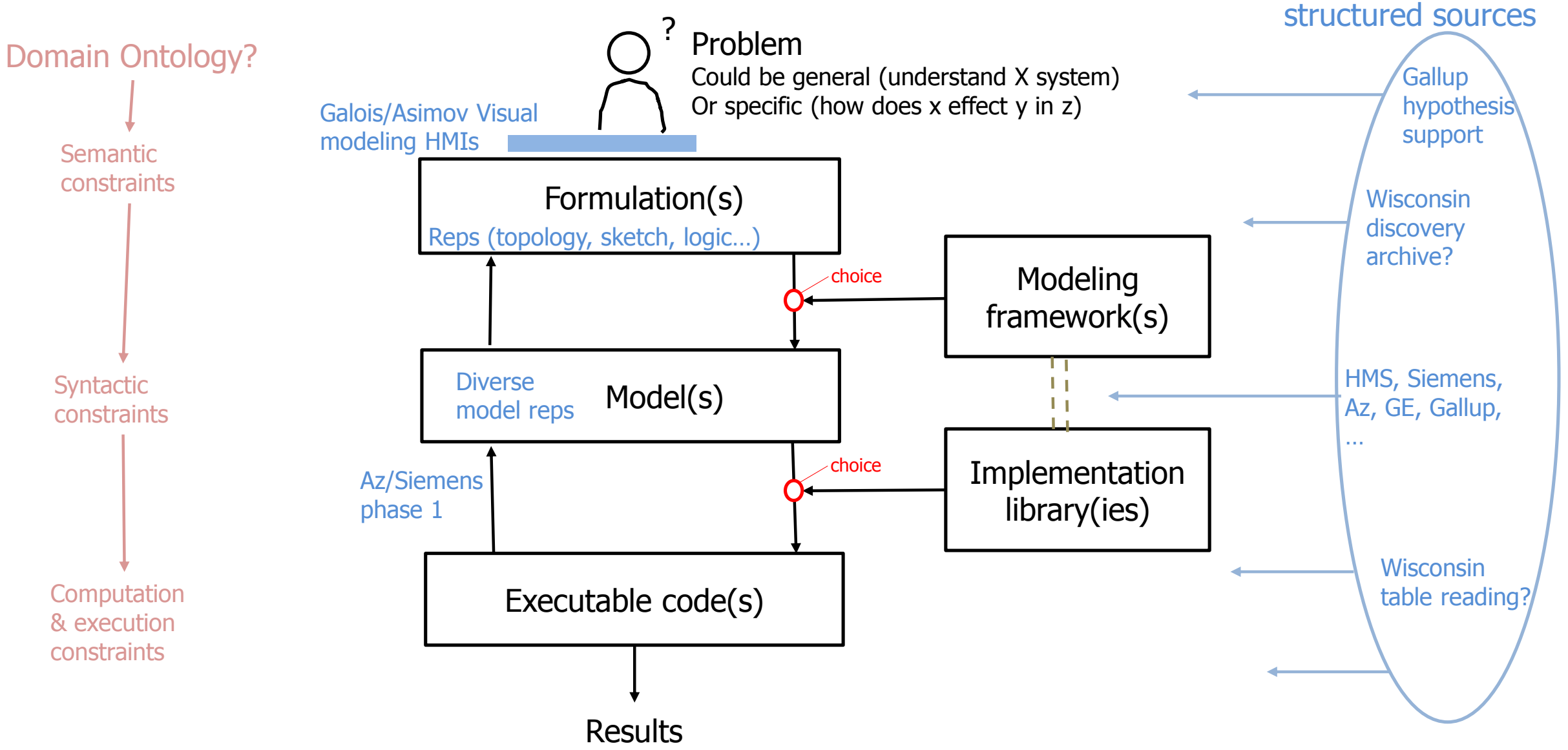
ASKE ANSWER Description
(Homework from Phase 2 Kick-off Meeting)

