

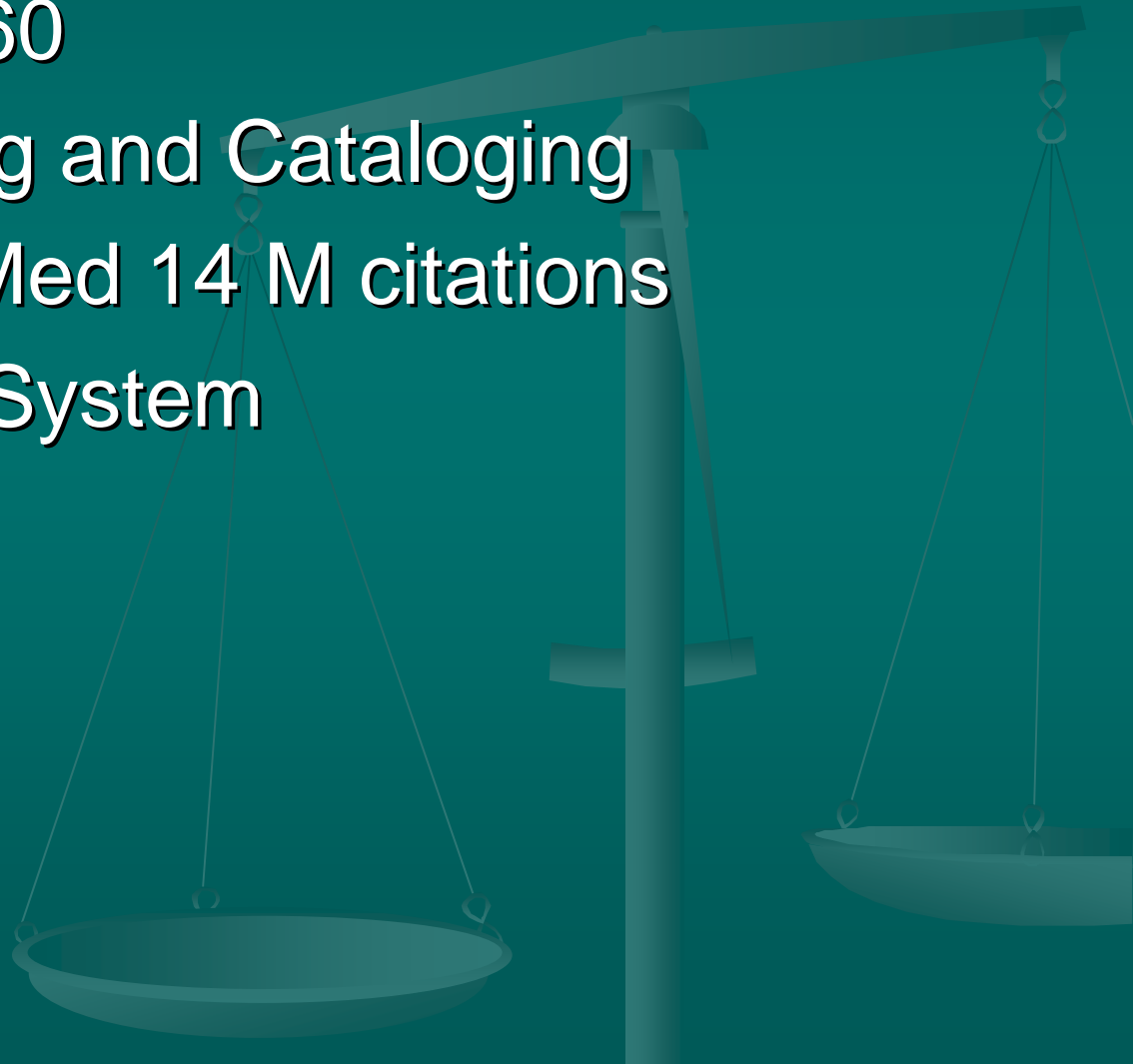


# ***Vocabulary Development and Maintenance***

The Example of MeSH

# The Medical Subject Headings

- First Edition 1960
- Used in Indexing and Cataloging
- MEDLINE/PubMed 14 M citations
- Early Retrieval System



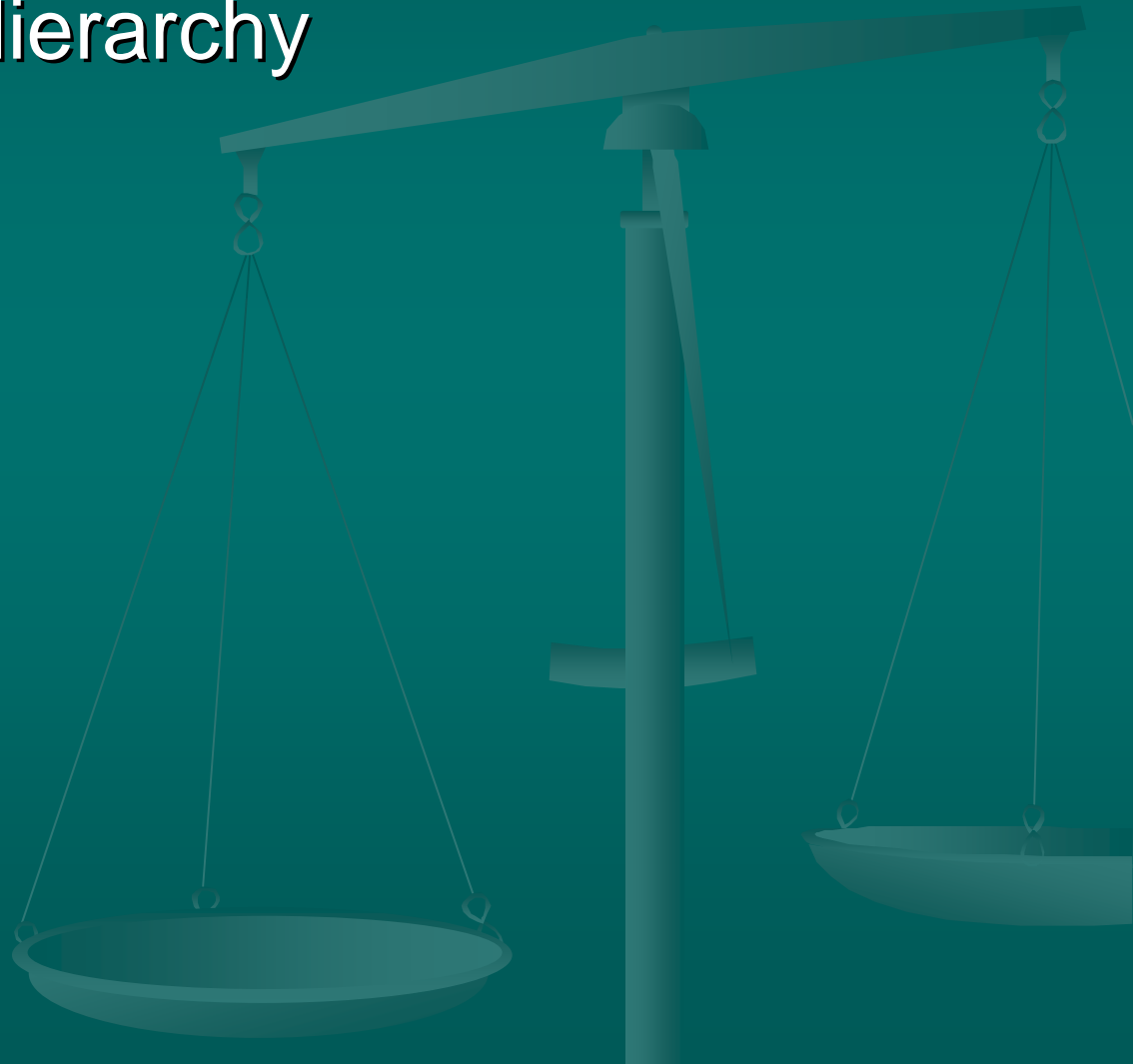
# Prescient Design

- Multiple Topics
- Simple Information Model
  - What Is Article About?
- Anticipated Databases



