Capturing Subject Matter Expertise for Automated Assisted Analysis

William F. Szewczyk

National Security Agency

Mathematics Research Group

Level Setting

Data science is the extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing.

— *NIST Big Data Interoperability Framework: Volume 1, Definitions (Sept. 2015)*

Level Setting (II)

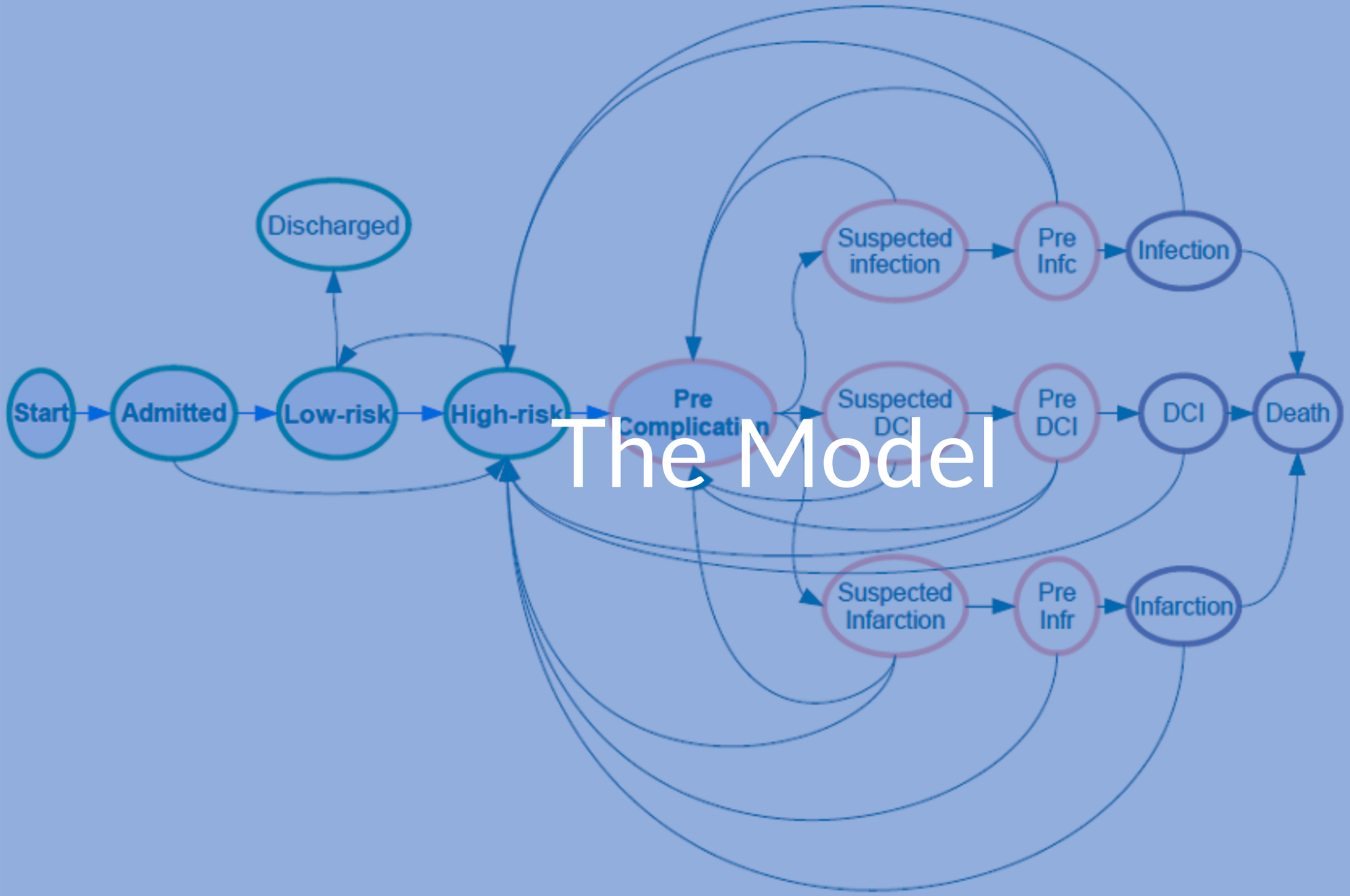
A data scientist is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the data life cycle.

