MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

University of Cyprus

MAI643 Artificial Intelligence in Medicine

Elpida Keravnou-Papailiou

January – May 2023

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU | Master programmes in Artificial Intelligence 4 Careers in Europe

AIM: from Knowledge-Intensive to Data-Intensive Applications

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

UNIT 2

AIM: from knowledge-intensive to data-intensive applications

CONTENTS

1. The AI challenges of the medical field
2. Influential knowledge-based systems in medicine
3. Probabilistic models
4. Knowledge acquisition: the use of ontologies
5. Intelligent data analysis in medicine

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

INTENDED LEARNING OUTCOMES

Upon completion of this unit on AIM: from knowledge-intensive to data-intensive applications, students will be able:

1. To appreciate the AI challenges of the medical field.
2. To list the major characteristics of a knowledge representation.
3. To point out key features of several influential knowledge-based systems in medicine, and to present their inferencing through hypotheses status transition diagrams.
4. To grasp the importance of causality in medical knowledge-based systems.
5. To overview the probabilistic models of Naïve Bayes, Bayesian networks, decision analysis and influence diagrams.
6. To explain the use of ontology systems in knowledge acquisition using the Protégé tool and the Unified Medical Language System (UMLS) as examples.
7. To illustrate the intelligent data analysis of medical data by presenting data abstraction methods and data mining through symbolic classification methods.

The AI challenges of the medical field

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Is medicine science or art?

- This philosophical question is frequently posed to show that expert clinicians often reach correct decisions based on intuition and hindsight rather than scientific facts
- Medical tasks, such as diagnosis and therapy, are by nature complex and not easily amenable to formal approaches
- Medical knowledge is inherently uncertain and incomplete
- Likewise, patient data are often ridden with uncertainty and imprecision, showing serious gaps; moreover, they could be too voluminous and at a level of detail that would prevent direct reasoning by a human mind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Over the years medicine has accumulated different kinds of knowledge

- The digitalis (from the purple foxglove) story:

digitalis, drug obtained from the dried leaves of the common foxglove (Digitalis purpurea) and used in medicine to strengthen contractions of the heart muscle. Belonging to a group of drugs called cardiac glycosides, digitalis is most used to restore adequate circulation in patients with congestive heart failure, particularly as caused by atherosclerosis or hypertension.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Development of medical knowledge from ancient times

- Traditional healers experimented with natural medications to treat symptoms
- Ancient Egyptian records pass on the knowledge that digitalis helps to treat certain types of congestive sickness – in that era there was no understanding that the heart pumped blood
- The empirical correlation between treatment with foxglove and improvements in some patients became part of medical knowledge
- The fact that digitalis affects of the heart was recognized much later (1785) by Withering who further characterized its therapeutic and toxic effects, and published a guide to its proper use, that remained the state of the art until real pharmacokinetic models supplanted it in the middle of the twentieth century

The evolution of understanding enables increasingly more sophisticated uses of such knowledge to improve medical care

Empirical correlation that holds up frequently enough to be clinically useful

↓

Associations may become interpreted as due to some mechanism whose operation is understood at some level of detail

↓

Possibly a more quantitative understanding of just how a disease develops, what to expect from its unchecked development (prognosis), how it generates the signs and symptoms associated with it (diagnosis) and how it responds to therapeutic interventions (treatment)

Still the mechanisms of many diseases are not understood in detail; yet useful knowledge of the above various sorts has been accumulated to help improve the lives of patients

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The computer-based performance of medical tasks poses many challenges

- It is not surprising that AI researchers were intrigued with the automation of medical problem solving from the early days of AI
- The technology of expert systems is largely founded on attempts to automate medical expert diagnostic reasoning
- Renowned are the Stanford experiments of the Heuristic Programming Project resulting in the MYCIN family of rule-based systems

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Care providers perform various tasks

- ❑Diagnose the cause of a problem
- ❑Predict its development
- ❑Prescribe treatment
- ❑Monitor the progress of a patient, and
- ❑Overall manage a patient

Their decisions should be as informed as possible; in the current age of information explosion, the only viable means of handling large amounts of information are computer based.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Change of focus

- The work of all care providers can benefit substantially from computer-based support
- In the early days, the biggest challenge was the modeling of knowledge for supporting tasks such as diagnosis, therapy and monitoring; to a certain extent this is still a challenge
- The information explosion has brought a drastic change of focus from knowledge-intensive to data-intensive applications and from systems that advise to systems that inform
- The biggest challenge now is the intelligent exploitation of medical data, whether they refer to clinical or demographic data

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Intelligent exploitation of data

- Can yield significant new knowledge, e.g., guidelines and protocols for the treatment of acute and chronic disorders, by summarizing all available evidence in the particular field
  - Evidence-based medicine
- Can provide accurate predictors for critical risk groups based on “low-cost” information
- Aims to provide means for the intelligent comprehension of individual patient’s data, whether such data are riddled with gaps, or are voluminous and heterogeneous in nature
  - Thus, closing the gap between the raw patient data and the medical knowledge to be applied for reaching the appropriate decision for the patient

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The shift in focus has not changed the ultimate objective

Which is to aid care providers reach the best possible decisions for any patient, to help them see through the consequences of their decisions/actions and if necessary to take rectifying actions as quickly as possible.

The change in focus though has given a new dimension of significance to clinical databases and in particular to the intelligent management and comprehension of such data.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Key medical tasks

- Diagnosis
- Therapy administration and monitoring
- Protocol- and guideline-based therapy
- Patient management

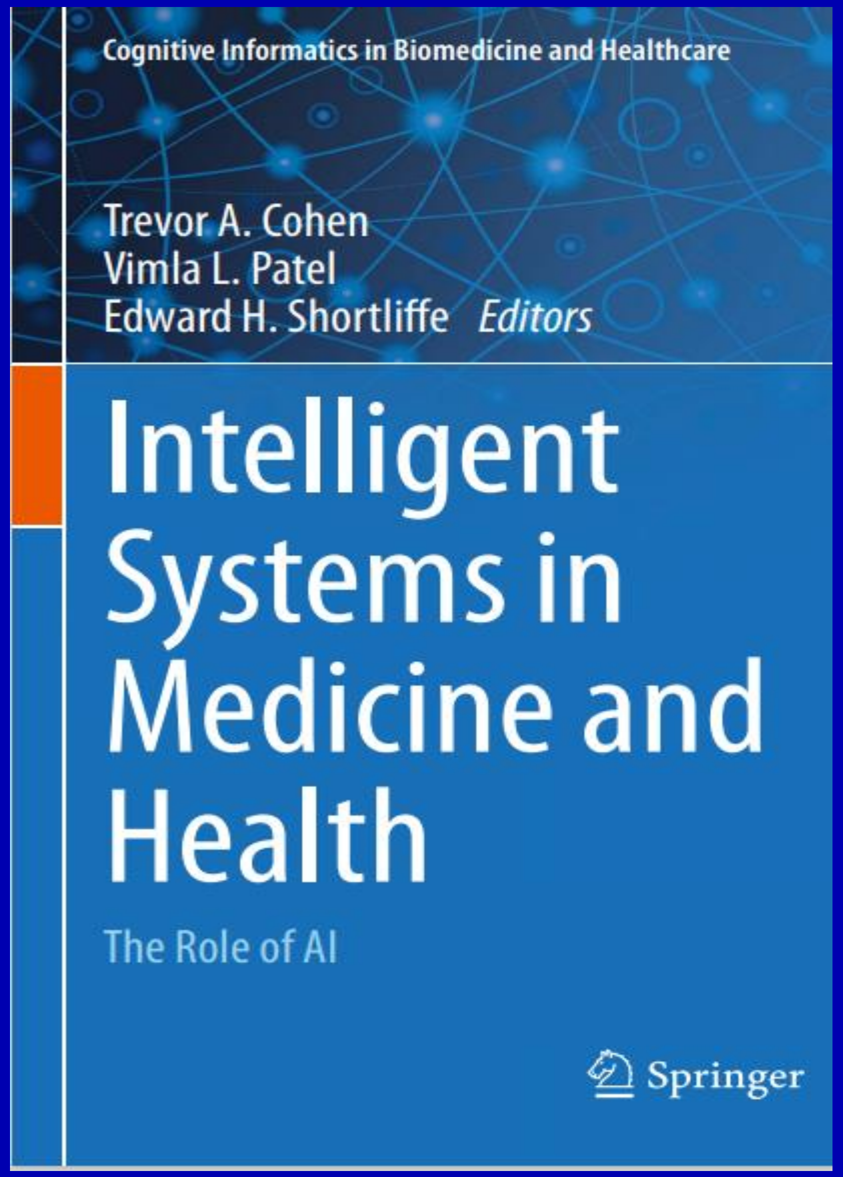
Clinical areas

- Cardiology
- Oncology
- Psychiatry
- Internal medicine
- Intensive care
- Cardiac surgery
- Orthopedics
- Urology
- Infectious diseases
- Anesthesiology
- Pediatrics
- Endocrinology