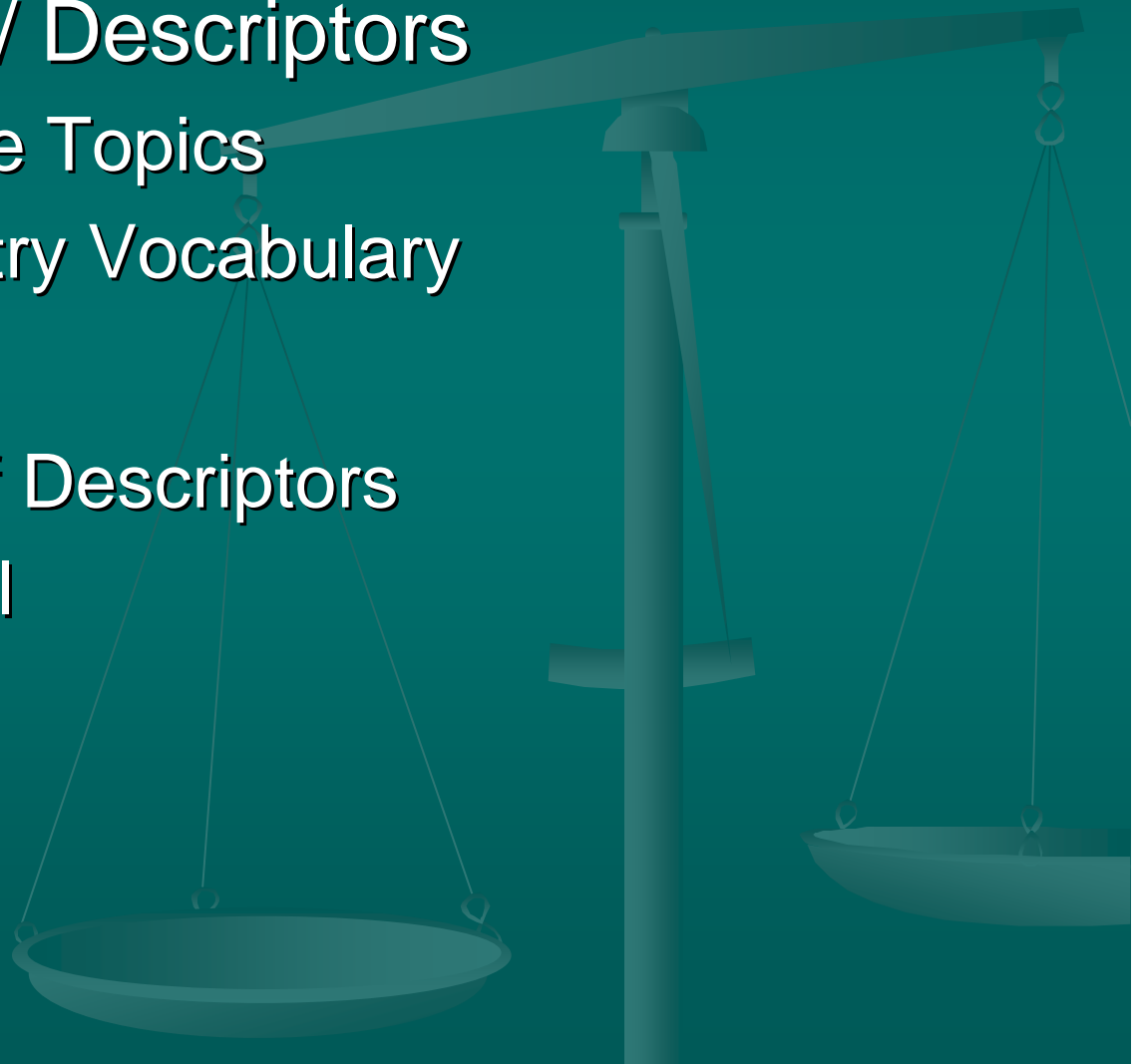
# Evolving Structure

- Headings in a Hierarchy
- Qualifiers
- Supplementals



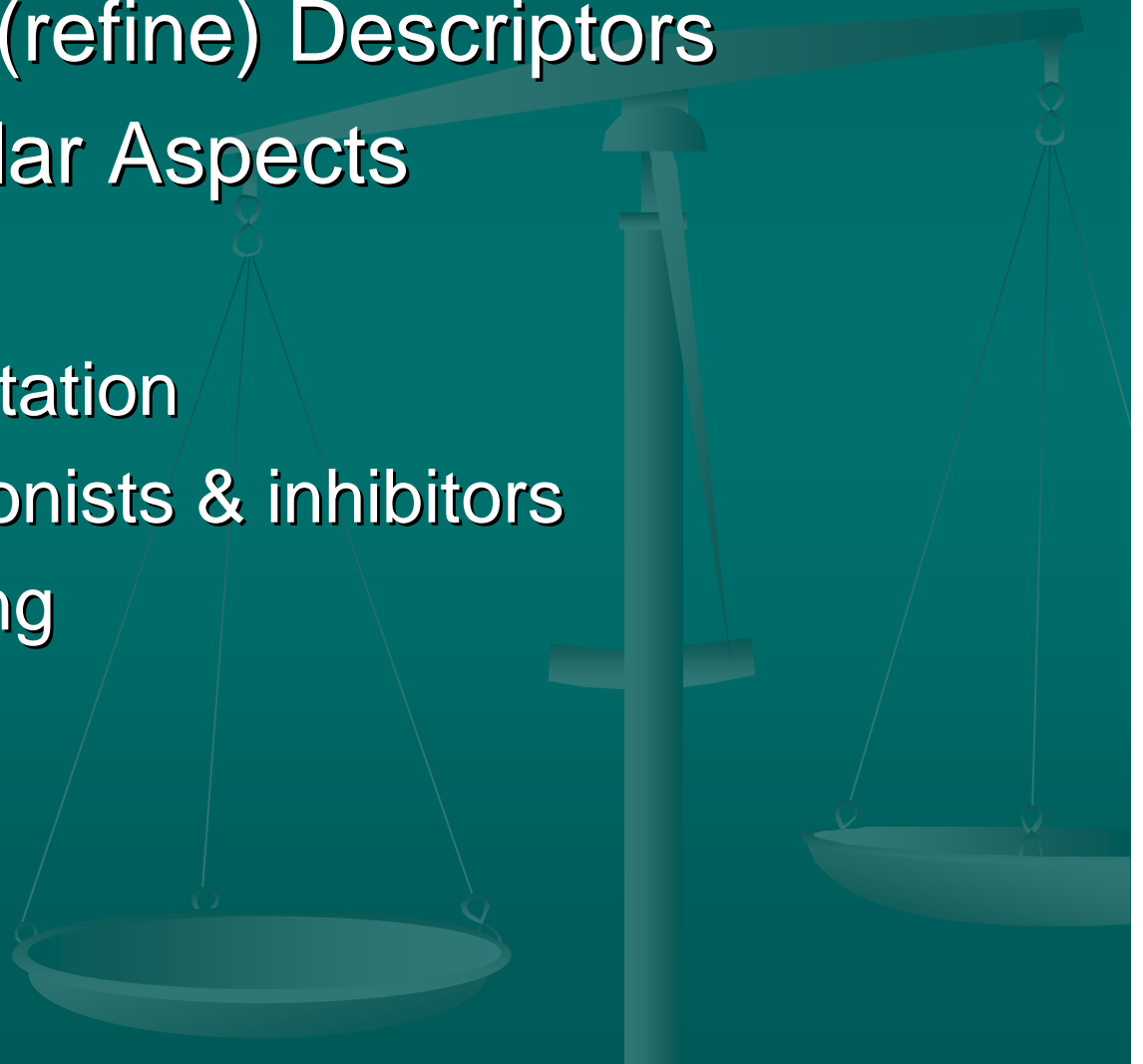
# Headings in a Hierarchy

- Main Headings / Descriptors
  - Used to Indicate Topics
  - The Role of Entry Vocabulary
- Hierarchies
