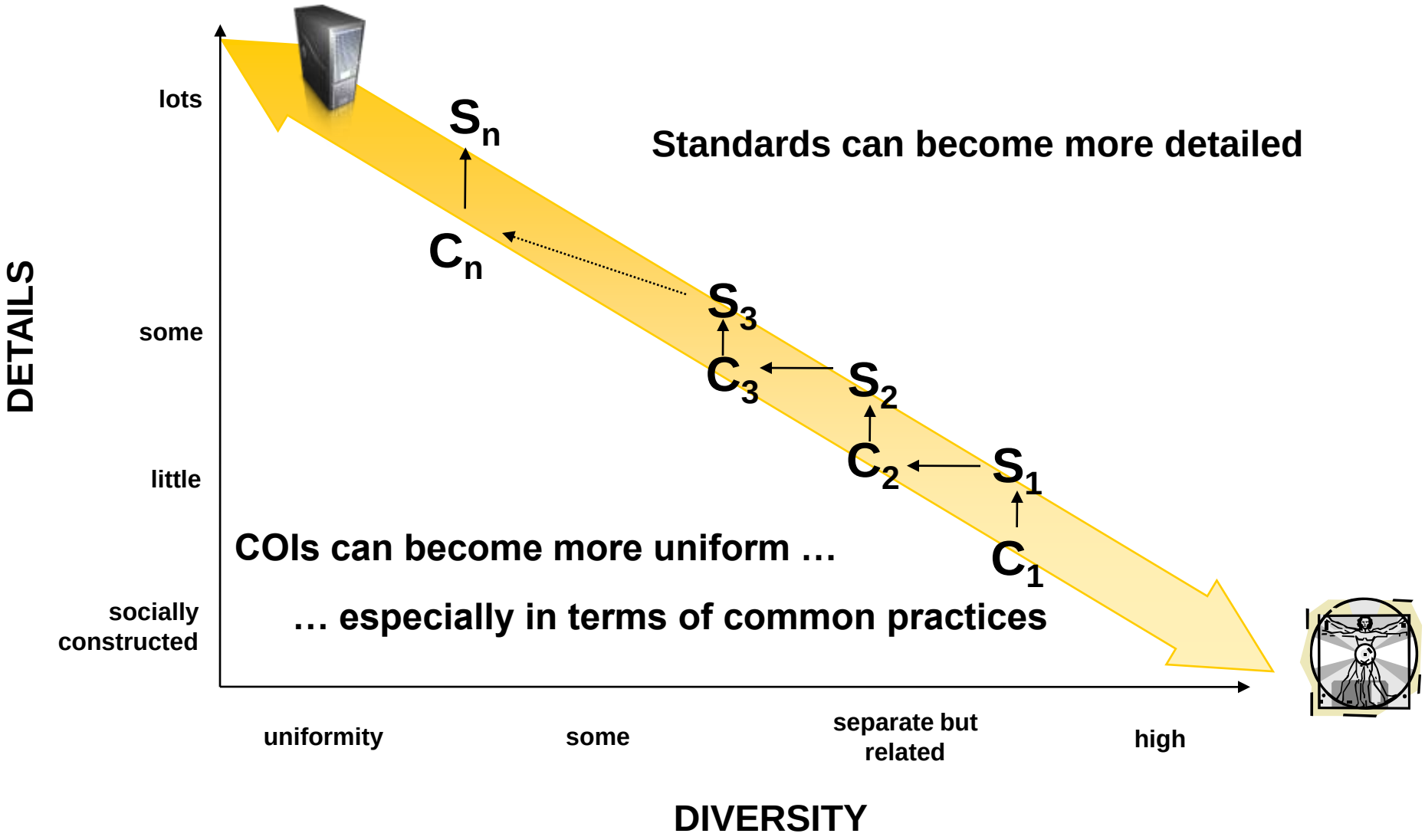
THEORY
The Detail-Diversity Tradespace