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe



Influential Knowledge-Based Systems in Medicine

(some material drawn from P. Szolovits and E. Alsentzer’s chapter in T.A. Cohen, V.L. Patel and E.H. Shortliffe (editors), Intelligent Systems in Medicine and Health: The Role of AI, Springer, 2022.)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Knowledge-Based Systems (KBS) in health care

- Try to reproduce in computer programs the ways in which human practitioners think about and handle difficult medical cases.
- The aim is not to replace human clinicians, but to improve their decision making by providing an automated "second opinion" in
  - Interpreting the available patient data
  - Choosing appropriate further tests and treatments
- Predominant examples of "augmented intelligence"
- Many of the landmark KBS were developed in the latter half of the twentieth century
- Researchers recognized early the potential value of learning from clinical databases, but at that time electronic medical (or health) records (EHRs) only existed in very few leading academic-affiliated medical centers.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

EHRs became widely adopted in 2009

- The Obama administration provided incentives under the Health Information Technology for Economics and Clinical Health (HITECH) Act
- Hospitals, clinics and practices were subsidized to install such systems
- Today, most major academic medical centers have repositories of case records documenting the conditions of, and the care given to millions of their patients, often going back in time for over a decade pre-2009
- Sharing of data is still a challenge, as institutions want to exploit their own data, before sharing them; hence regional, national and international repositories are far less common
  - A few exceptions: MIMIC, Physionet, eICU
  - Also, some large national repositories: UK Biobank, US “All of US”
- Machine Learning methods can be used to learn new knowledge from such collections of data, and integrate it with what is already known

Knowledge Representation

- Extensively discussed in the MAI611 course
  - Predicate Logic, Semantic Nets, Frames, Rules

- Major characteristics of a knowledge representation
  1. No representation can encompass all characteristics and associations of some entity, e.g., "cancer"
  2. A representation makes a set of ontological commitments: what real-world things can be represented in a computer? Logical adequacy
  3. A representation is tied to a fragmentary theory of intelligent reasoning
  4. A representation must be sufficiently efficient computationally to be practically useful; heuristic adequacy
  5. A representation must serve as a medium of human expression, allowing people and the computer to communicate their knowledge to each other; acquisitional adequacy

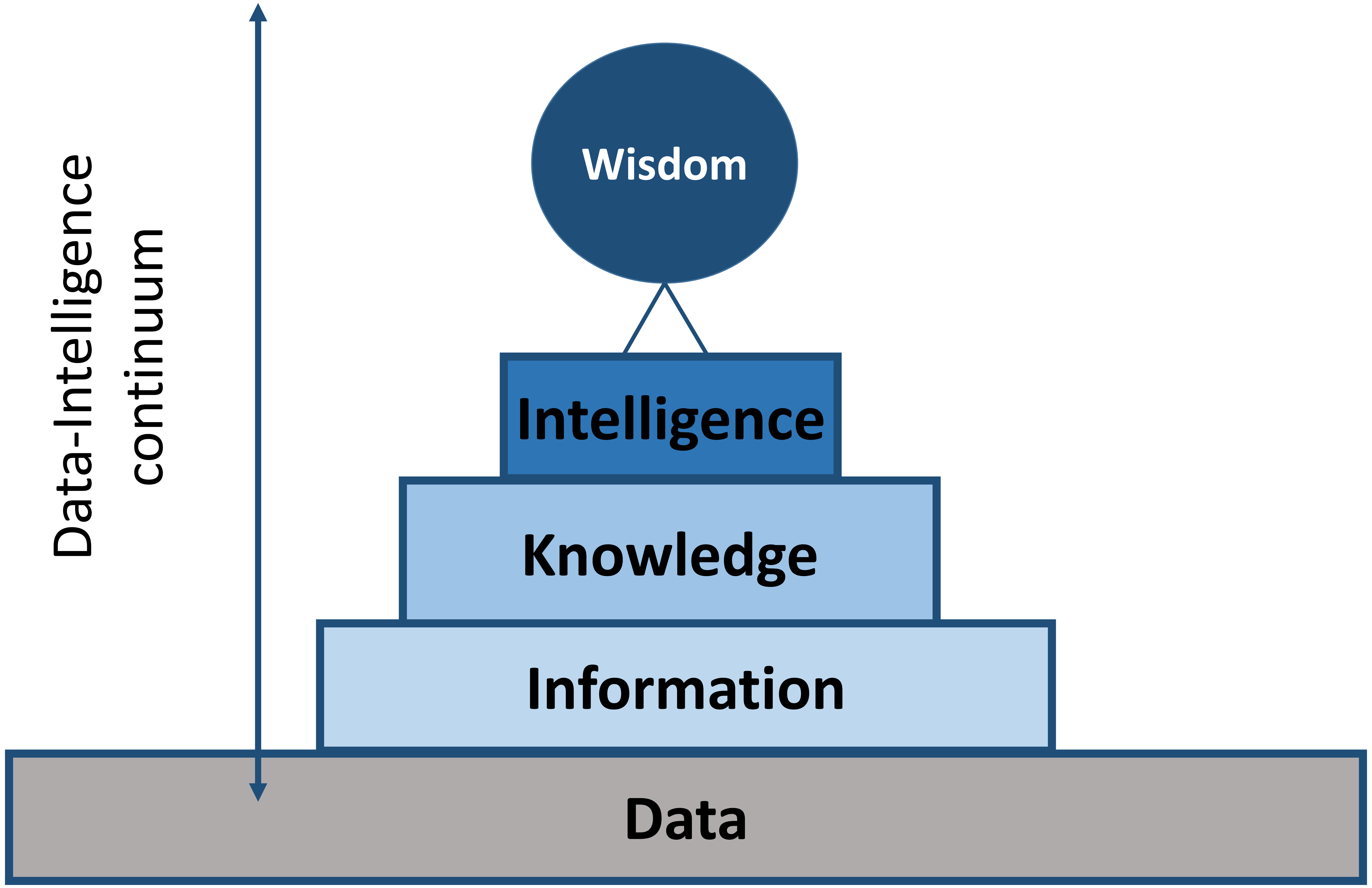
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Knowledge-Based System

- Is a system which manipulates “knowledge” in order to perform a task or tasks.
- The knowledge in a knowledge-base is in a highly structured symbolic form which represents a model of the relationship between knowledge elements and the uses to be made of them.
- The performance of a knowledge-based system depends both on the quality of its knowledge (structure, completeness, validity, consistency, etc.) and the ways in which this knowledge is applied.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



From A.C. Chang, Intelligence-Based Medicine: AI and Human Cognition in Clinical Medicine and Healthcare, Academic Press, 2020.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Influential knowledge-based systems in medicine

- MYCIN
- NEOMYCIN
- INTERNSIT-1, CADUCEUS, QMR
- PIP
- CASNET
- ABEL

These are primarily diagnostic systems. Diagnostic problem solving is a difficult task to model, especially when multiple failures/disorders are involved.

MYCIN, NEOMYCIN and INTERNIST-1 already discussed in the MAI611 course.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

A note on time representation and reasoning

- The modeling of time is one of the challenges regarding the computer-based automation of diagnostic reasoning.
- Some of the shortcomings of the pioneering medical diagnostic systems were attributed to their inability to model and reason with time; time was ignored, or it featured in a very implicit way in their knowledge bases and the patient data processed by them.
- For example, Minsky's original motivation for frame representations was to address a difficult technical problem in reasoning about actions, which require some representation of state or time:
  - Frame systems permitted facts to persist across states unless the action that moved from one state to another, explicitly altered an attribute
  - In medical applications, the use of frames was mainly to represent prototypical situations, e.g., a frame for a disease would have attributes that represented its typical signs, symptoms, predisposing factors, laboratory findings, drug or surgical treatments, etc., often in the aggregate called findings or manifestations
- Temporal clinical diagnosis will concern us at a later unit of the MAI643 course.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MYCIN

- **Function**: Diagnosis of, and recommendation of treatment for, antimicrobial infections
- **Knowledge Representation**: Rules
- **Notable features**: The rule-based framework and its explanatory facilities
- **Inference**: MYCIN diagnoses patients by reasoning entirely in a backward, deductive, fashion

- **Performance evaluation:**
  - 90% of MYCIN's therapy recommendations were found acceptable by panels of Stanford infectious disease experts
  - 97% agreement with Stanford and national experts regarding the program's identification whether the patient had a significant infection; 77% agreement regarding the identity of infecting organisms, and 73% agreement regarding the appropriate therapy