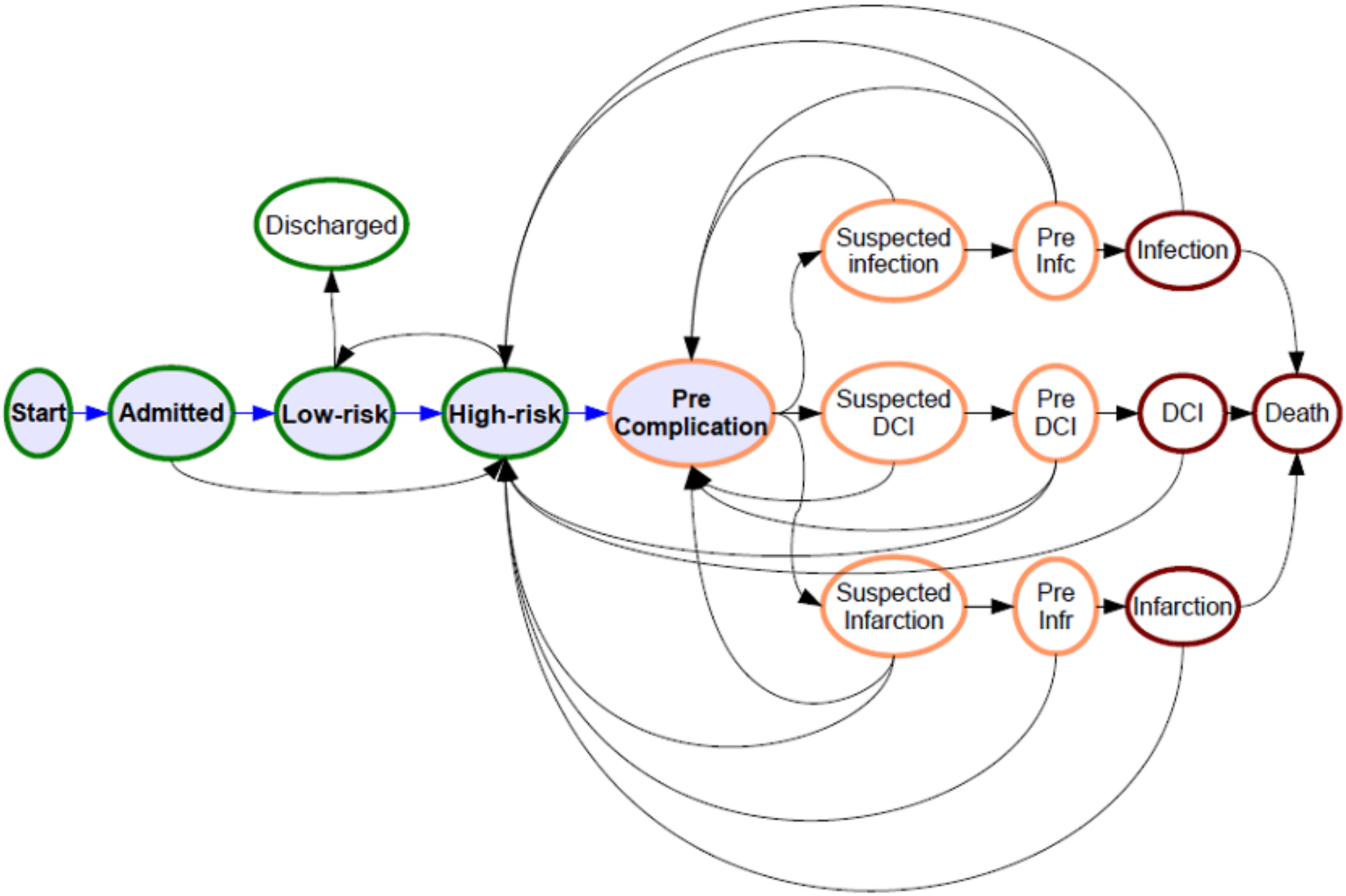
— *Ibid.*

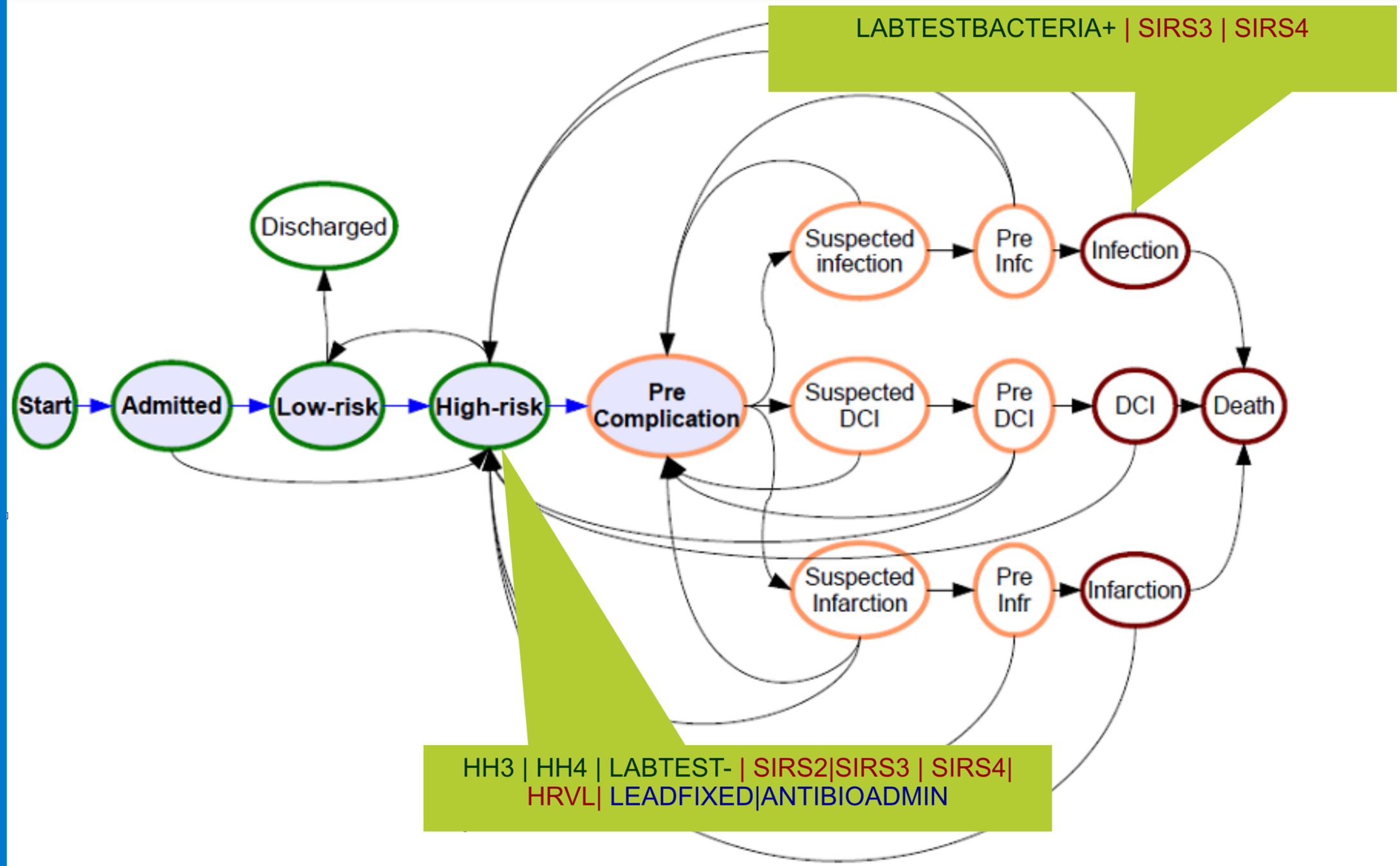
Let’s Play a Game!