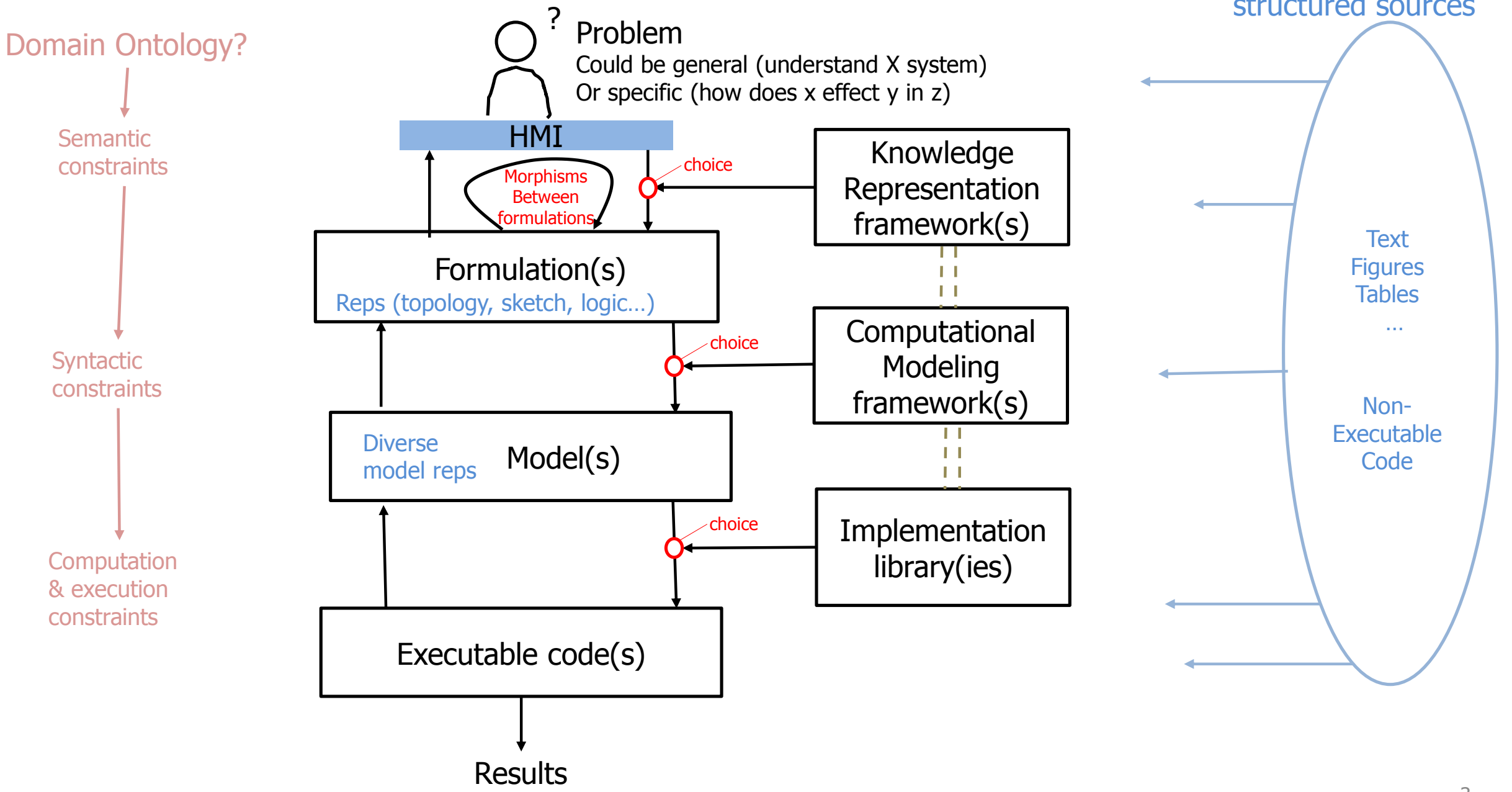
May 24, 2019

Andrew Crapo
Narendra Joshi
Nurali Virani
Varish Mulwad

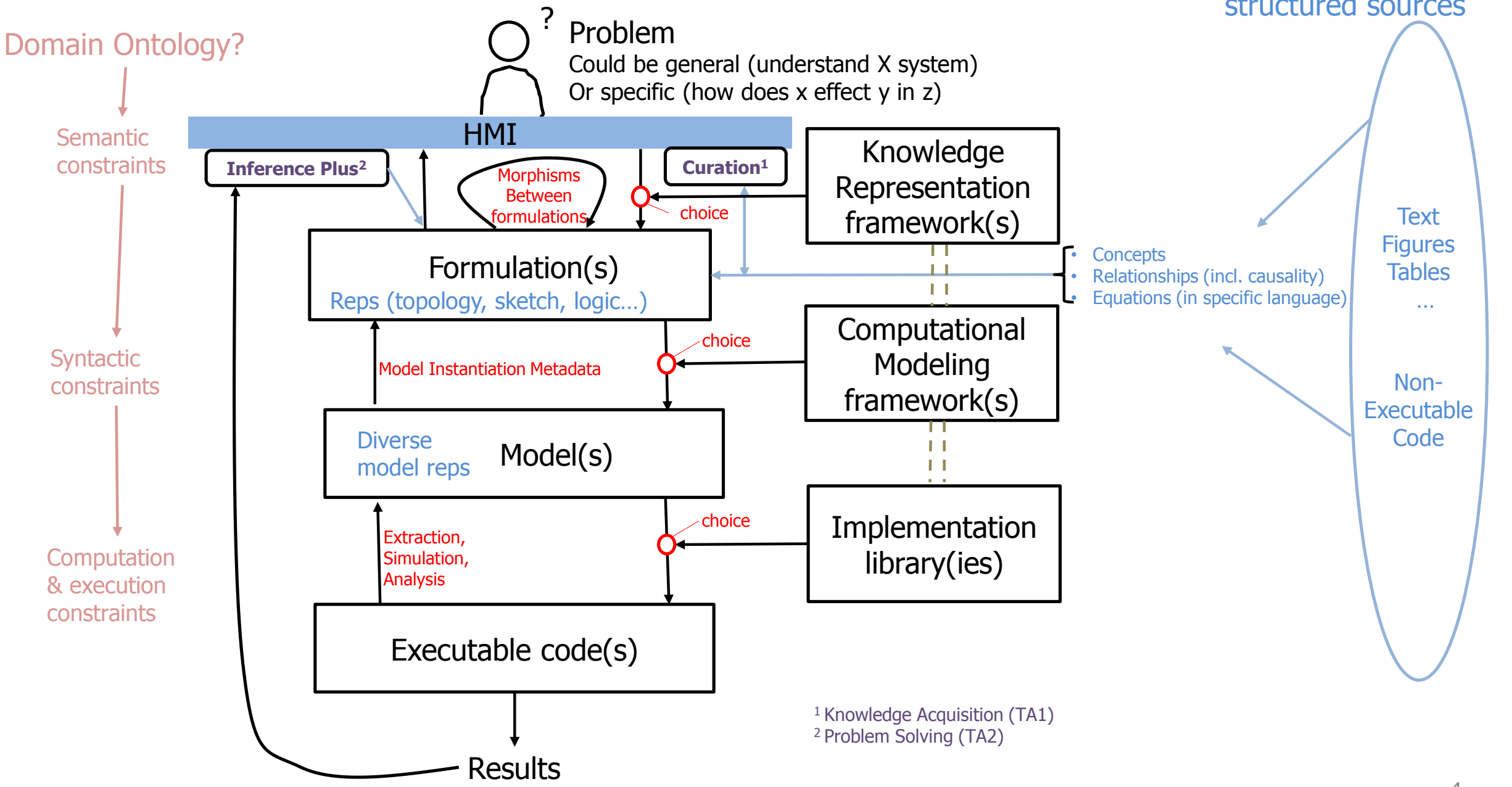
Revised sketch (original)



Refactored sketch (2 levels of framework, non-executable code, KR morphisms)



GE Work in Context (emphasis on HMI, human-in-the-loop extraction, curation, and use)



[1] Knowledge Acquisition (TA1)
[2] Problem Solving (TA2)

Description of GE ASKE TA1 ANSWER in Refactored Sketch Terms

Scientific Knowledge Extracted to Semantic Representations

- As part of **knowledge extraction from code**, the ANSWER system extracts characteristics of the code from the parser's AST, creates from this a semantic model of the code, and reasons over this model to decide which portions of the code to extract, convert to Python, and add to the knowledge graph.
- As part of **knowledge extraction from semi-structured sources**, the ANSWER system extracts scientific concepts and equations that appear in text. Additionally, it also extracts the mapping between the equation variables and their respective concepts. The system also links the extracted concepts with existing entities in Wikidata. The text extraction modules have been trained on concepts and equations that appear on the web pages of NASA's hypersonic index.
- The extracted concepts, equations, and code extracts are modeled using an OWL ontology (**W3C OWL is one of our knowledge representation frameworks**). While the concepts, along with the variables that appear in an equation (including the return variable) can be modelled in OWL, we currently model the equation expression using Python (which is represented as a text string in OWL).