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Further on MYCIN’s performance evaluation .. Expert disagreements

- Because experts disagreed among themselves, MYCIN’s recommendations were considered reasonable even in some cases where they did not match the experts’ majority recommendation
- It is interesting to note that Stanford and national panels at times disagreed about proper treatment for a case and thus about whether MYCIN’s conclusions were appropriate, the program agreeing more often with the local experts, probably reflecting practice differences at different institutions

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

NEOMYCIN

- **Function**: To explicitly represent strategic knowledge and thus provide an efficient basis for teaching diagnostic reasoning and interpreting student behavior
- **Knowledge Representation**: Rules (object and meta) and frame-like structures
- **Notable features**: The representation of its reasoning knowledge in an abstract fashion in terms of tasks and meta-rules; the generation of the strategy tree and the strategic explanations
- **Inference**: NEOMYCIN abduces etiologies and important immediate state categories from the user observations; hypotheses are tested through their expectations

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

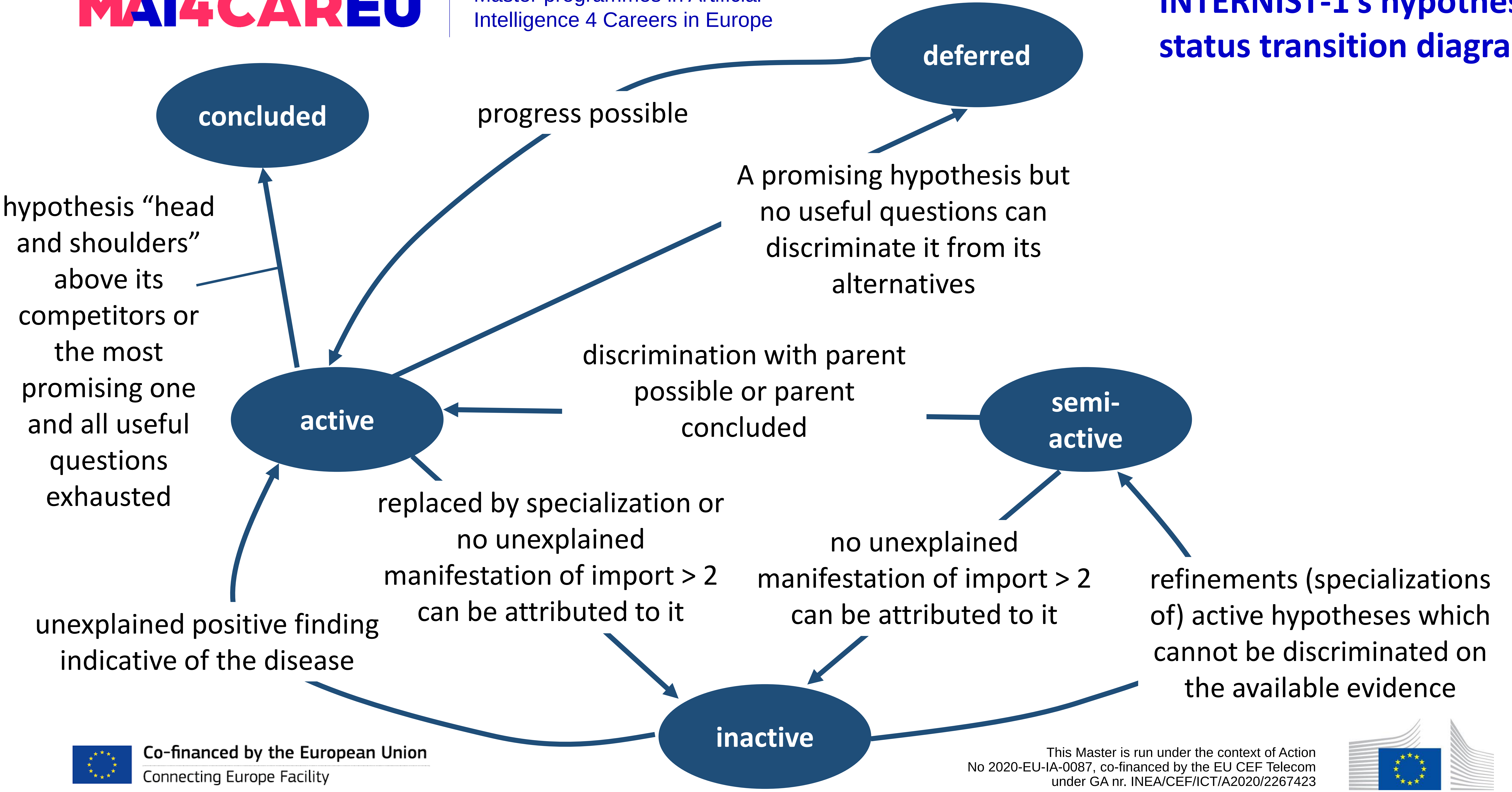
INTERNIST-1

- **Function**: Diagnosis of internal medicine
- **Knowledge Representation**: Frame-like structures
- **Notable features**: The formation of differential diagnosis; the synthesis of differential diagnoses via the partitioning heuristic; the information acquisition strategies
- **Inference**: INTERNIST-1 abduces hypotheses using the differential diagnosis lists of the “unexplained manifestations”; competing hypotheses are investigated by testing the deductions drawable from them

- INTERNIST-1’s main innovation is its clever partitioning heuristic for forming possible multiple differentials in the diagnosis of a complex case: **If a lower-scoring disease could explain either the same or a subset of the observed manifestations of the top-scoring disease, then it would be considered a competitor and be part of the differential**

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

INTERNIST-1's hypothesis status transition diagram



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

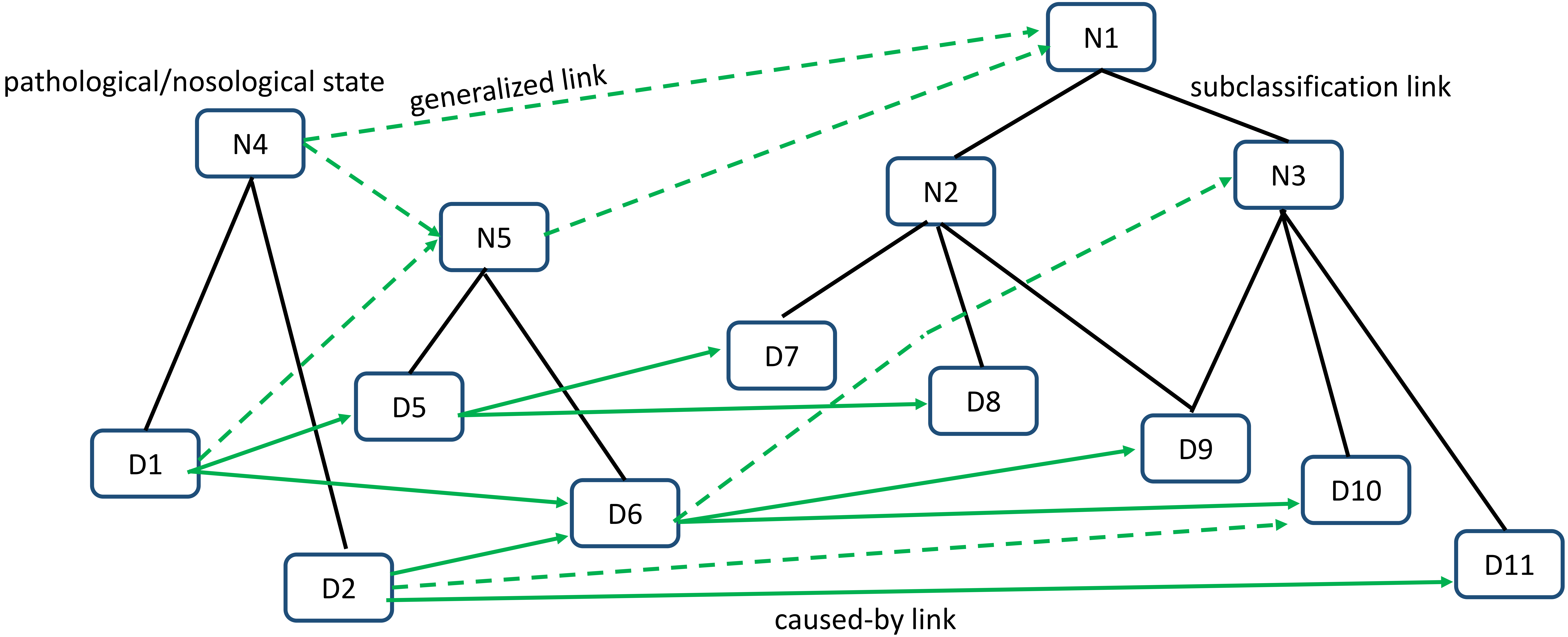
CADUCEUS

- **Function**: Diagnosis of internal medicine
- **Knowledge Representation**: Causal-taxonomical network
- **Notable features**: The restructuring of INTERNIST-1’s knowledge to permit two necessary and synergistic dimensions to the diagnostic reasoning; the formation of the initial problem contexts via special links (constrictors) from data to a point in the diagnostic space
- **Inference**: CADUCEUS abduces hypotheses from constrictor observations and established pathological states; the system differentiates competing hypotheses by testing the deductions drawable from them. More specifically, it extended INTERNIST-1’s differential diagnosis strategy to become a search through a space of complex hypotheses; applying Occam’s razor (aiming for a parsimonious explanation) various search techniques could be used to explore different ways to combine evidence and partial hypotheses into a unified whole