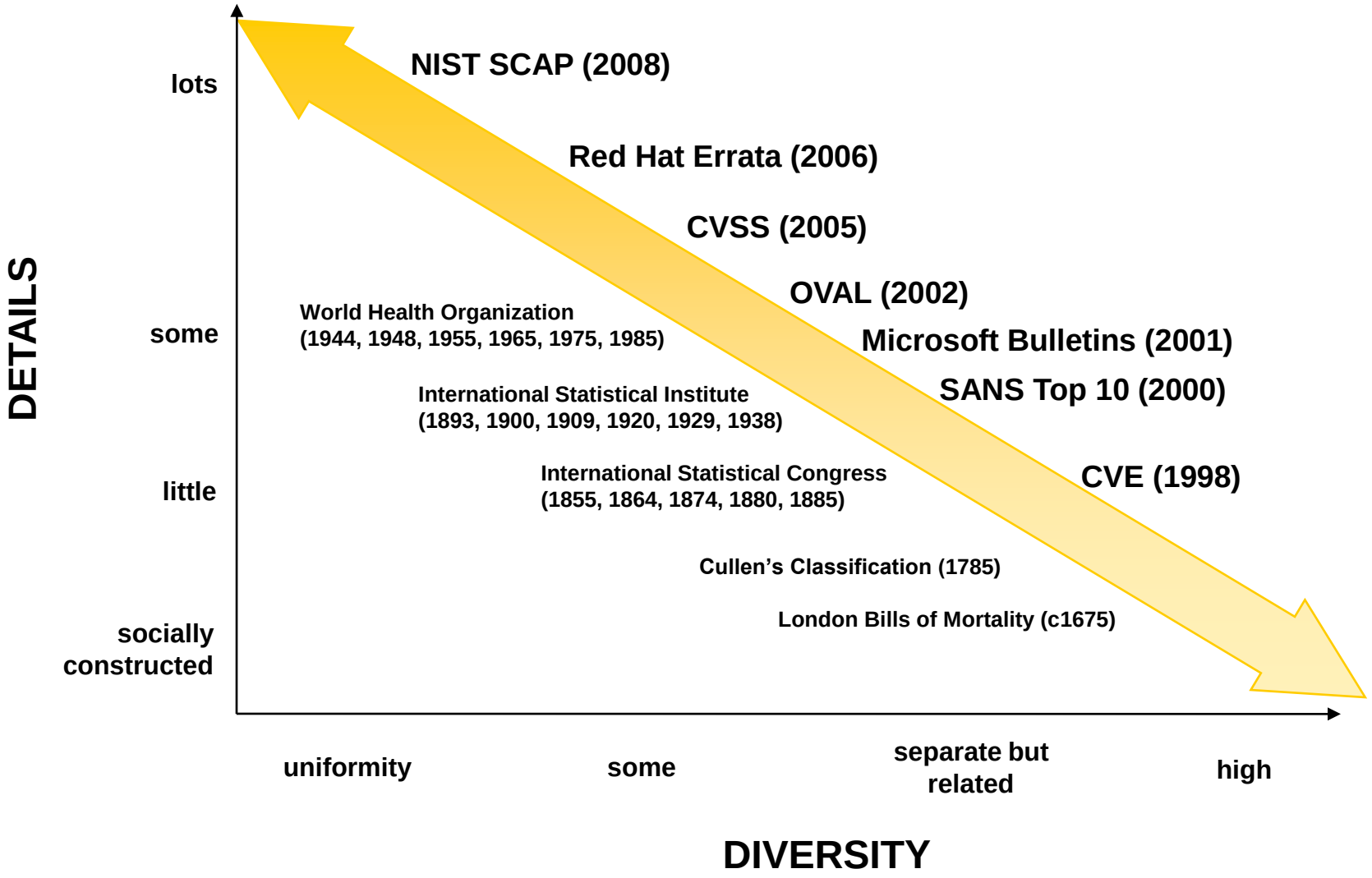
RECOMMENDATION
Don’t waste resources by going here