  - Arrangement of Descriptors
  - Polyhierarchical
- 22,000



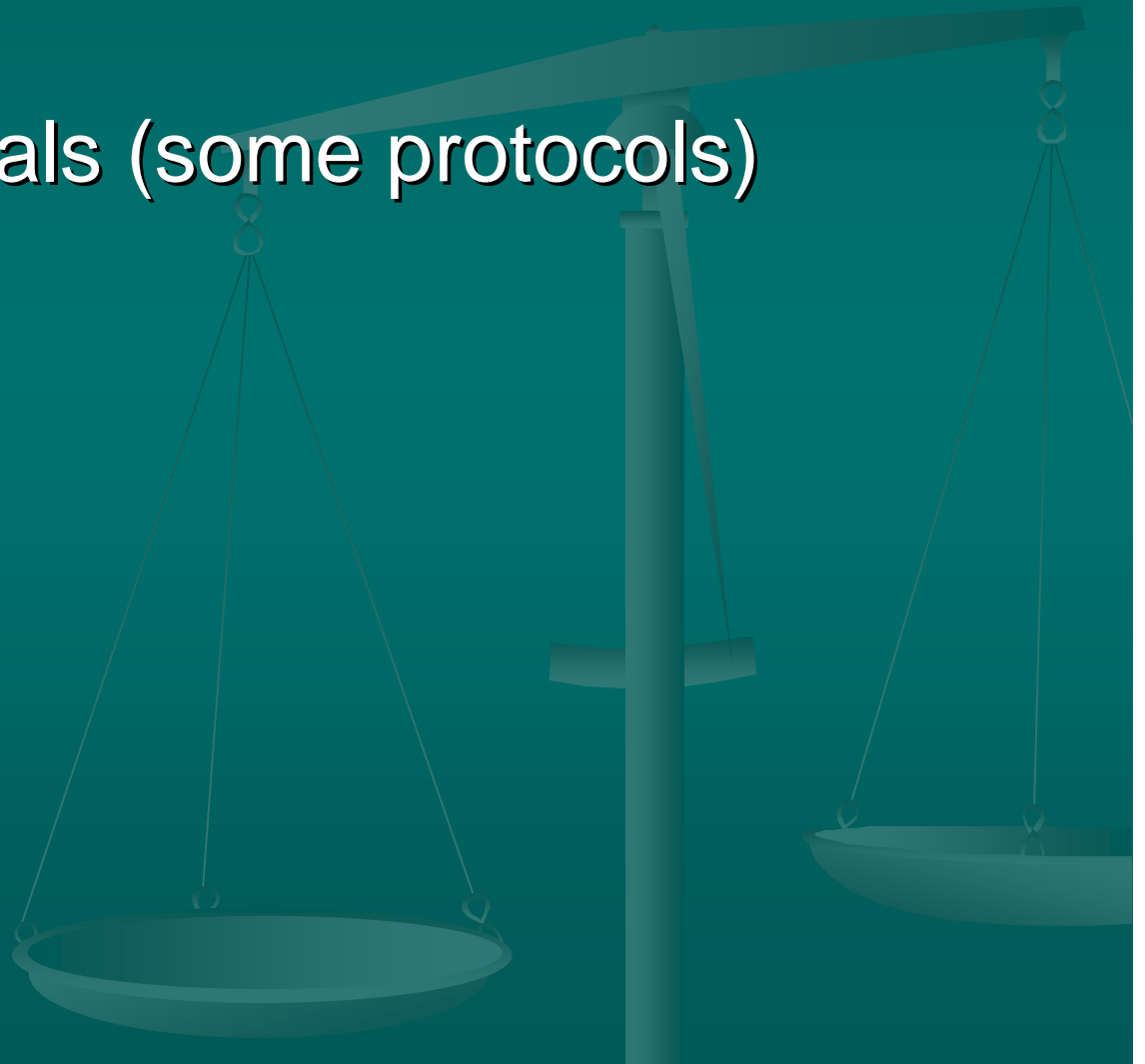
# Qualifiers

- Used to Modify (refine) Descriptors
- Indicate Particular Aspects
- Examples
  - Heart/transplantation
  - Calcitriol/antagonists & inhibitors
  - Aspirin/poisoning
- 87