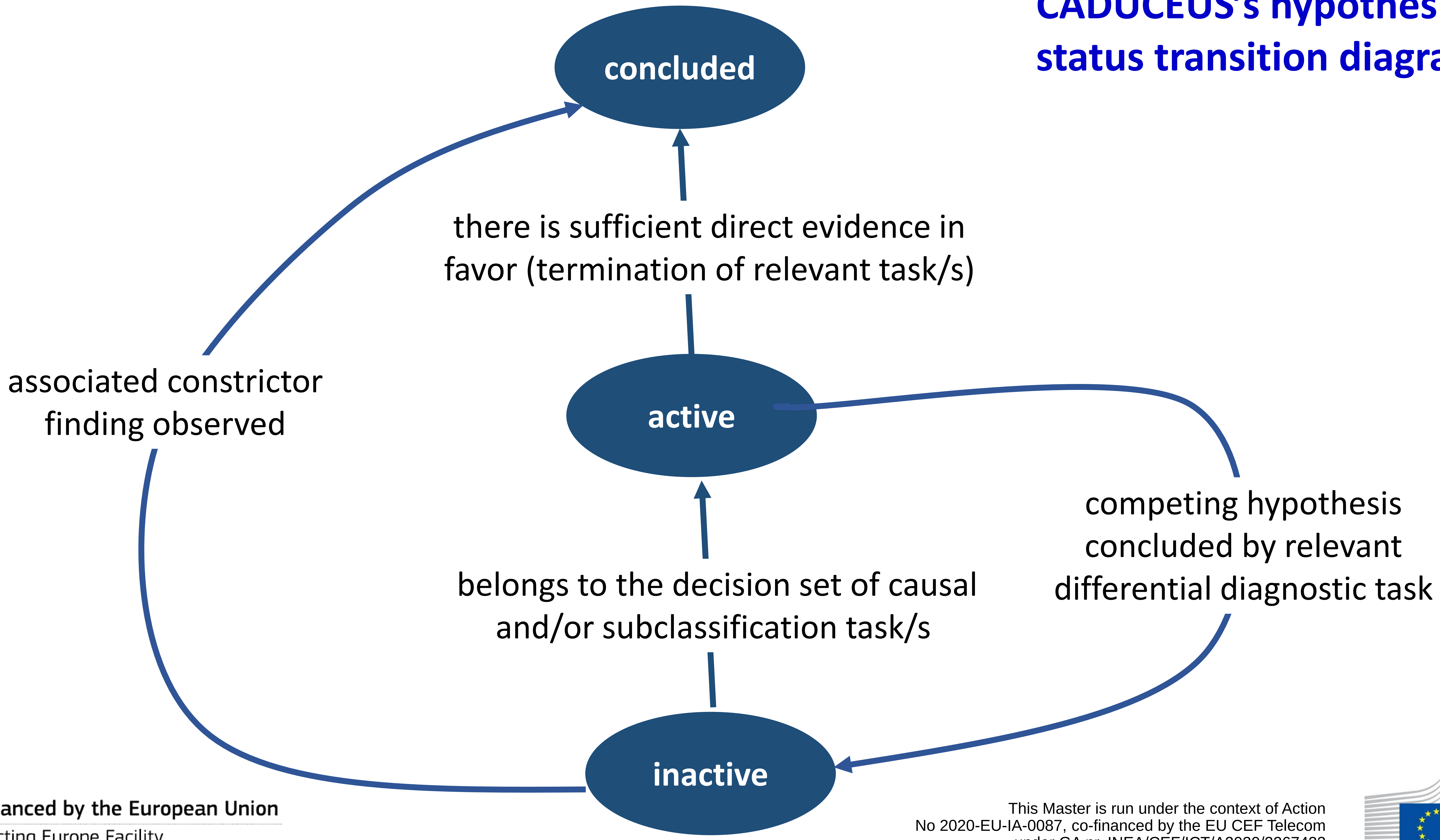
- Unfortunately, the full knowledge base for this proposed program was not constructed, though it remained an inspiring set of ideas

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

CADUCEUS' taxonomies and generalized links



CADUCEUS's hypothesis status transition diagram



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

QMR (Quick Medical Reference)

- The INTERNIST-1 algorithm was later implemented on the then new personal computer as QMR using the valuable knowledge base expanded to cover over 750 diagnoses and 5500 manifestations
- Its subsequent license to a company to produce a commercial product was unsuccessful except as an educational tool
- DXPLAIN, a program with similar structure, was developed much later by the Massachusetts General Hospital and was successfully used as a teaching tool in medical schools, but not for clinical support

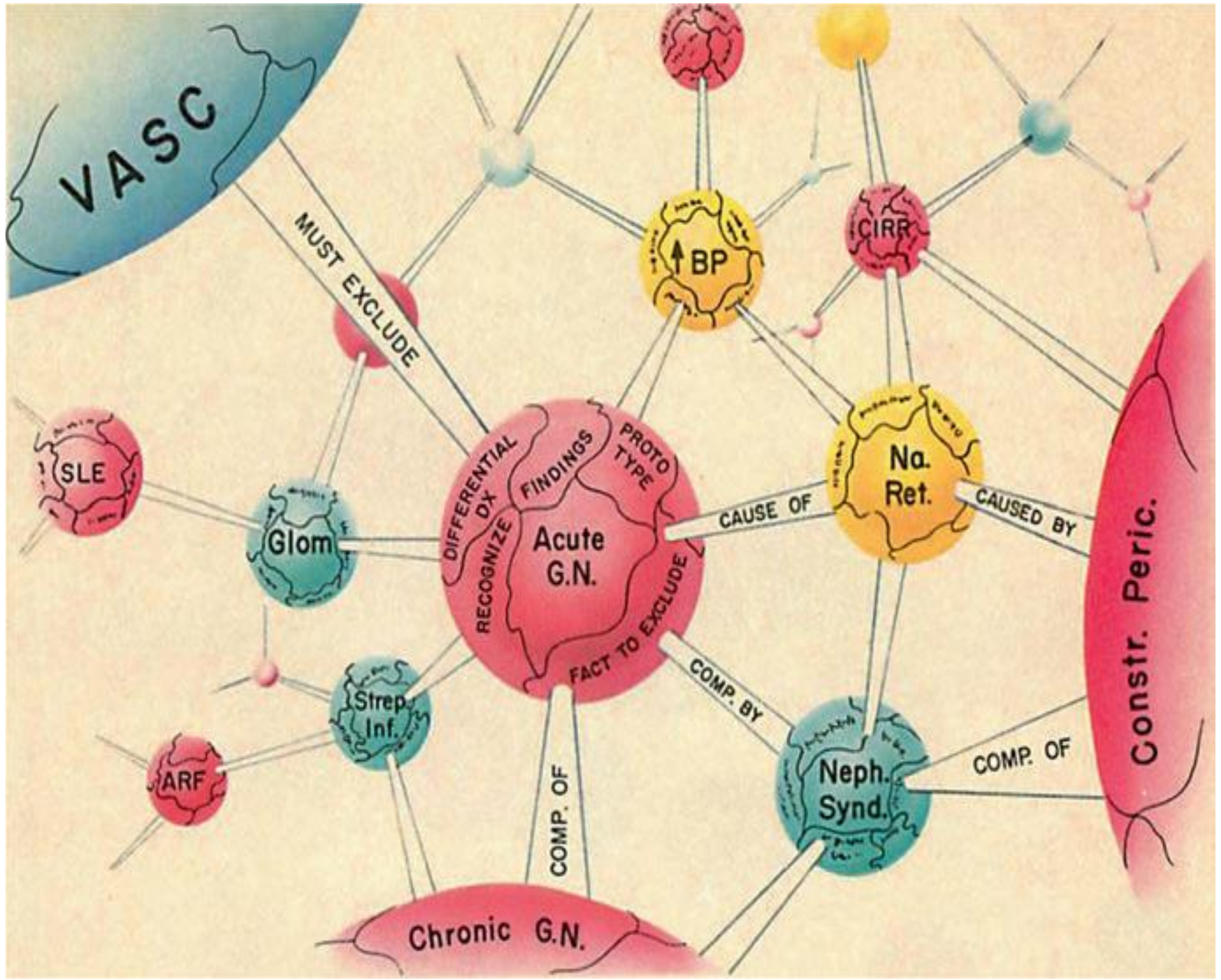
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