THEORY
Co-Evolution of Standards and COIs

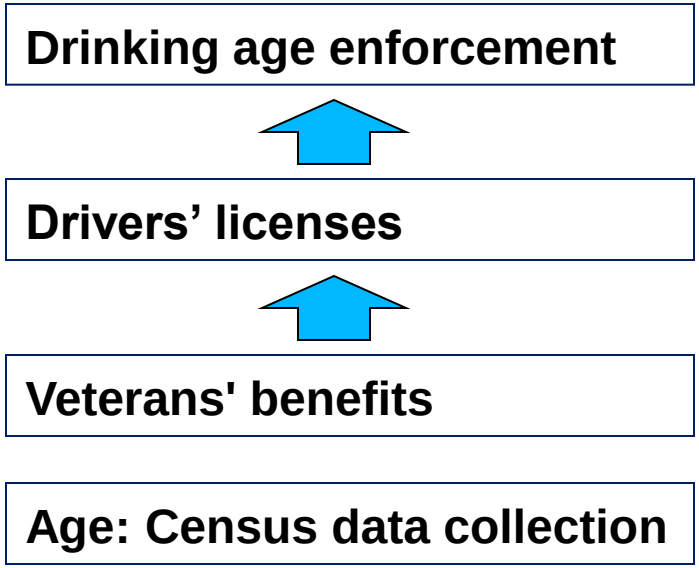


THEORY
Two Examples of Co-Evolution

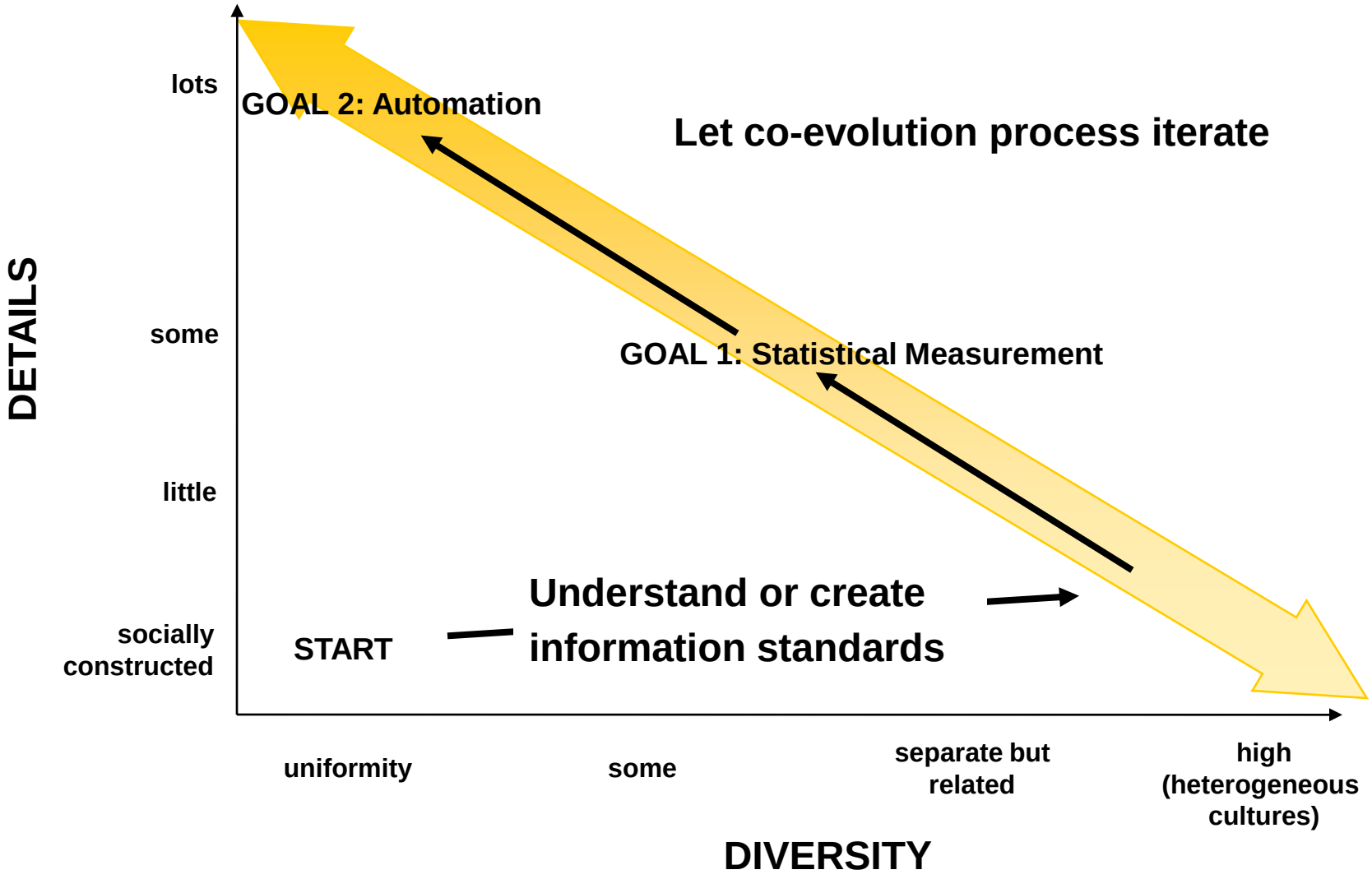


THEORY