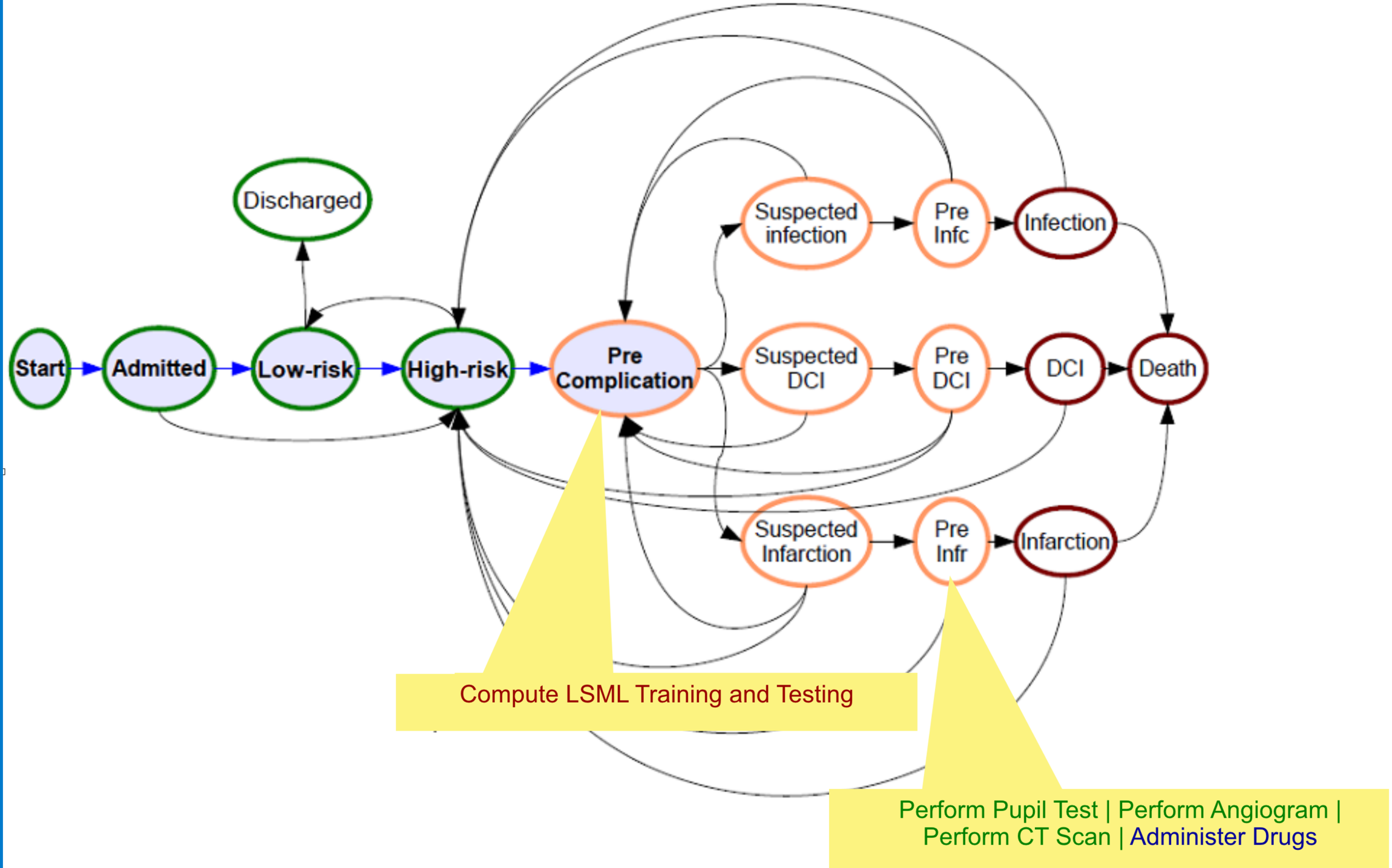
- I’ll give you observations
- You give me states of reality
- I’ll make it easy
  - I’ll let you ask questions
  - I’ll let you take actions
  - I’ll even let show the possible states