PIP (Present Illness Program)

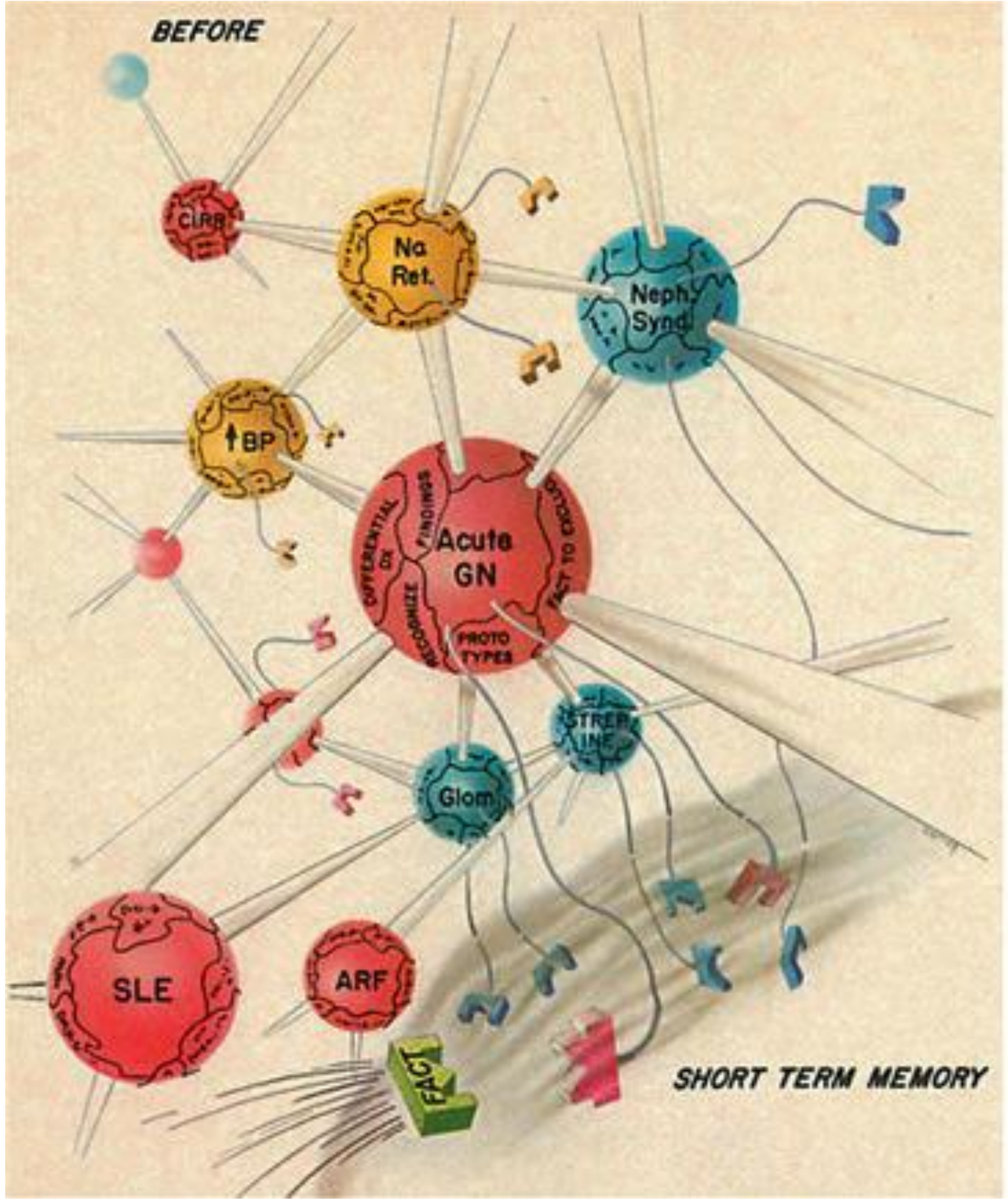
- **Function**: Simulates the behavior of an expert nephrologist in taking the history of the present illness of a patient with underlying renal disease
- **Knowledge Representation**: Frames/Semantic nets
- **Notable features**: The formation of contexts for problem solving via findings that act as abductive triggers to hypotheses; the shifts of focus via links to hypotheses with similar expectations (links to the differential diagnosis)
- **Inference**: PIP abduces hypotheses via trigger links; also disorders complementary to hypothesized disorders are abduced on the evidence of any of their associated typical expectations (findings). Hypothesized disorders are explored by testing the deductions (expected observations) inferable on that hypothesis.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

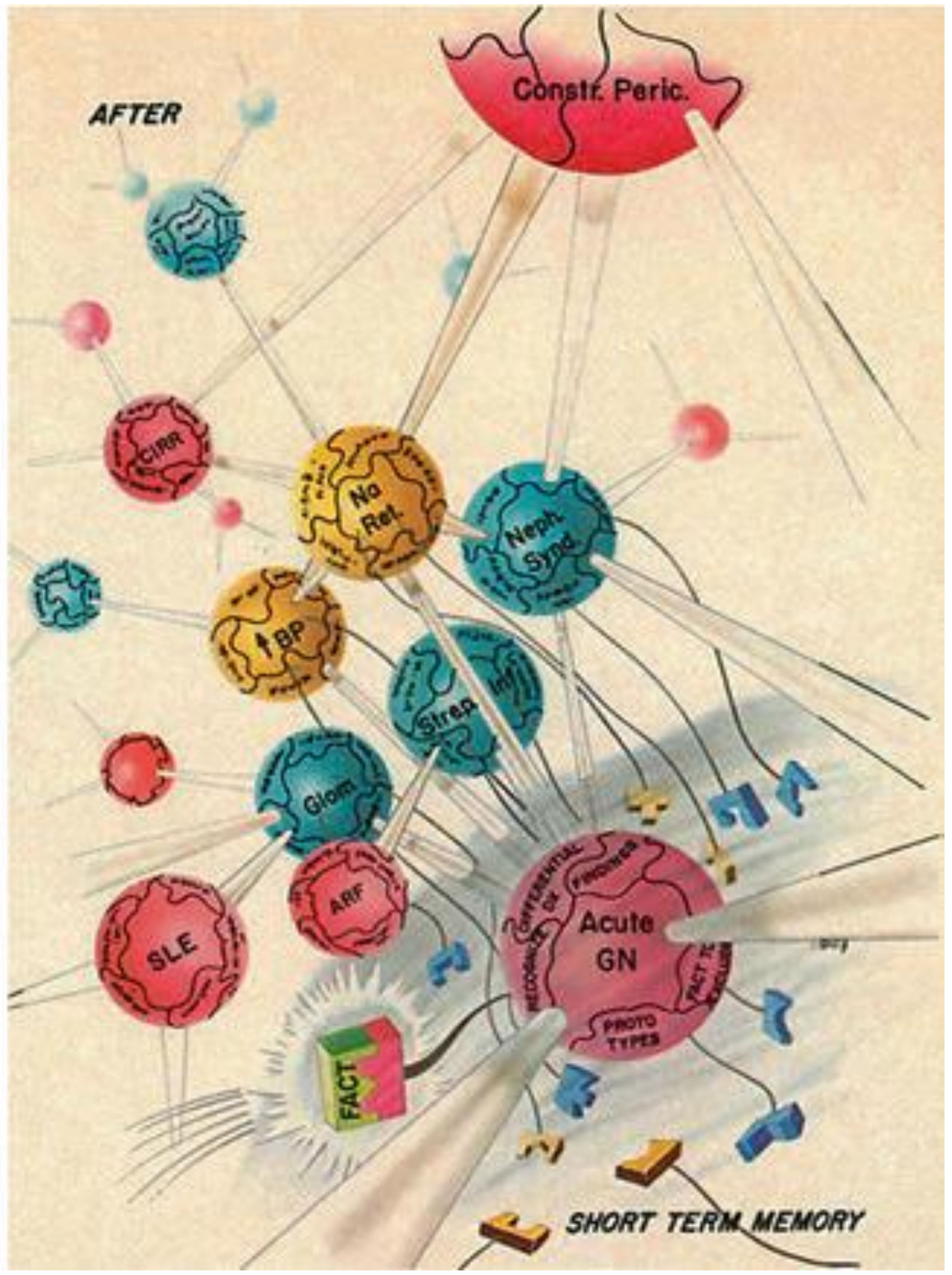


PIP’s associative (long term) memory

- Consists of a rich collection of knowledge about diseases, signs, symptoms, pathologic states, “real-world” situations, etc.
- Each point of access into the memory allows access to many related concepts through a variety of associative links shown as rods; each rod is labeled to indicate the kind of association it represents.
- Red spheres denote disease states, green spheres denote clinical states (e.g., nephrotic syndrome) and yellow spheres denote physiologic states (e.g., sodium retention).