The Solidification of Standards

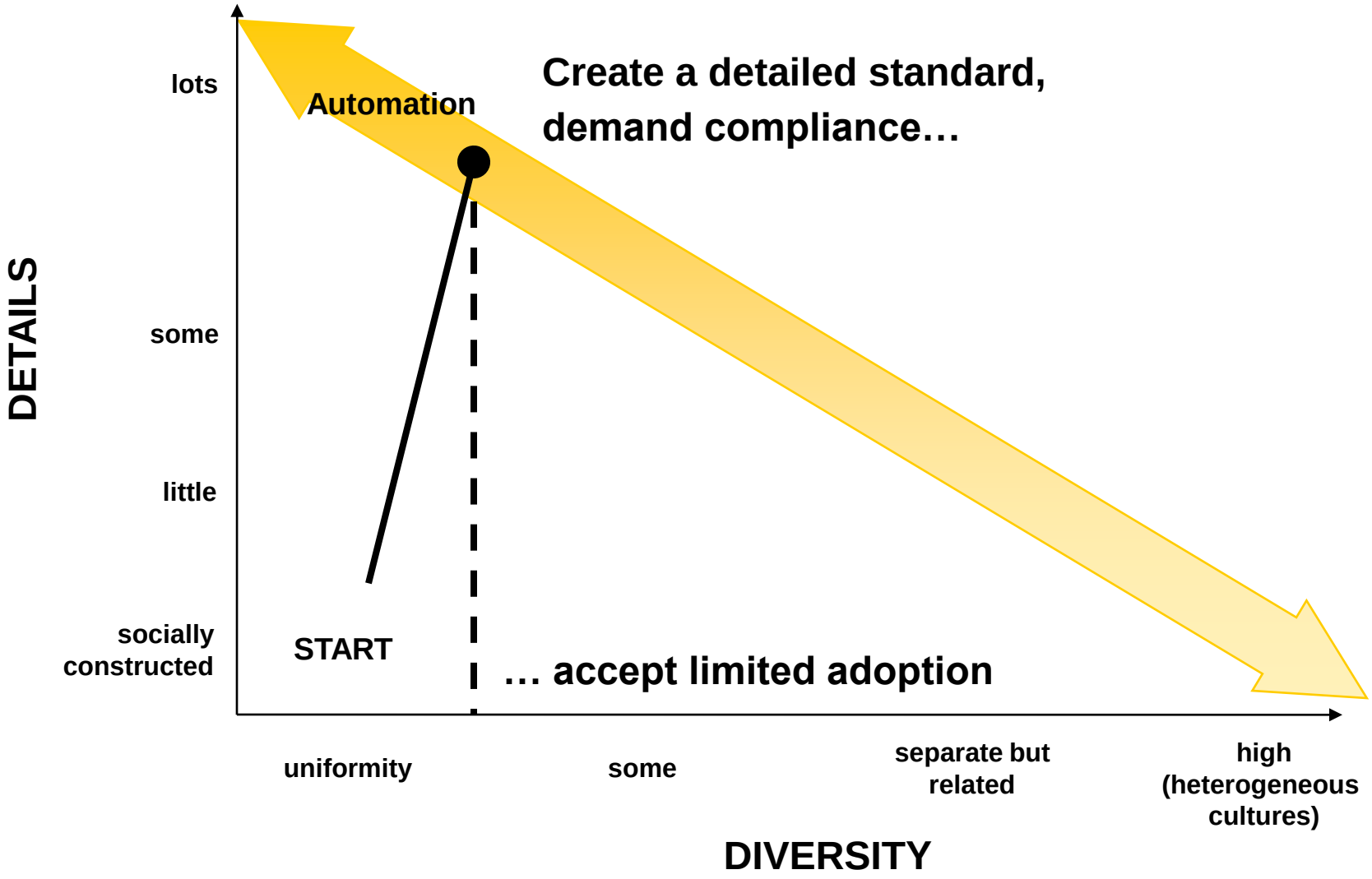
- Standards can build on other standards
- "Russian doll" effect
- This "locks in" base standard