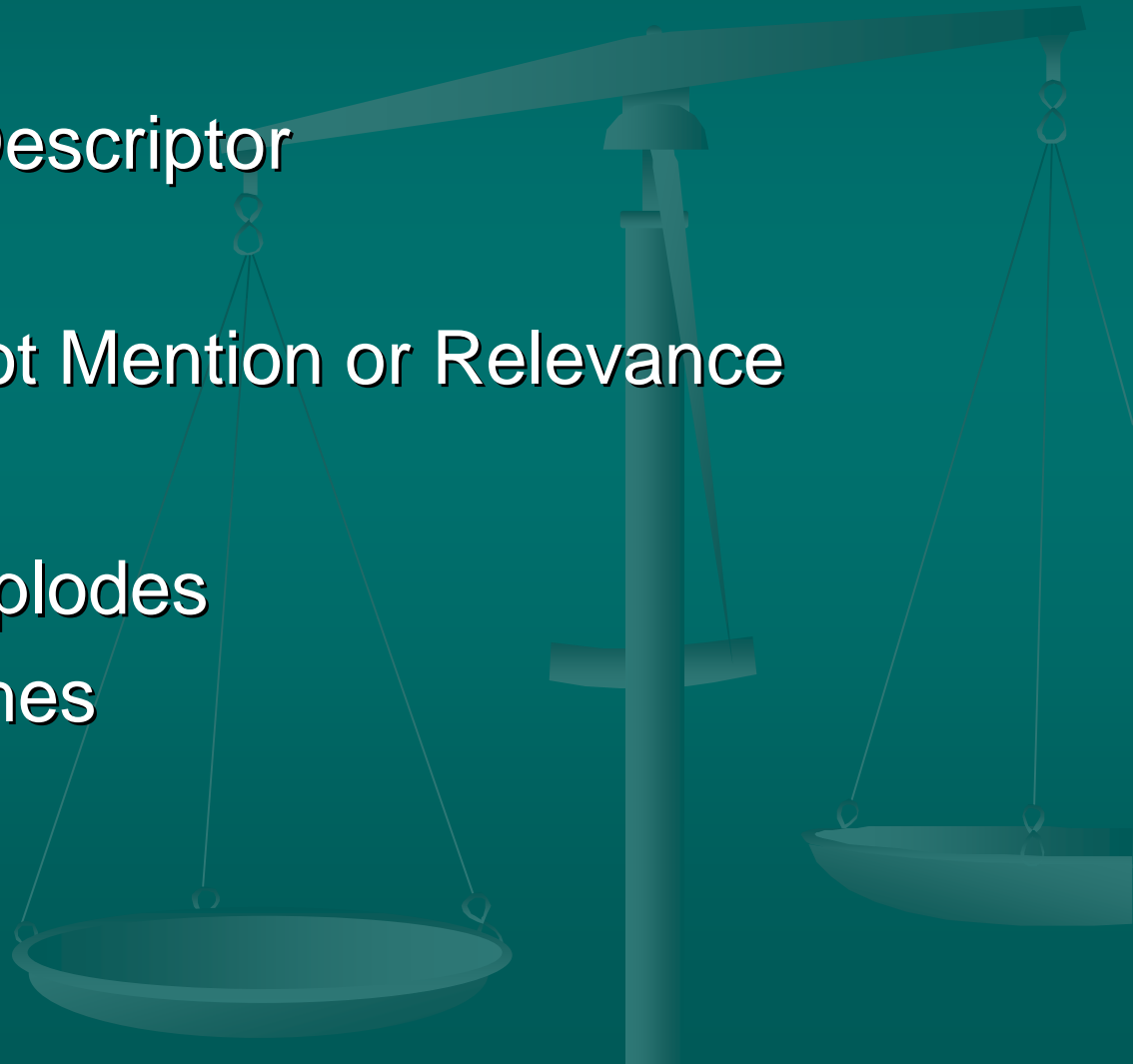
# Supplementals

- Added Daily
- Largely Chemicals (some protocols)
- 140,000



# Use of MESH

- Indexing
  - Most Specific Descriptor
  - Coordination
  - "Aboutness", not Mention or Relevance
- Searching
  - Hierarchical Explodes
  - Boolean Searches



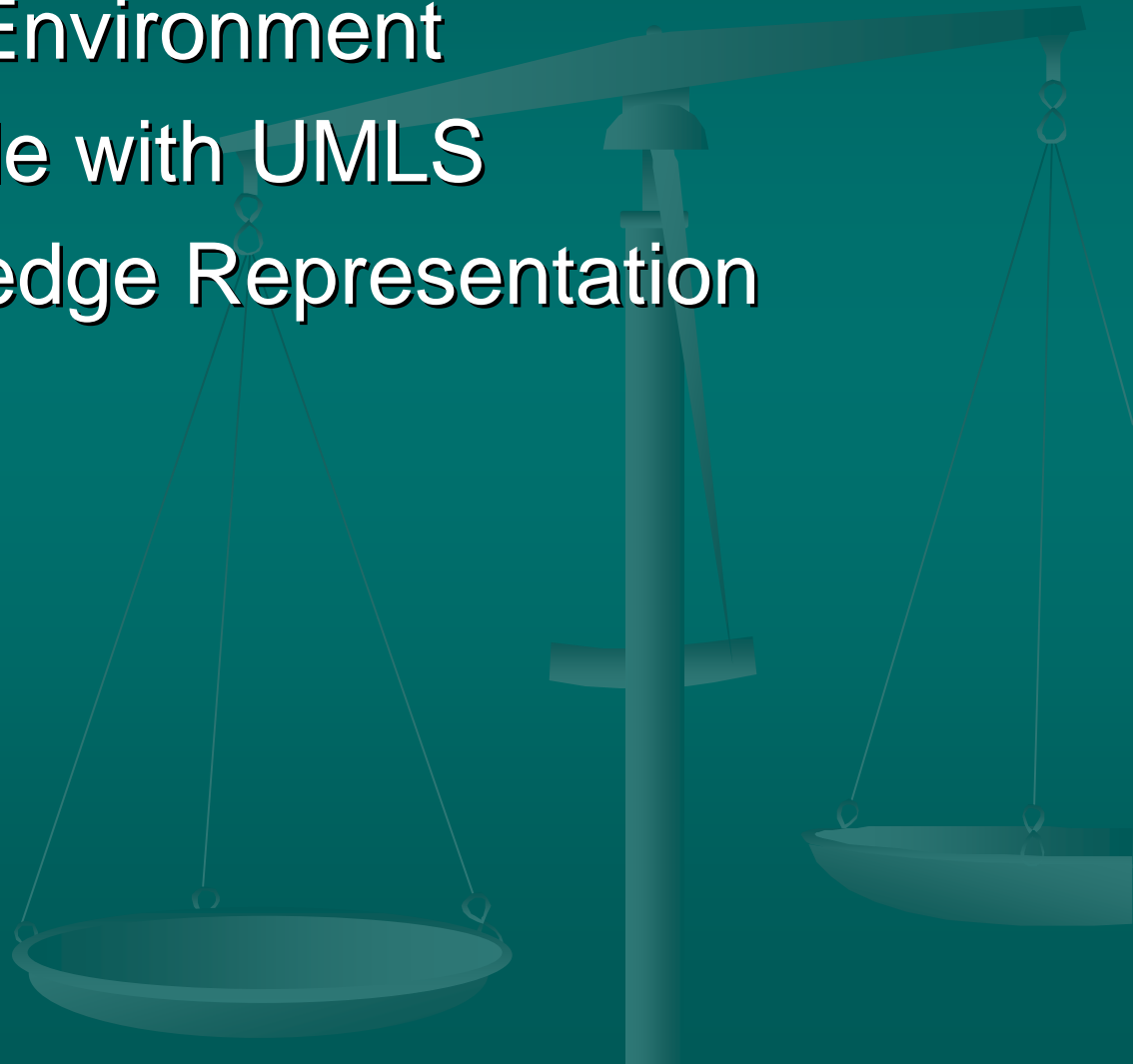
# Information Model

- What Citation Is About
- Topical Indexing
- Descriptive Indexing
  - Check Tags
  - GenBank Numbers
  - Grant Numbers
  - Support