BEFORE



AFTER

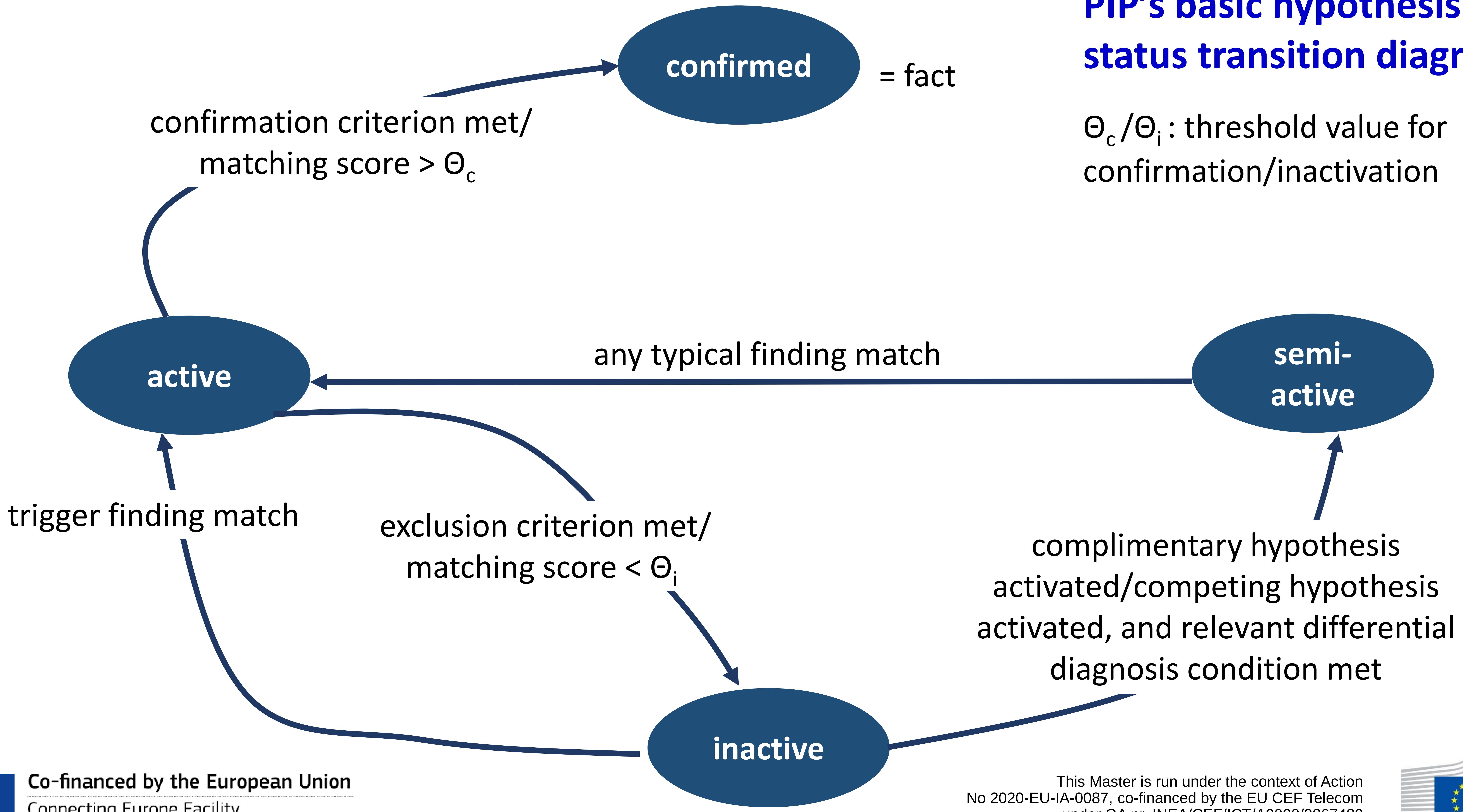
Hypothesis Generation in PIP

BEFORE: no hypothesis in short-term memory; tentacles (daemons) from some frames in long-term memory extend into the short-term memory where each constantly searches for a matching fact

AFTER: the matching of fact and deamon causes the movement of the full frame of “Acute GN” into short-memory; as a secondary effect, frames adjacent to the activated frame move closer to short-term memory and can place additional daemons therein.

PIP's basic hypothesis status transition diagram

Θ_c / Θ_i : threshold value for confirmation/inactivation

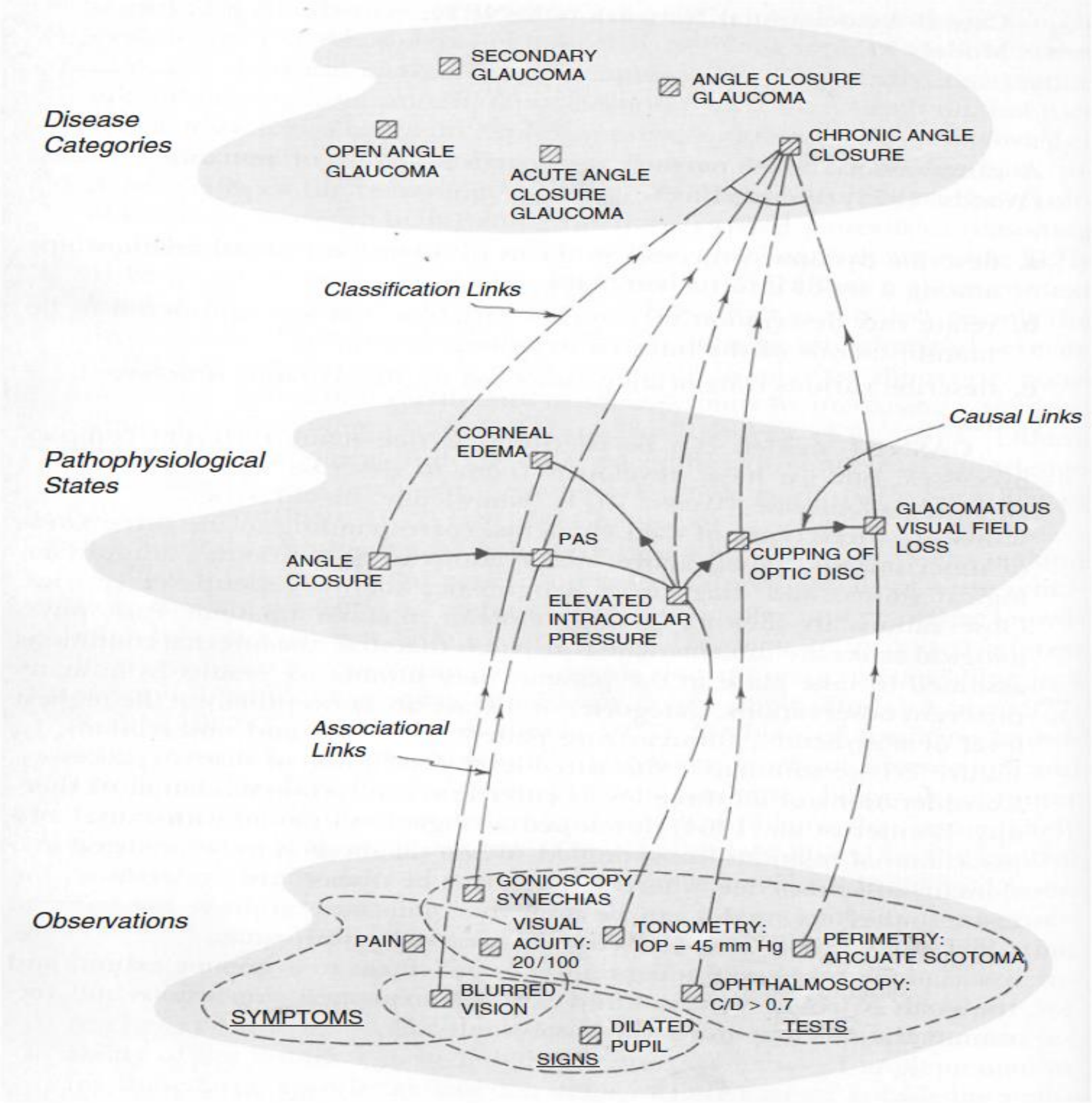


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

CASNET (Causal ASsociational NETwork)

- **Function**: Long-term management of diseases whose mechanism is well known
- **Knowledge Representation**: Causal-associational network
- **Notable features**: The division of its knowledge into distinct planes of observations and pathophysiological states; the separation between a belief measure and a promise measure of its state hypotheses; the instantiation of disease pathways in the pathophysiological plane
- **Inference**: CASNET abduces state hypotheses from direct evidence and established causal, consequent states; hypotheses are inductively supported by established causal, antecedent states. Hypotheses are infirmed/confirmed by testing the observations expected on them.