RECOMMENDATION
Take the long view



RECOMMENDATION
If immediate interoperability is needed (1)



RECOMMENDATION
If immediate interoperability is needed (2)

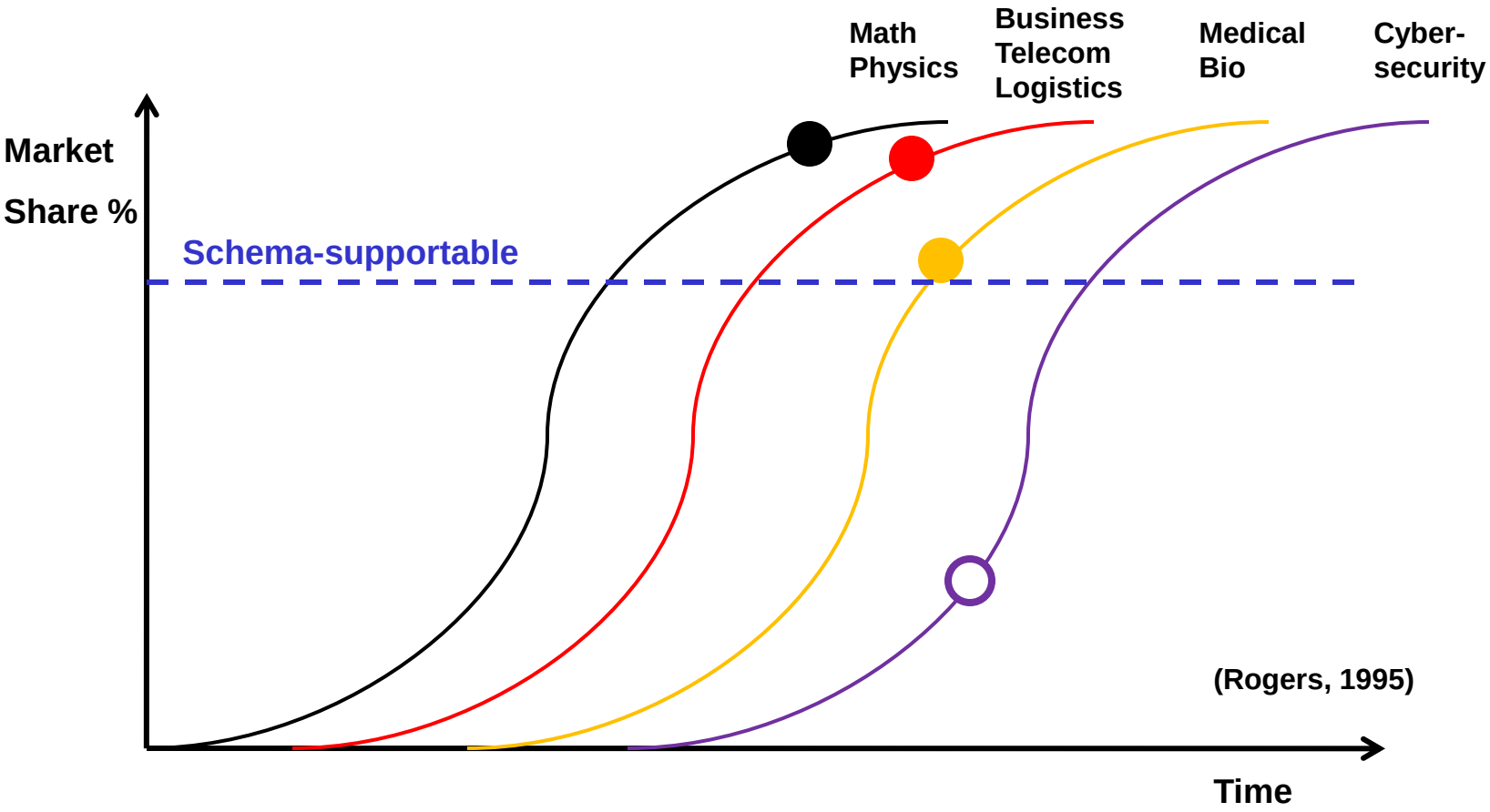
- Accept that interoperability
  - Demands high detail
  - Tolerates little diversity
- Marshal political and economic power
  - Expect vigorous push back on details
  - Be ready to limit participation to reduce diversity
  - Expect inconsistent statistics
- Anticipate split between proprietary solution and market
  - “Russian doll” effect will “lock in” your design choices
  - External standards may evolve differently
  - Expect retooling costs to adopt market standards later on
  - Explains traditional Federal/Commercial market split