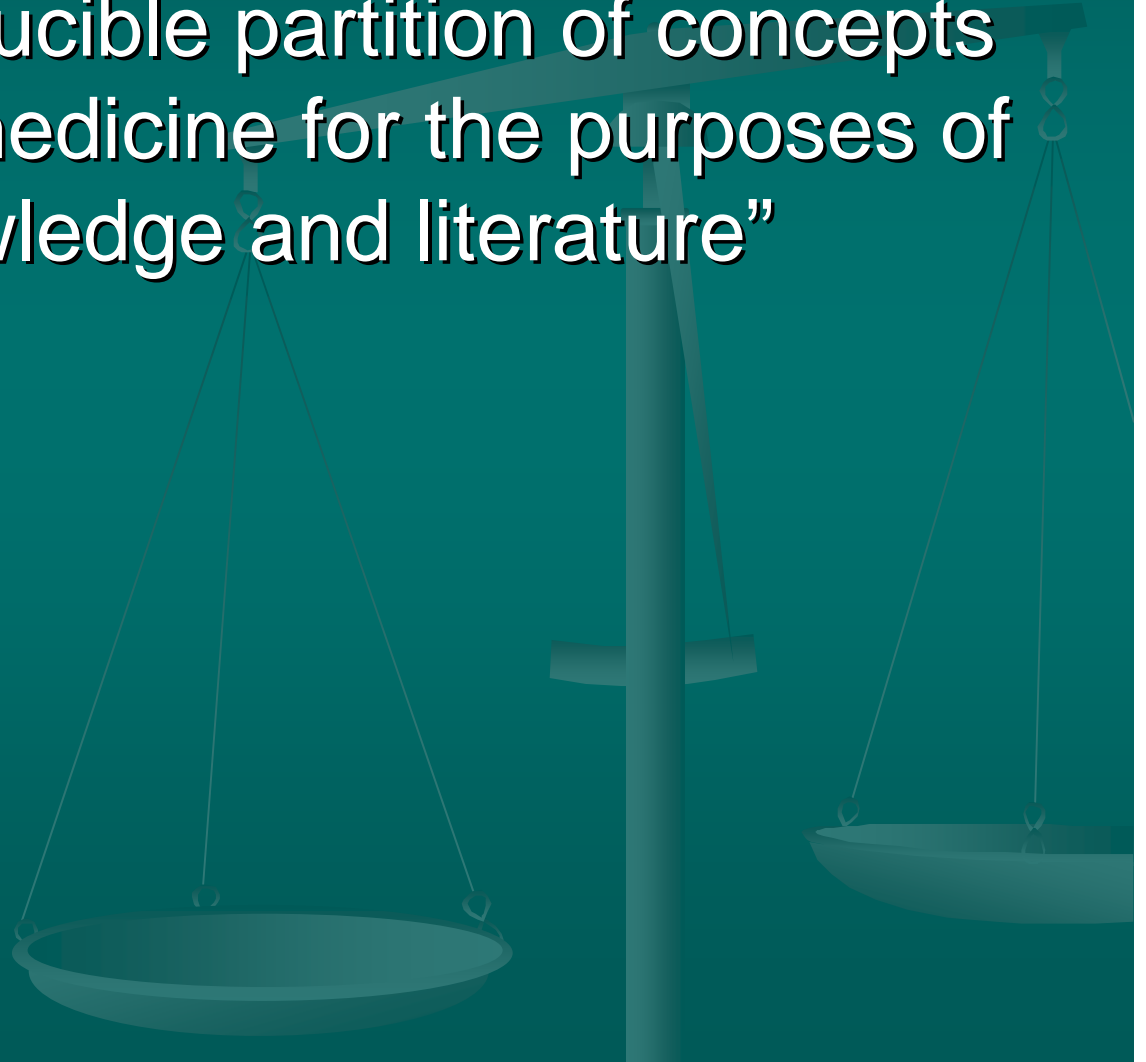
# The Challenge

- Off Mainframe Environment
- More Compatible with UMLS
- Address Knowledge Representation Critiques



# Mission Statement

“Provide a reproducible partition of concepts relevant to biomedicine for the purposes of organizing knowledge and literature”