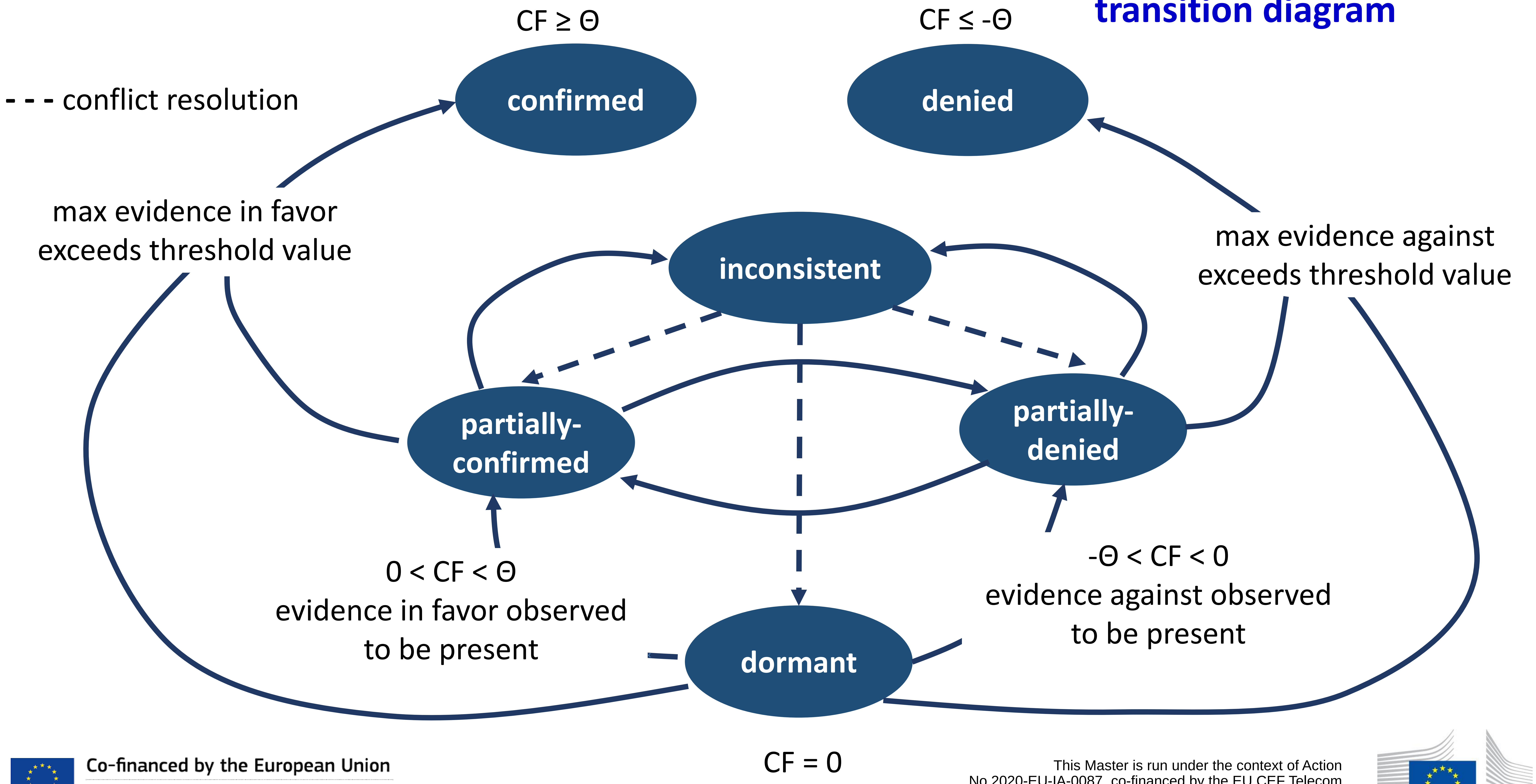


CASNET’s three level-description of a disease process:

- Causality is the central relation featuring in the middle level of pathophysiological states
- Observations (symptoms, signs, tests) could be associated with pathophysiological states or disease categories

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

CASNET's state status transition diagram



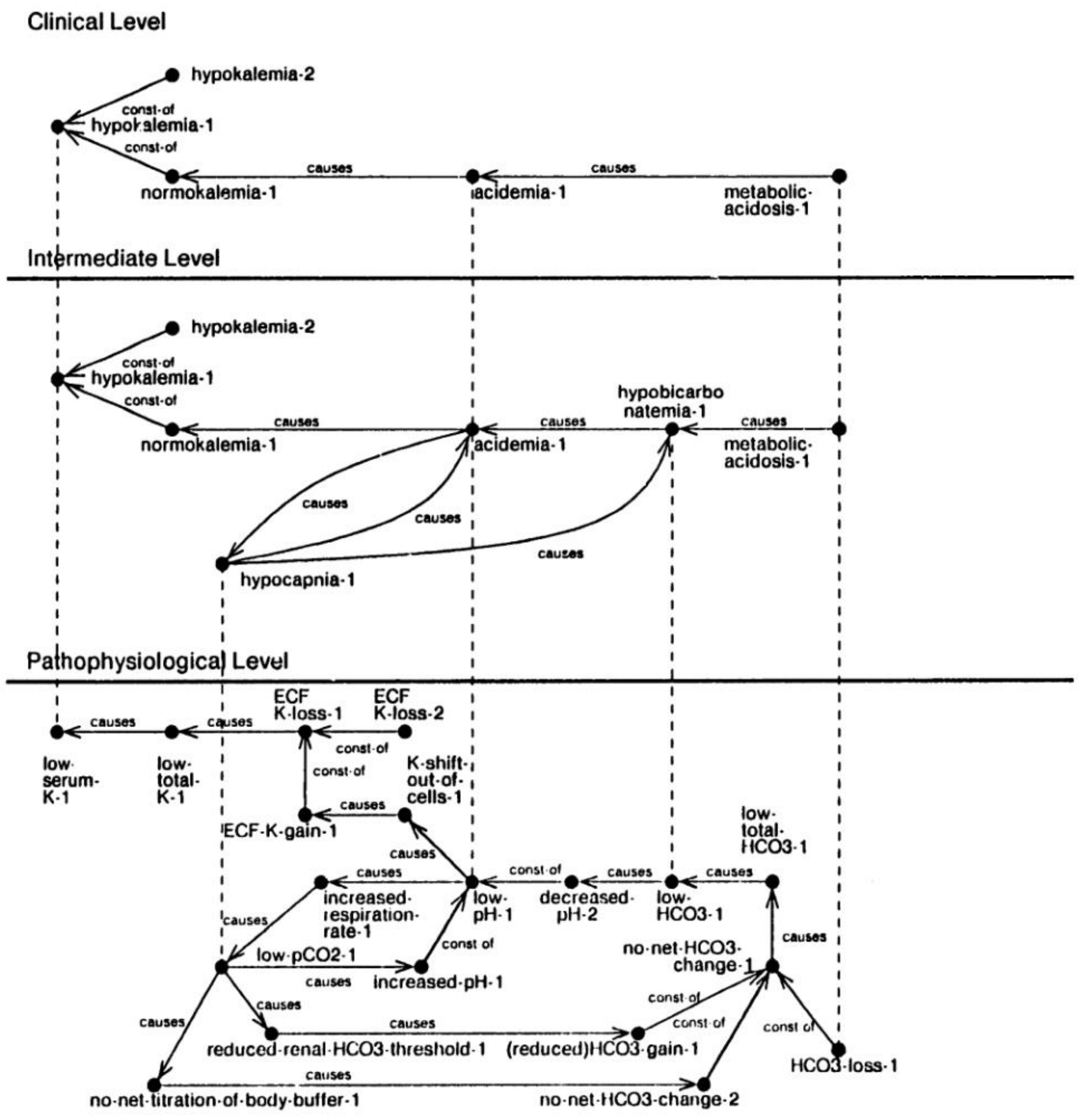
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