THEORY
Success in Applied Computation

- Why has computation succeeded so well in some disciplines and struggled in others?

- Easy and early
  - Arithmetic / Physics
- Doable and later
  - Business (IBM), Communication (AT&T), Logistics (WWII)
- Hard and current
  - Medical, Bio/Genome, AI
- Really, really hard (…or impossible?)
  - Cyber-security management

THEORY
Adoption of Innovations



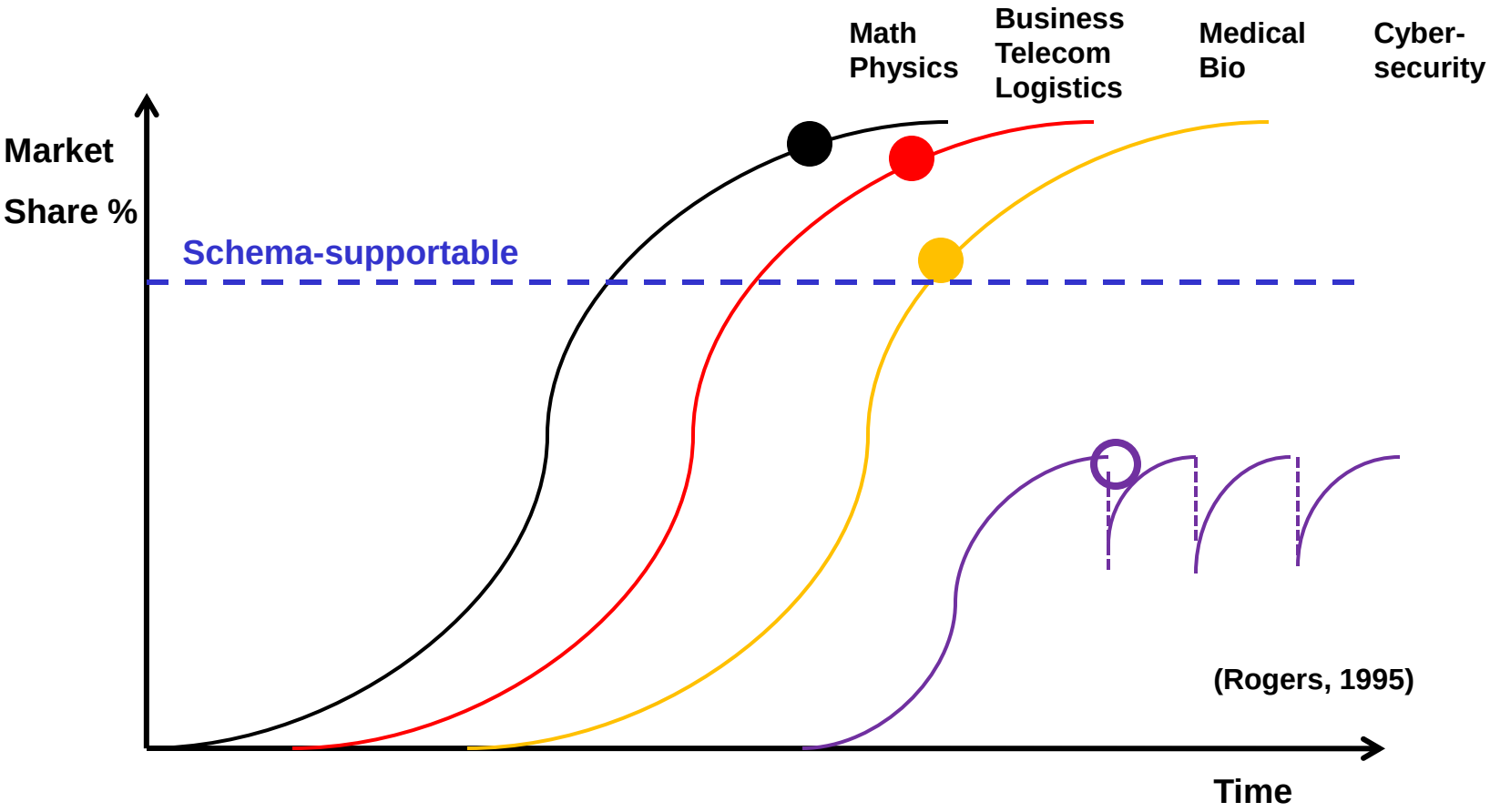
Schema supported by a field only after adoption of innovations has stabilized