The Model


Discharged
Start
Admitted
Low-risk
High-risk
Pre
Complicati
Suspected
infection
Pre
Infc
Infection
Suspected
DCI
Pre
DCI
DCI
Death
Suspected
Infarction
Pre
Infr
Infarction


Discharged
Start
Admitted
Low-risk
High-risk
Pre Complication
Suspected infection
Pre Infc
Infection
Suspected DCI
Pre DCI
DCI
Death
Suspected Infarction
Pre Infr
Infarction


LABTESTBACTERIA+ | SIRS3 | SIRS4
Discharged
Start
Admitted
Low-risk
High-risk
Pre Complication
Suspected infection
Pre Infc
Infection
Suspected DCI
Pre DCI
DCI
Death
Suspected Infarction
Pre Infr
Infarction
HH3 | HH4 | LABTEST- | SIRS2|SIRS3 | SIRS4|
HRVL| LEADFIXED|ANTIBIOADMIN


Discharged
Start
Admitted
Low-risk
High-risk
Pre Complication
Suspected infection
Pre Infc
Infection
Suspected DCI
Pre DCI
DCI
Death
Suspected Infarction
Pre Infr
Infarction
Compute LSML Training and Testing
Perform Pupil Test | Perform Angiogram |
Perform CT Scan | Administer Drugs

Now multiply by about 10 or 20 and find places to automate

This works, but...

- It really doesn't scale
- It's hard to identify places for code reuse
- Finding design patterns is difficult

What is needed is a **repeatable** method for decomposing the analytic process into finer pieces, while avoiding the temptation to drop immediately into code.

What We Propose

- A process and artifacts for understanding, decomposing, and communicating the process of analysis for a specific problem.
- It provides a set of phases and artifacts designed to capture the analysis process, individual blocks of analysis, and decompose the individual components which assist in the analysis.
- Its output is intended to become input to a traditional systems analysis and design methodology for building systems.
- Specific focus is on systems that can codify the knowledge that analysts continuously adjust and apply to repetitive analysis tasks (e.g. anomaly detection, state change, prediction, etc).
  - Achieve agreement with analysts of analysis they perform and how it is being broken down and described (analysis dependencies)
  - Communicate complex analysis dependencies with System Designers
  - Identify and detail specific advanced analysis patterns for use by System Designers

Description

- Divided into three phases which promote a focus on the problem at hand, identifying where and what analysis is performed, and identifying the dependencies and details of the analysis
  - Phases:
    - **Problem (Mission) Definition and Refinement**
      Define an unambiguous statement and definition of the analysis problem including specific outcomes expected.
      Capture constraints and assumptions of the mission
    - **Conceptual Analysis Design**
      Capture a high level decomposition of the analysis that is performed to achieve the Mission problem
      Capture the Blocks of Analysis and their dependencies on one another.
    - **Logical Analysis Design**
      Logical decomposition of conceptual analysis design. Logical blocks focus on the purpose of the logic, not the implementation.
      Capture the flow of information and transformations that are observed to perform analysis and produce a resulting hypotheses and actions.

Phase One: Problem Definition

- Possibly the hardest part of the methodology
  - Specific and precise - not general, broad or obscure
  - Concise Definition - one or two sentences
  - Refinement provides supporting information and questions below
  - Clear - not vague, ambiguous or confusing
  - Goal-oriented - stated in terms of desired outcomes
  - Key Questions
    1 Who is the recipient of the analysis results
    2 What is the question you are trying to answer | What is the situation you are trying to be aware of
    3 What do the answers / results of this analysis look like?
    4 Why are you doing this? What is the benefit