- Extracted knowledge is curated in collaboration with the user (not yet implemented) to determine what to actually add to the knowledge repository and with what provenance (source, trust, etc.) The equations and relationships are also converted into a **K-CHAIN** model, a **computational modeling framework** that is part of the ANSWER system.

Scientific Knowledge Instantiated in Computational Models

- Knowledge-consistent hybrid AI network or K-CHAIN is a **computational modeling framework**, that uses a **computational graph (a.k.a. a dataflow graph)**, to represent computations in terms of the dependencies between individual operations and variables. In this modeling framework, we first define the computational graph using equations or relations from semantic knowledge, then run parts of the graph for inference processing. This modeling framework also allows us to represent random variables (with associated distributions) as nodes of the graph to enable **Probabilistic programs** and **Bayesian network** models, which will be explored in Phase 2.
- This computational modeling framework uses **the implementation library TensorFlow** to create executable code. TensorFlow supports multiple programming languages to create and run the computational graphs. We use the **Python** programming language to automatically create the computational graphs in TensorFlow based on exact knowledge in the form of equations and inexact knowledge in the form of dependencies from text. These implemented models can then be run in TensorFlow sessions to support problem solving as well as guided curation in knowledge acquisition.

Human-Machine Interaction

- We are developing the **Dialog** grammar and editor to enable mixed-initiative communication between user and system for both curation and problem solving, the latter shared with TA2. Dialog is augmented with various graphical representations to communicate information from the system to the user.