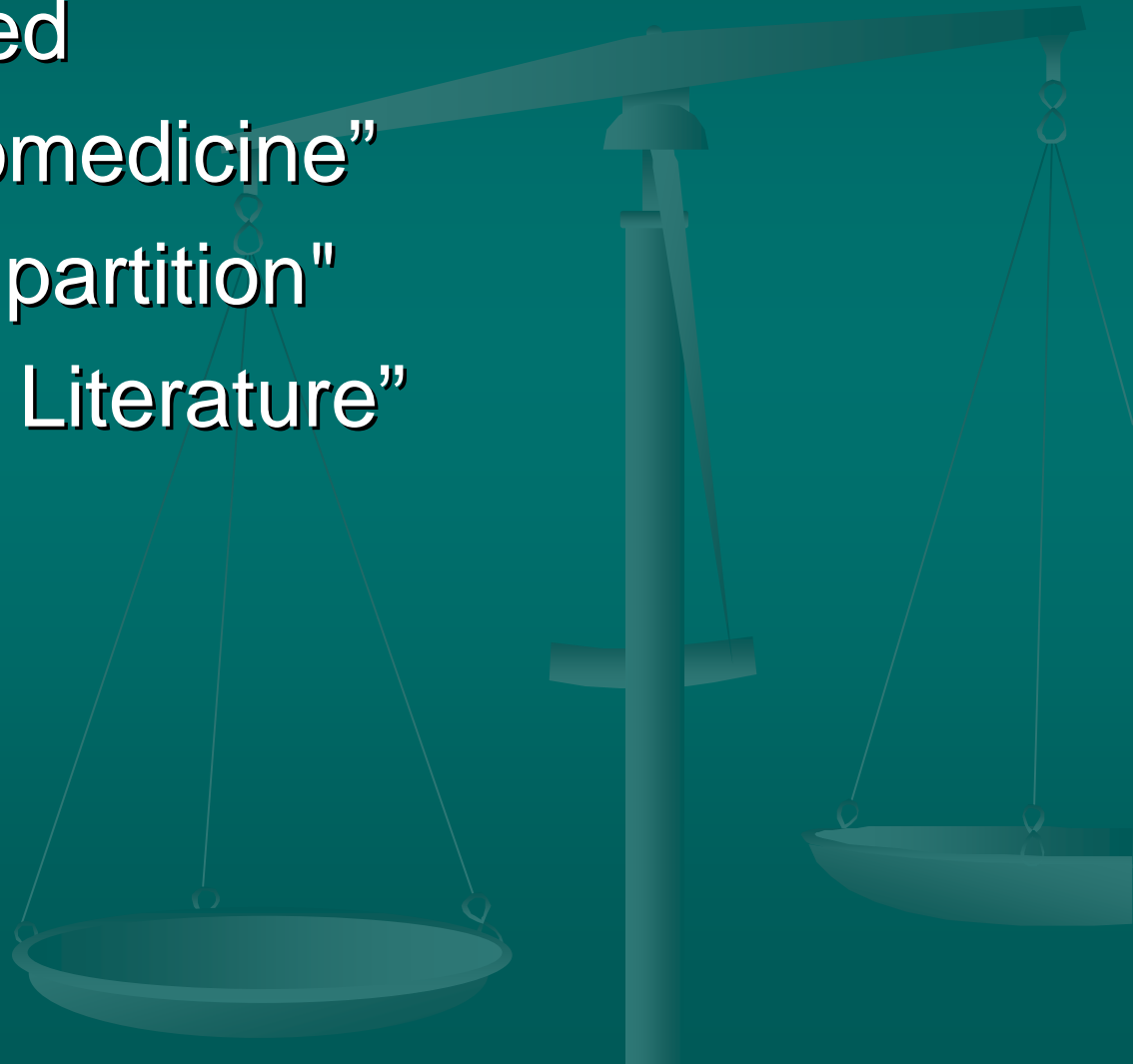
# Reproducible

- Understandable
- Consistent - predictable
- Current
- Valid



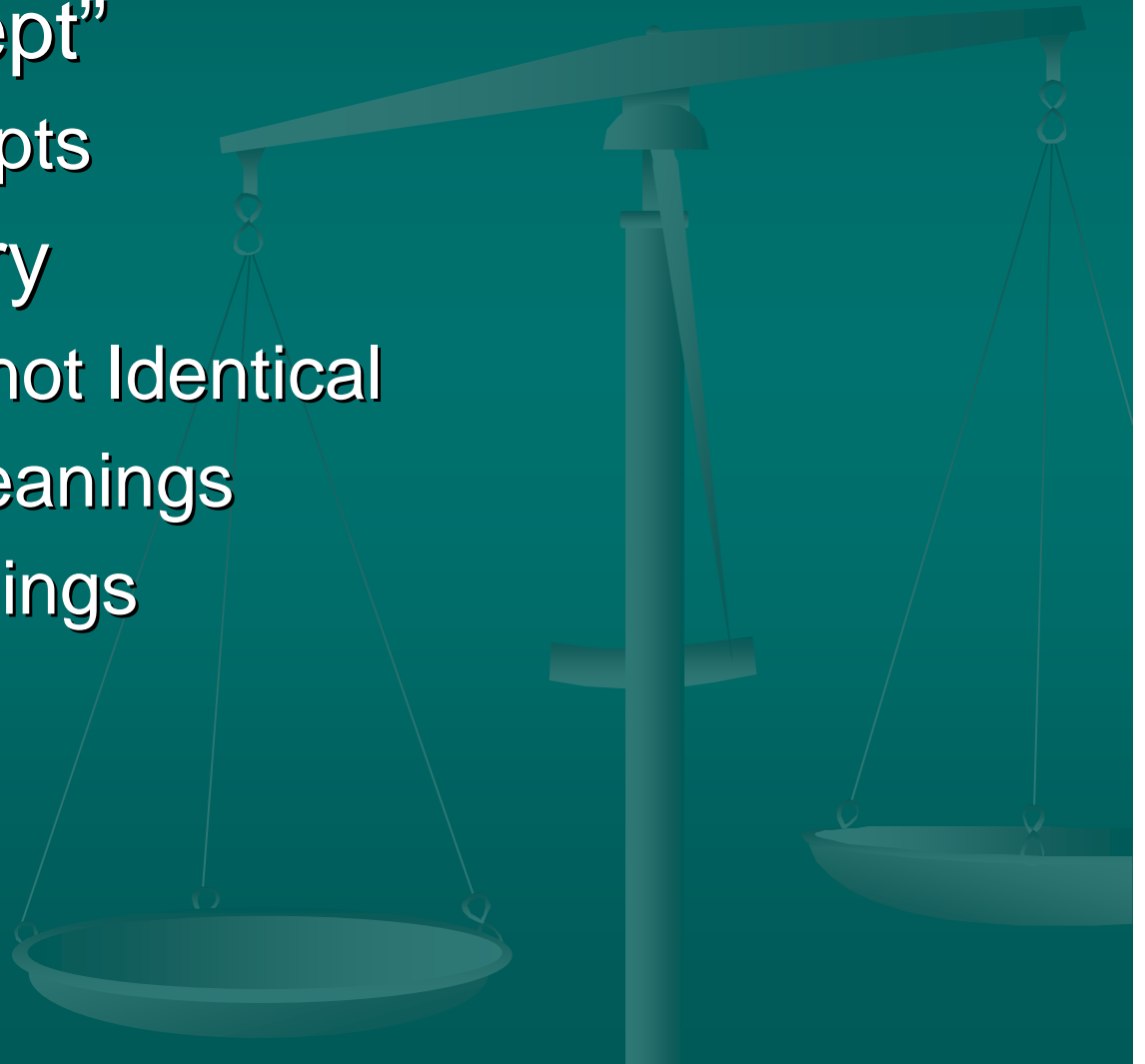
# Other Notions

- Concept Oriented
- “Relevant to Biomedicine”
- "A reproducible partition"
- "For Organizing Literature"



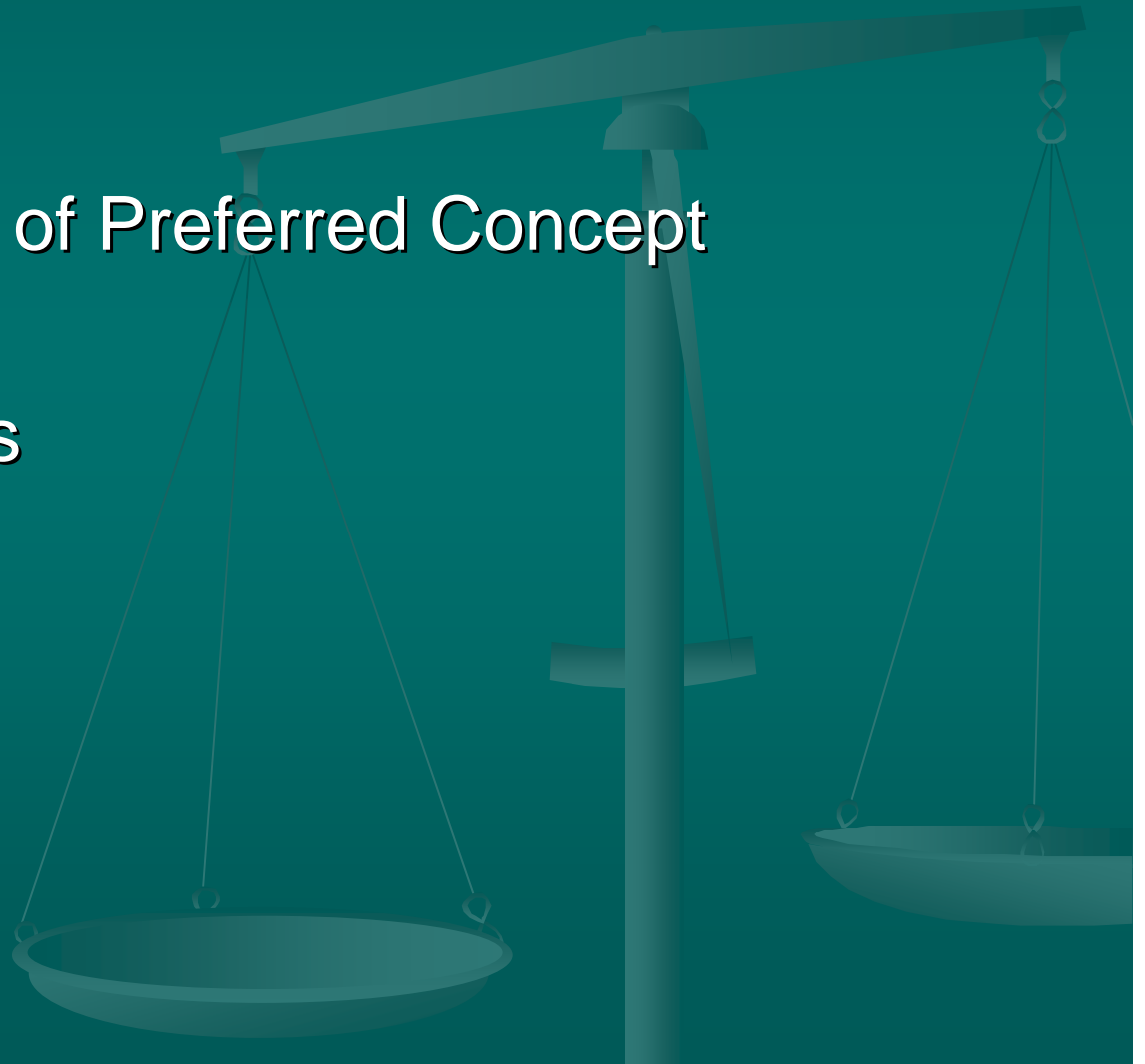
# What Is a Descriptor?