Problem Definition

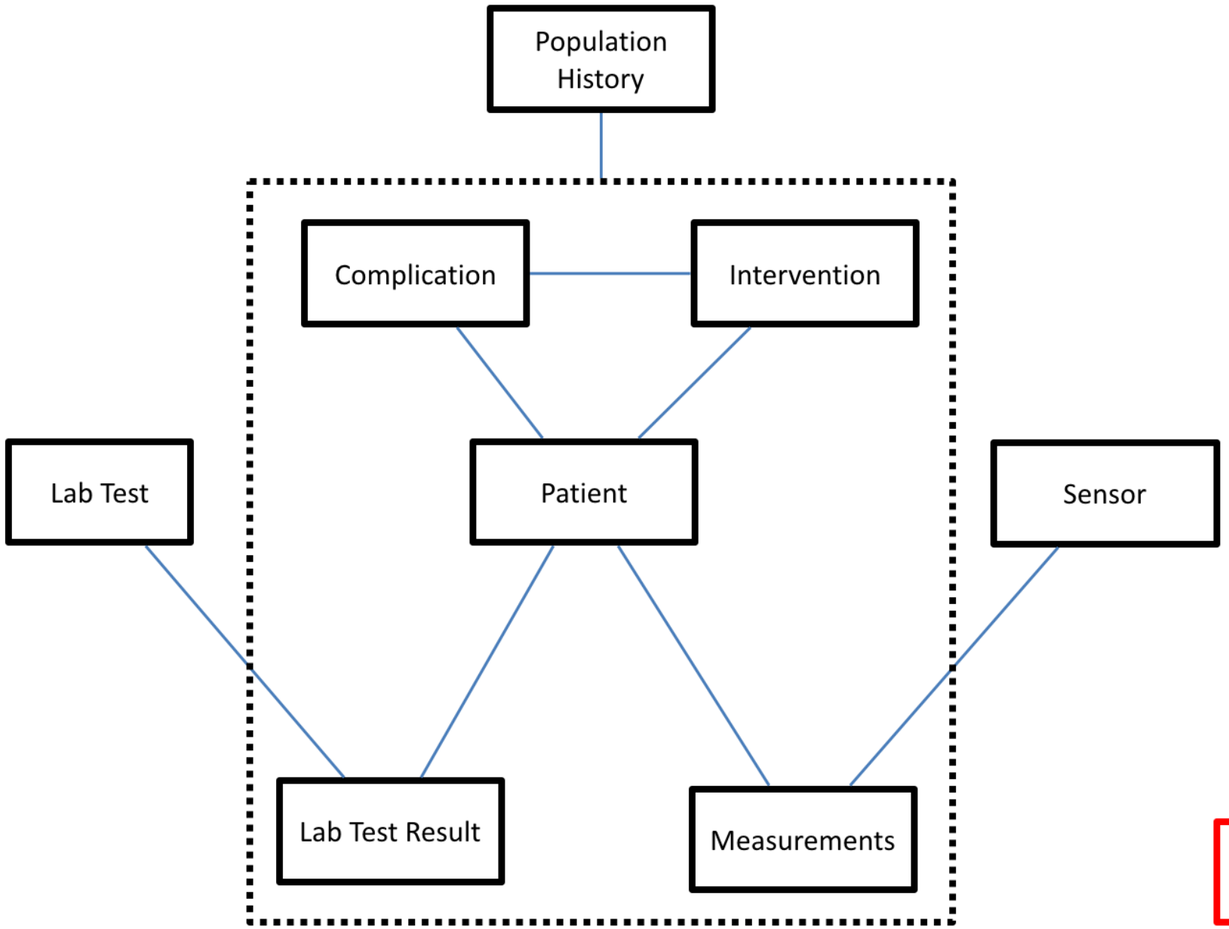
- Predict and Detect complications on patients in the Neuro-ICU as early as possible
- To support interventions
- Recipient: Health care providers (doctors/nurses)
- Report: likely complication with time frame