THEORY
Three Factors of Stability

- **Professionalization – Credentialing, standardized procedures and methods, standardized literature**
- **Balance of power – Asymmetries of control among stakeholders over resources, funding, environment, strategy, etc.**
- **Infrastructure – Foundational base of technology and related practices supporting the field**

- **Stable schema would only work if:**
  1) **Established profession agrees on appropriateness**
  2) **Negotiated truce among stakeholders about power**
  3) **Technology holds still long enough**

THEORY
Adoption of Innovations – Redux



Continuous innovation renders new technology adoption unstable; limits effective schema development & use

MITRE

RECOMMENDATIONS
Some Cautionary Assertions

- **Cyber-security standards are comparatively unstable**
  - **Relative lack of professionalization (and of stable concepts)**
  - **Unresolved power relationships**
  - **Rapidly changing technologies undermine infrastructural foundation**

- **Cyber-security "schema" efforts are attempts to replace traditional human-oriented standardization process**
  - **Short-*cuts* turn into short-*circuits***

- **Machine-oriented standards are not well suited to organize human thought**

RECOMMENDATIONS – Summary

- Think in social and organizational terms
  - Think about how information standards will be used
  - Distinguish between machine-, interface- & Human-oriented standards
- Delay design choices: Use known types as design patterns
- Accept programmatic differences in funding cycles
  - Machine-oriented standards: funding spikes for releases
  - Human-oriented standards: on-going production costs
- Don't waste resources in the "Zone of Infeasibility"
  - Avoid forced compliance and death by committee
- Invest in human-oriented standards for long term agreement
- Limit involvement to achieve faster interoperability
  - Marshal political and economic power to hold your ground
  - Anticipate splitting the market and possible retooling costs
- Beware of persistent factors of instability

White Papers

- Anne Rawls, David Mann, Angela Garcia, Gary David, Matt Burton, “Simple Enumerations. Ethnomethodology and MITRE Information Assurance Data Standards”, Il Mulino, 2009. - Issue 1 : Vol. Etnografia e ricerca qualitativa.

- David Mann and JoAnn Brooks, “Information Standards and Their Use: Implications for Design Patterns”, Case: 10-1789, The MITRE Corp, 2010.

- David Mann and JoAnn Brooks, “The Relationship between Human and Machine-Oriented Standards and the Impact to Enterprise Systems Engineering”, Case: 10-2335, The MITRE Corp, Forthcoming.