- A “Fuzzy Concept”
  - Class of Concepts
- Entry Vocabulary
  - Equivalent but not Identical
  - Overlapping meanings
  - Narrower Meanings



# Three Levels of Objects - Attributes at Each Level

- Descriptors
  - Hierarchies
  - Preferred Term of Preferred Concept
- Concepts
  - Semantic Types
  - Definitions
- Terms
  - Spelling
  - Lexical Tags



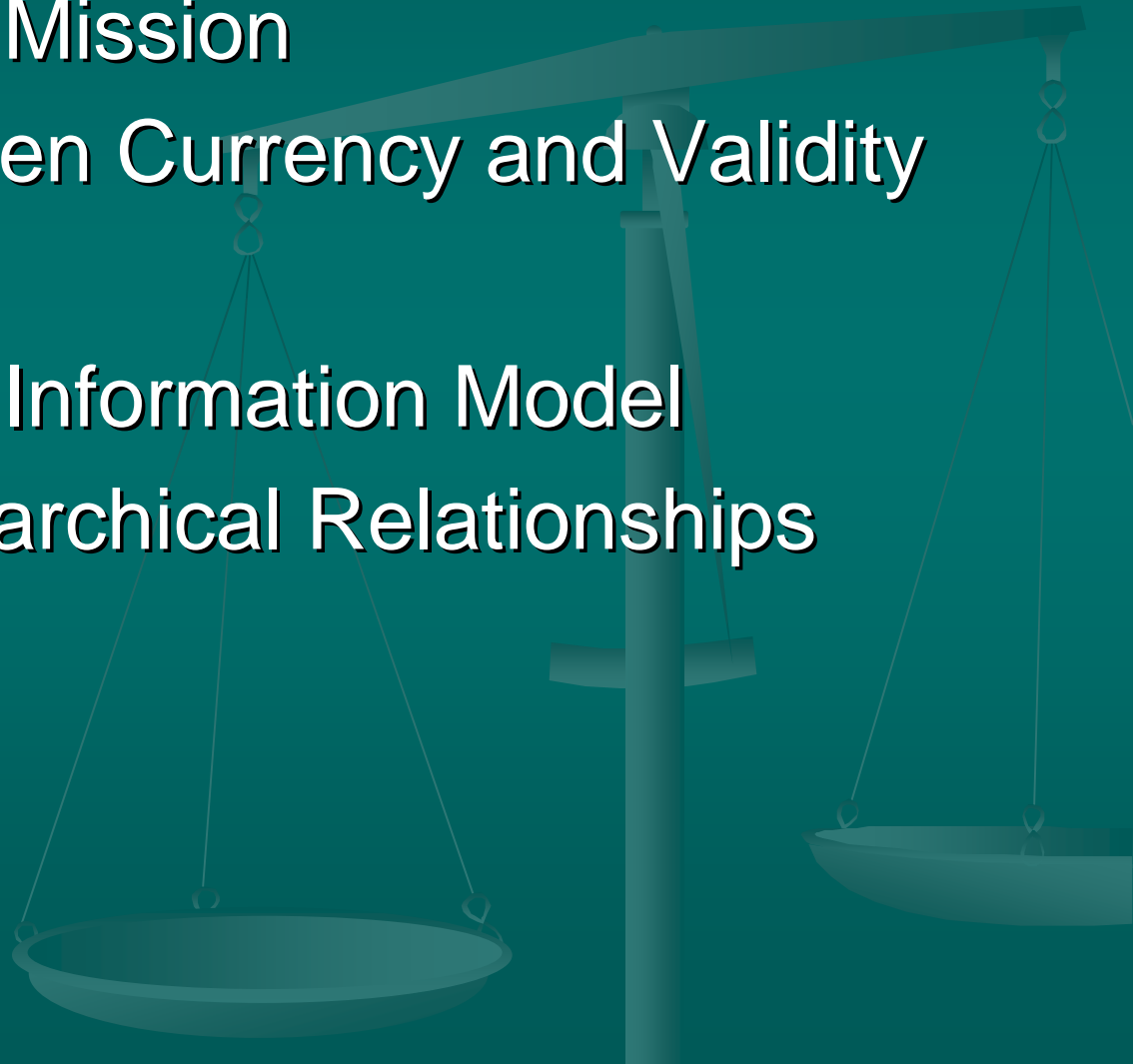
# Workflow

- Sources of Changes
- Analyst
- Internal Review
- External Review



# Criteria for Changes

- Consistent with Mission
- Trade-off between Currency and Validity
- URU Criteria
- Consistent with Information Model
- Soergelian Hierarchical Relationships