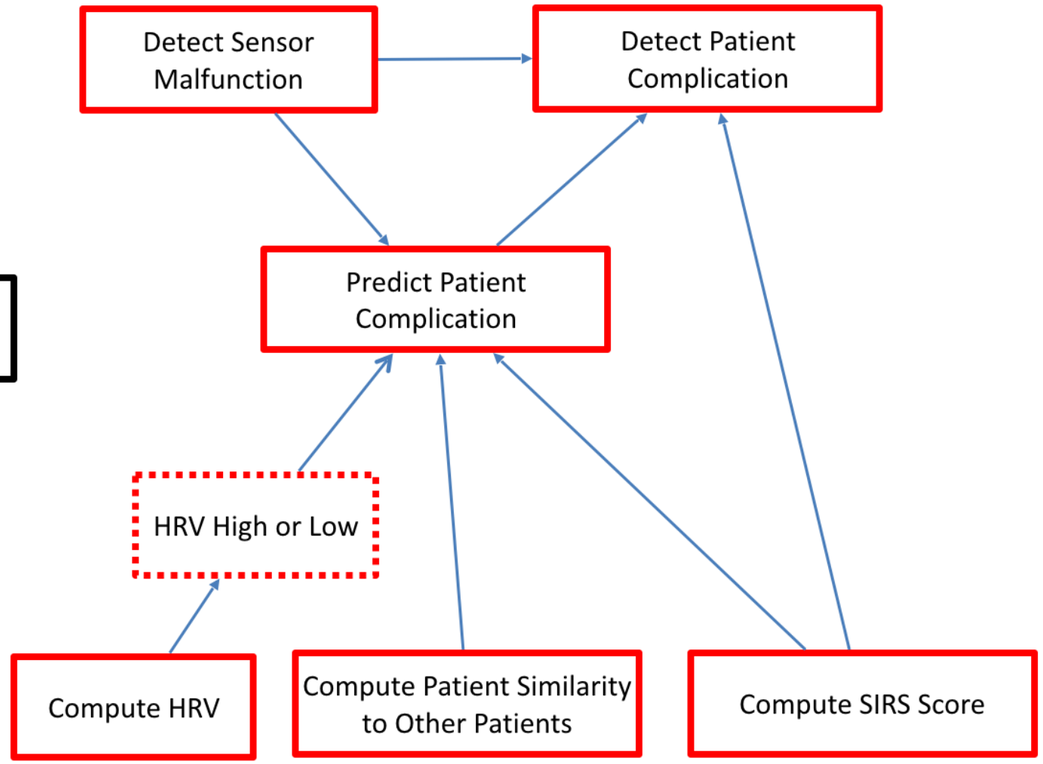
Phase 2: Conceptual Analysis Design

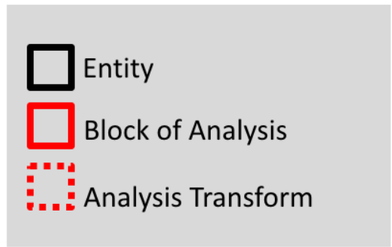
- Capture a high level decomposition of the analysis that is performed to achieve the Mission problem
  - Capture the subset of conceptual data design of the mission problem for the purpose of clarifying terms and grammar of the mission (Entity Types, Entity Relationships and their descriptions)
  - Capture the Blocks of Analysis , Analysis Result Dependencies (Conceptual Analysis Model) and descriptions of each (Conceptual Analysis Description Document)
  - Limit scope to that of the Problem Definition

Conceptual Analysis Design

- Candidate Entities:
  - Patient & demographics, sensor, infection, infarction, DCI, ECG, temperature, equipment alarms
- Candidate Blocks of Analysis:
  - Calculate SIRS score
  - Calculate Heart Rate Variability


Population History
Complication
Intervention
Lab Test
Patient
Sensor
Lab Test Result
Measurements

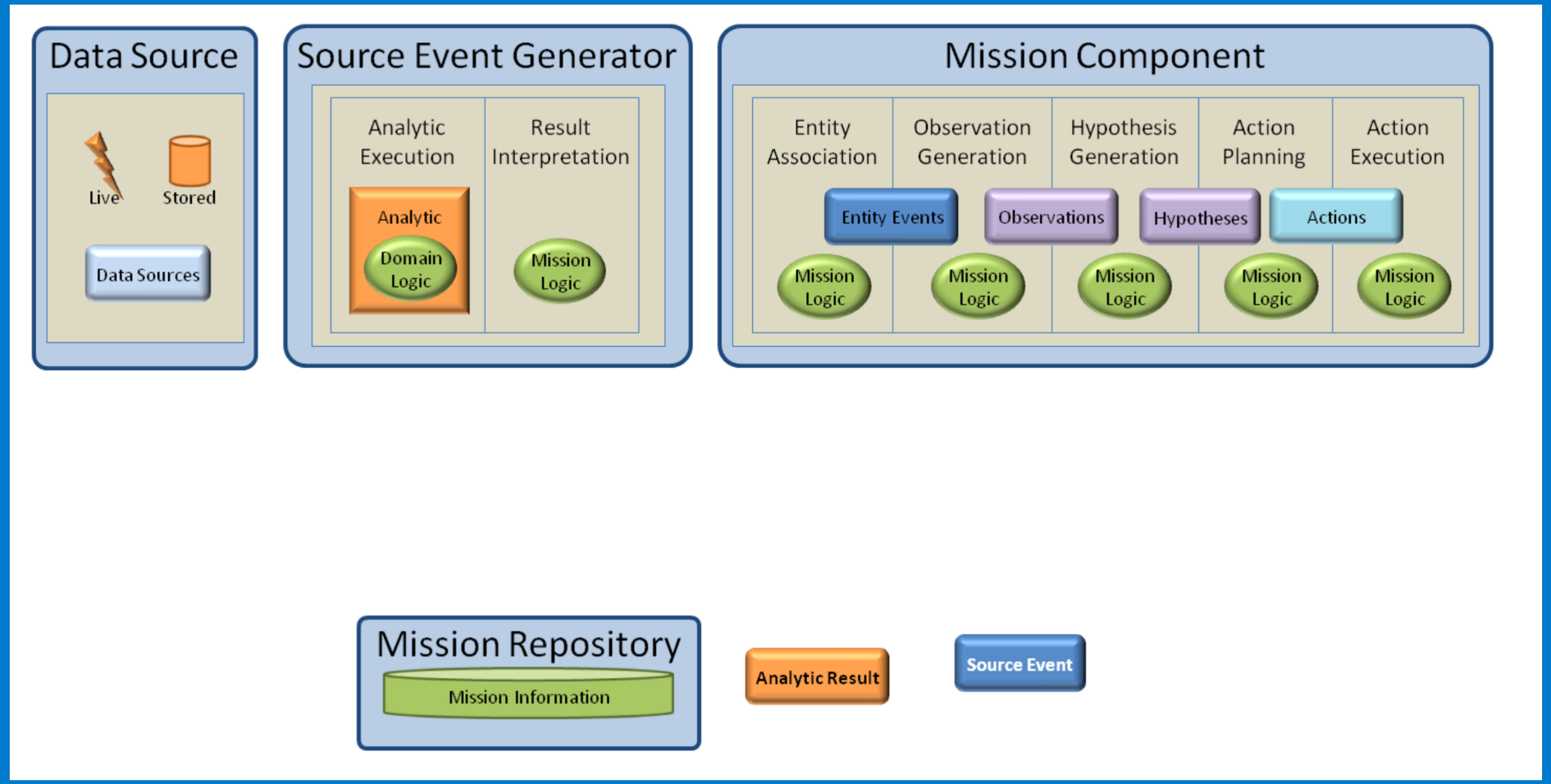

Detect Sensor Malfunction
Detect Patient Complication
Predict Patient Complication
HRV High or Low
Compute HRV
Compute Patient Similarity to Other Patients
Compute SIRS Score