# Differing Viewpoints

- Users
- Indexers
- Search Engine
- Citation Maintenance



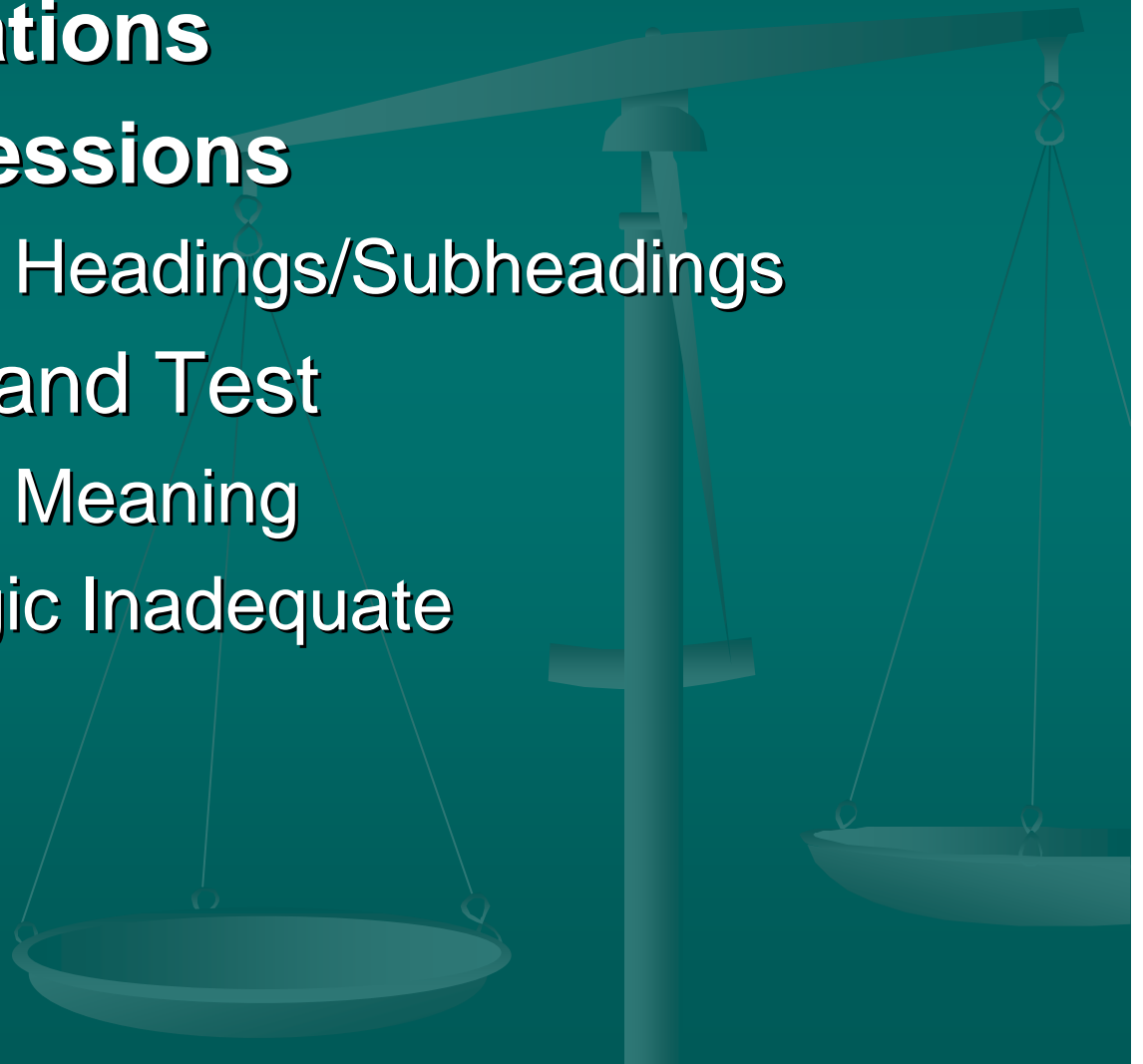
# Representational Integrity

- One Meaning - One Place
- More than Non-redundancy
- No Unrecognized Synonymy
- No Unrecognized Expressions



# Representational Integrity

- **Entry Combinations**
- **Complex Expressions**
  - Coordination of Headings/Subheadings
- **How to Identify and Test**
  - Each Citation a Meaning
  - Description Logic Inadequate

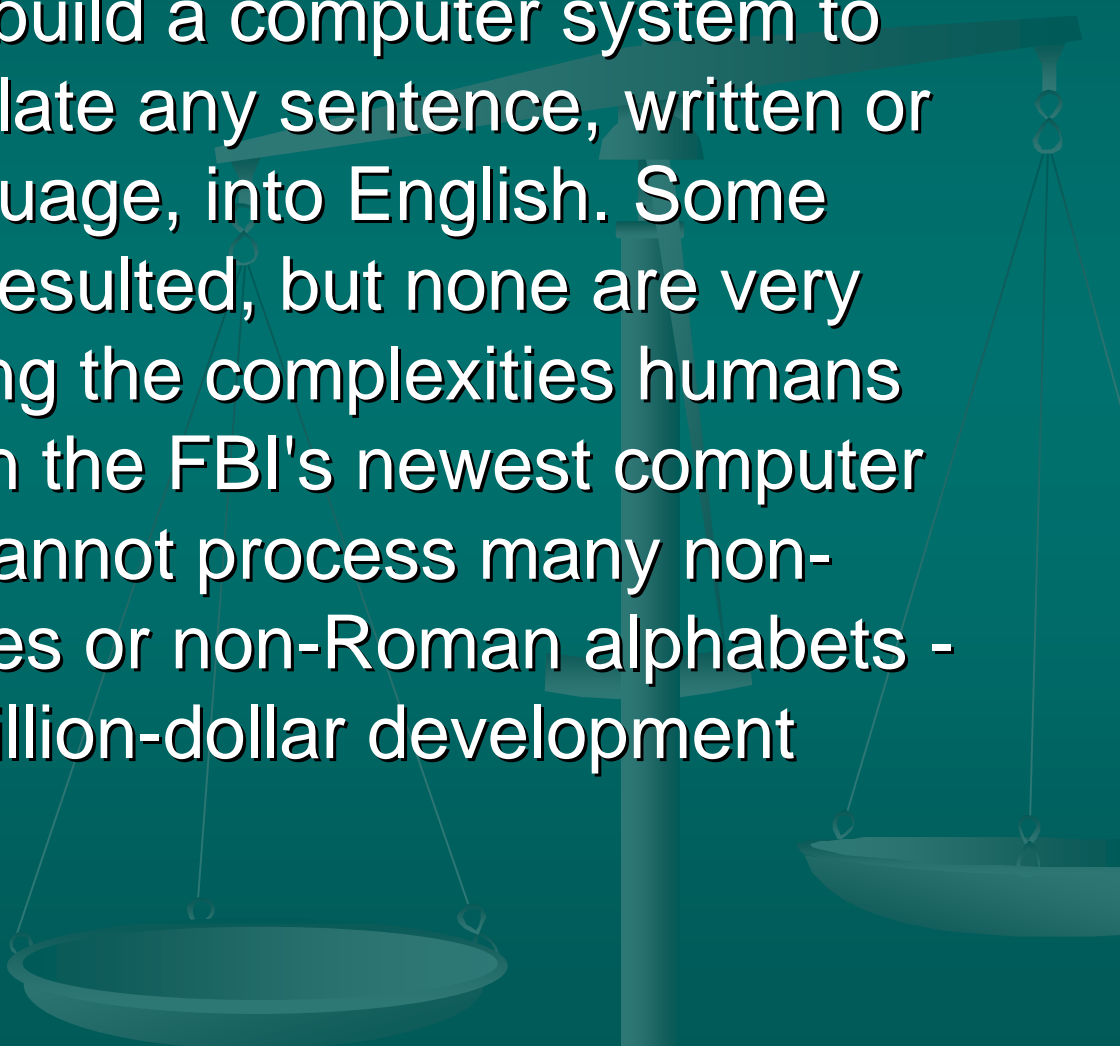


# Citation Maintenance



- An Update Model
  - Not Just Changes in MeSH
  - Repercussions for MEDLINE/PubMed
- No Retrospective Indexing
- Year End Processing
  - Automatic Tasks
  - Manual Tasks
- Now Published

The federal government has spent huge sums of money trying to build a computer system to automatically translate any sentence, written or spoken in any language, into English. Some limited tools have resulted, but none are very effective at decoding the complexities humans easily handle. Even the FBI's newest computer network (Trilogy) cannot process many non-European languages or non-Roman alphabets - - despite its 600-million-dollar development price.



The goal of universal machine translation is being replaced by a different vision -- one that involves using a variety of tools that assist rather than replace humans. The new objective is to link inexperienced translators to online dictionaries and expert assistance.

Translation in the Age of Terror, MIT Technology Review, March.