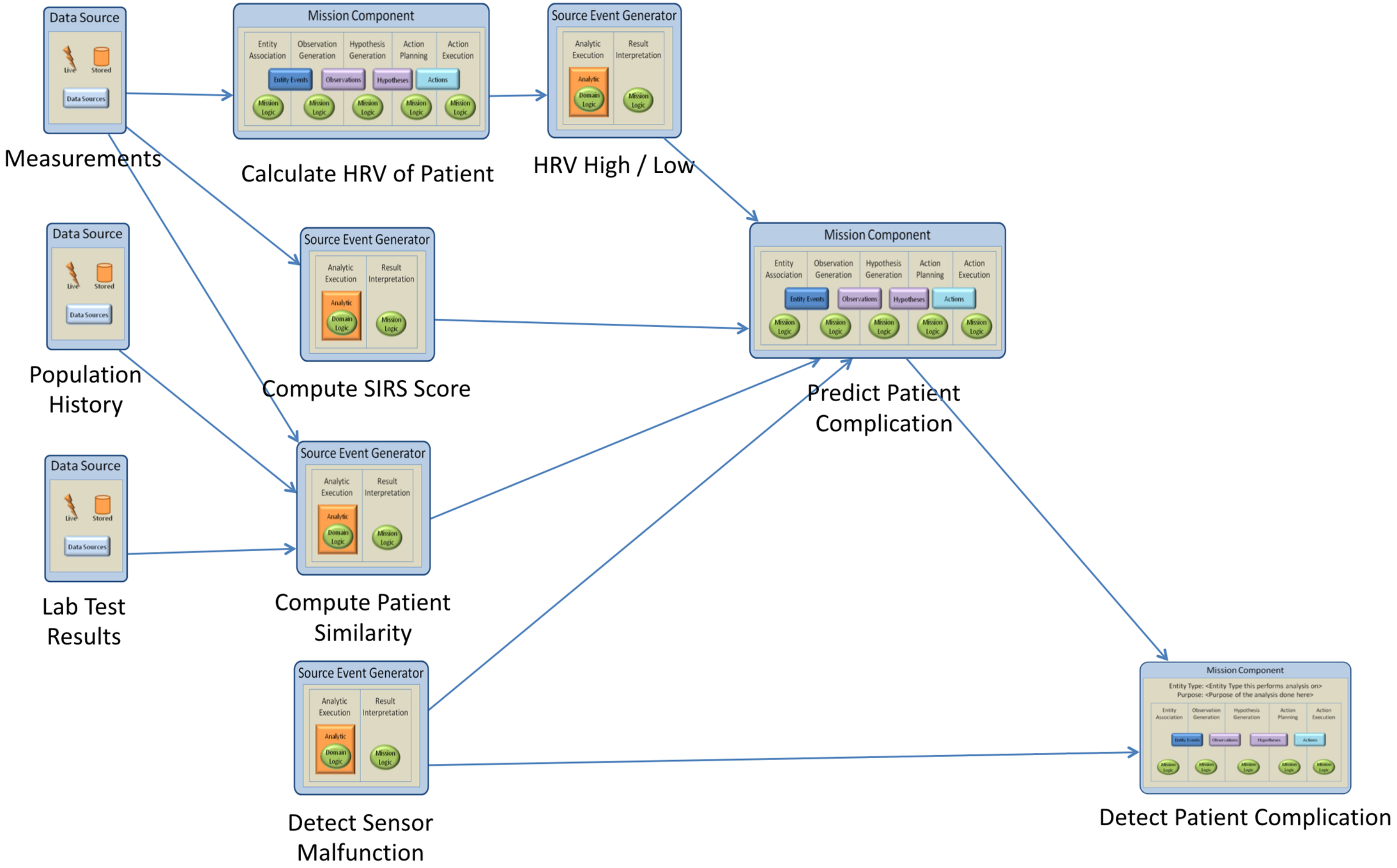

Entity
Block of Analysis
Analysis Transform

Phase 3: Logical Analysis Design

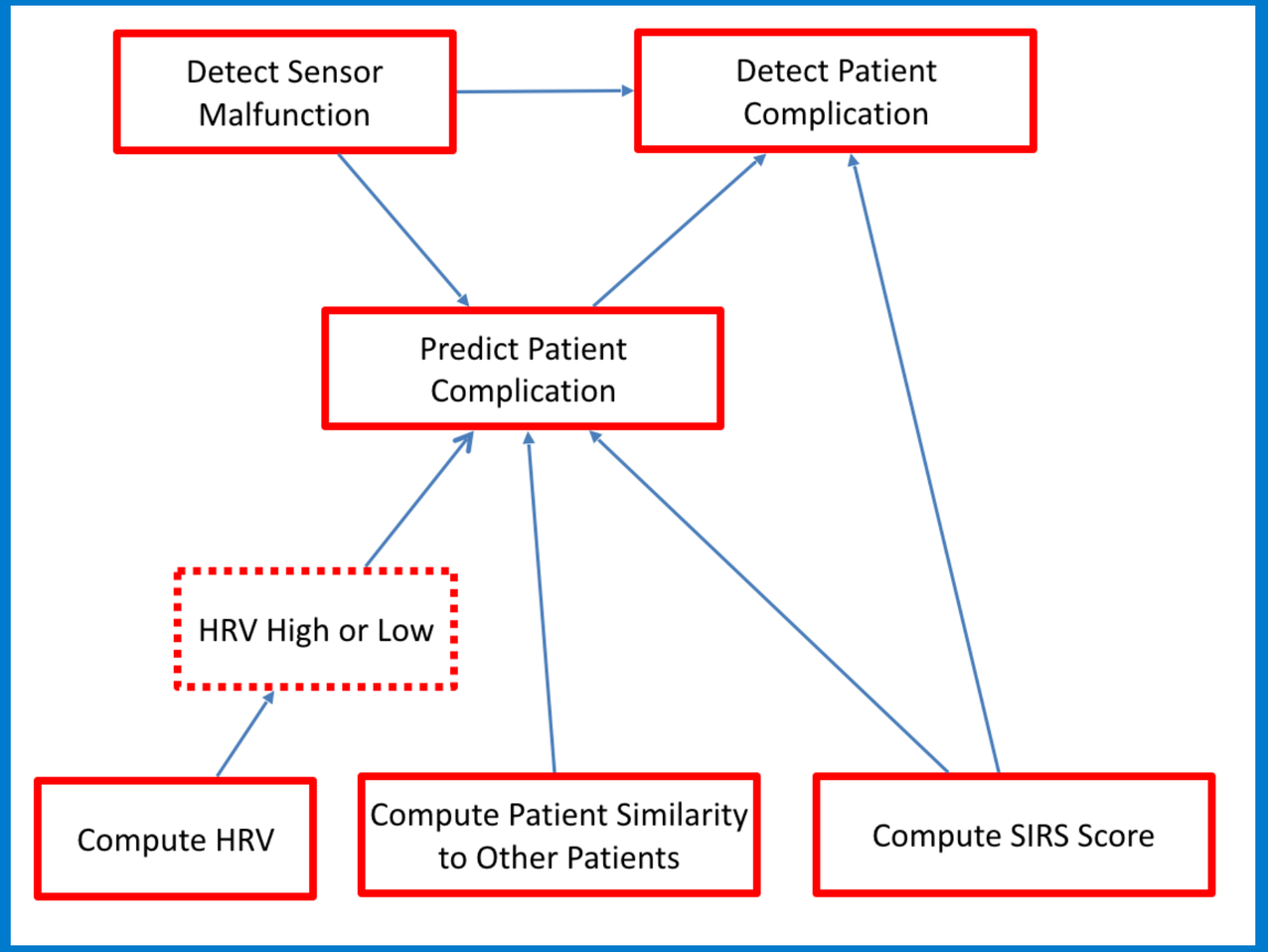
- Develop decompositions of the Blocks of Analysis to capture the flow of information and transformations that are observed to perform analysis and produce a resulting hypotheses and actions.
- Identify Data Sources, Source Object Generators, Mission Components, and their inter-dependencies
- Identify the Object of Analysis for each Mission Component
- Capture the Source Objects, Entity-Associated Objects, Observations, Hypotheses, and Actions for each Mission Component.

Logical Analysis Design Components




Data Source
Live
Stored
Data Sources
Measurements
Mission Component
Entity Association
Observation Generation
Hypothesis Generation
Action Planning
Action Execution
Entity Events
Observations
Hypotheses
Actions
Mission Logic
Calculate HRV of Patient
Source Event Generator
Analytic Execution
Result Interpretation
Analytic
Domain Logic
HRV High / Low
Data Source
Population History
Source Event Generator
Compute SIRS Score
Mission Component
Predict Patient Complication
Data Source
Lab Test Results
Source Event Generator
Compute Patient Similarity
Source Event Generator
Detect Sensor Malfunction
Mission Component
Entity Type: <Entity Type this performs analysis on>
Purpose: <Purpose of the analysis done here>
Detect Patient Complication

Pattern for Discovery



Takeaways

- Described a process to capture subject matter knowledge
- Demonstrated ability to switch domains with only minor modifications
- Easier to discover common patterns across applications