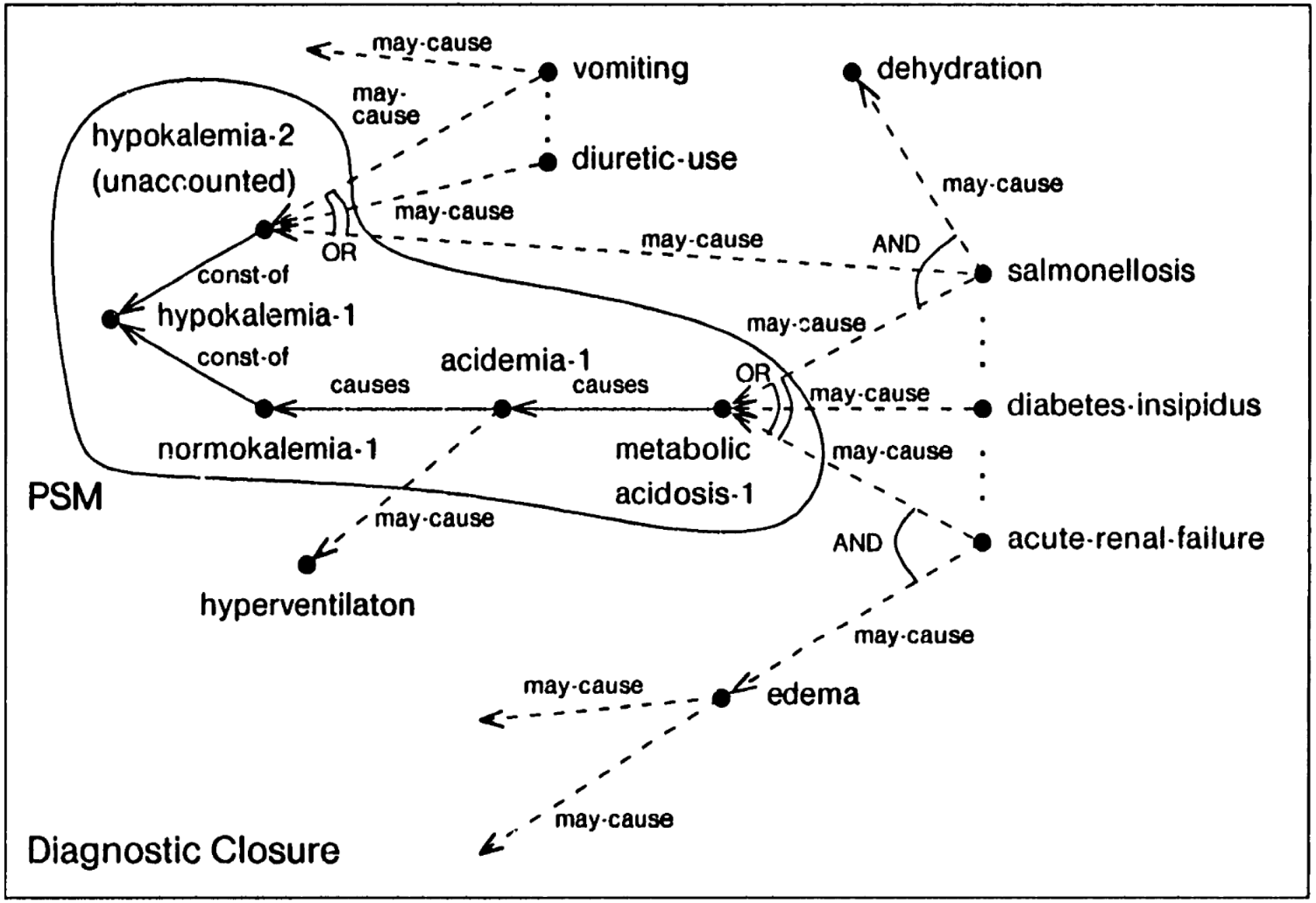
ABEL (Acid-Base and Electrolyte system)

- **Function**: Diagnosis of acid-base and electrolyte disorders
- **Knowledge Representation**: Causal networks at different levels of abstraction
- **Notable features**: The representation of a disease phenomenon at different levels of detail. The exploitation of the notion of causality in several ways: to organize the patient facts and disease hypotheses to deal with the effects of more than one disease present in a patient and to provide the basis for explanations. The capturing of the notions of adequacy and simplicity of a diagnostic possibility and hence allowing for not numeric belief measures as criteria for diagnostic reasoning.
- **Inference**: The initial hypotheses are abduced from electrolyte data using the acid-base nomograph. The construction of the diagnostic closures consisting of projecting backwards and forwards along the causal networks involves abducing hypotheses and deducing their expectations.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



ABEL's causal networks at the pathophysiological, intermediate and clinical levels



An example of diagnostic closure in ABEL

PSM – Patient Specific Model

Probabilistic Models

(material drawn from P. Szolovits and E. Alsentzer’s chapter in T.A. Cohen, V.L. Patel and E.H. Shortliffe (editors), Intelligent Systems in Medicine and Health: The Role of AI, Springer, 2022.)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Some Probabilistic Models

- ❑ **Naïve Bayes**
- ❑ **Bayesian Networks**
- ❑ **Decision Analysis and Influence Diagrams**

- ❑ Uncertainty lies at the heart of diagnostic reasoning

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Naïve Bayes

- Bayes Rule: $P(H/E) = (P(E/H)\ P(H))\ /\ P(E)$

- Some of the earliest diagnostic efforts used Naïve Bayes models to assess the impact of patient evidence (observations, laboratory test results, etc.) on the probability of disease hypotheses

- **Typical assumption**: The patient had just one disease and all the manifestations of that disease were **conditionally independent** of each other, depending only on what the actual disease was; this is the reason the models were referred to as "**naïve**"

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Uncertainty requirements of naïve models

- *A priori* probability distribution over the possible diseases
- Conditional probability distributions for each manifestation given each disease

P(D1)	**P(D2)**	**P(D3)**	**P(D4)**	**P(D5)**

P(M1/D1)	**P(M1/D2)**	**P(M1/D3)**	**P(M1/D4)**	**P(M1/D5)**
P(M2/D1)	**P(M2/D2)**	**P(M2/D3)**	**P(M2/D4)**	**P(M2/D5)**
P(M3/D1)	**P(M3/D2)**	**P(M3/D3)**	**P(M3/D4)**	**P(M3/D5)**
P(M4/D1)	**P(M4/D2)**	**P(M4/D3)**	**P(M4/D4)**	**P(M4/D5)**

$$P(E/H) = P(E \cap H)\ P(H)$$

- Such models are appropriate for diagnosis **acute illnesses** because newly presenting facts about a patient are likely to be caused by one rather than multiple diseases; hence they are not appropriate for complex cases where multiple diseases are typically simultaneously present

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Optimizing question ordering in Naïve Bayes models

- **Entropy minimization heuristic** for dynamically selecting the optimal next question about manifestations, thus allowing the program to ask for only a small fraction of all possible manifestations known in the model and to quickly reach a probability distribution that shows one disease as highly likely:
  - Choose the question to ask that **minimizes the expected entropy** of the probability distribution resulting from asking that question
  - Say that a question has k answers; using Bays Rule the posterior probability distribution of the diseases is computed for each of the possible answers; compute the entropy of each distribution and weight each entropy by the probability of getting that answer, thus calculating the expected **information gain** from asking that question
  - The conditional independence of manifestations allows the application of this method

A note on entropy in AI

Entropy calculates the impurity of the group that helps to make a better split of data. It is a measure that is used to check the homogeneity of the data, calculated using the following formula:

Entropy = $\sum -p_i \log_2^{(p_i)}$

where p_i is the probability of the ith class

Note also that the entropy minimization heuristic does not take into consideration the **"cost"** entailed in answering a given question from the perspective of the patient (discomfort, risk, time required, monetary cost, etc.)

Bayesian Networks

- Allow the presence of multiple diseases
- Each disease has *a priori* probability, independent of the others
- Each finding depends on some subset of the diseases, but the findings are conditionally independent
- When a finding can be caused by multiple diseases, the conditional probability table for that finding must have an entry for all possible combinations of the presence or absence of each causing disease; to combat the given exponential explosion the **noisy-or assumption** is adopted:
  - A finding is absent only if none of its possible causes actually cause it

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Noisy-or assumption

- A finding, S, has just one possible disease cause, D:
  - P(S) = P(D) P(S/D)
- A finding, S, could be caused by any of D_1, D_2, …, D_k:
  - $P(S/d_1,d_2,\ldots,d_k) = 1 – (1 – P(d_1)P(S/d_1))$

$$(1 – P(d_2)P(S/d_2))$$

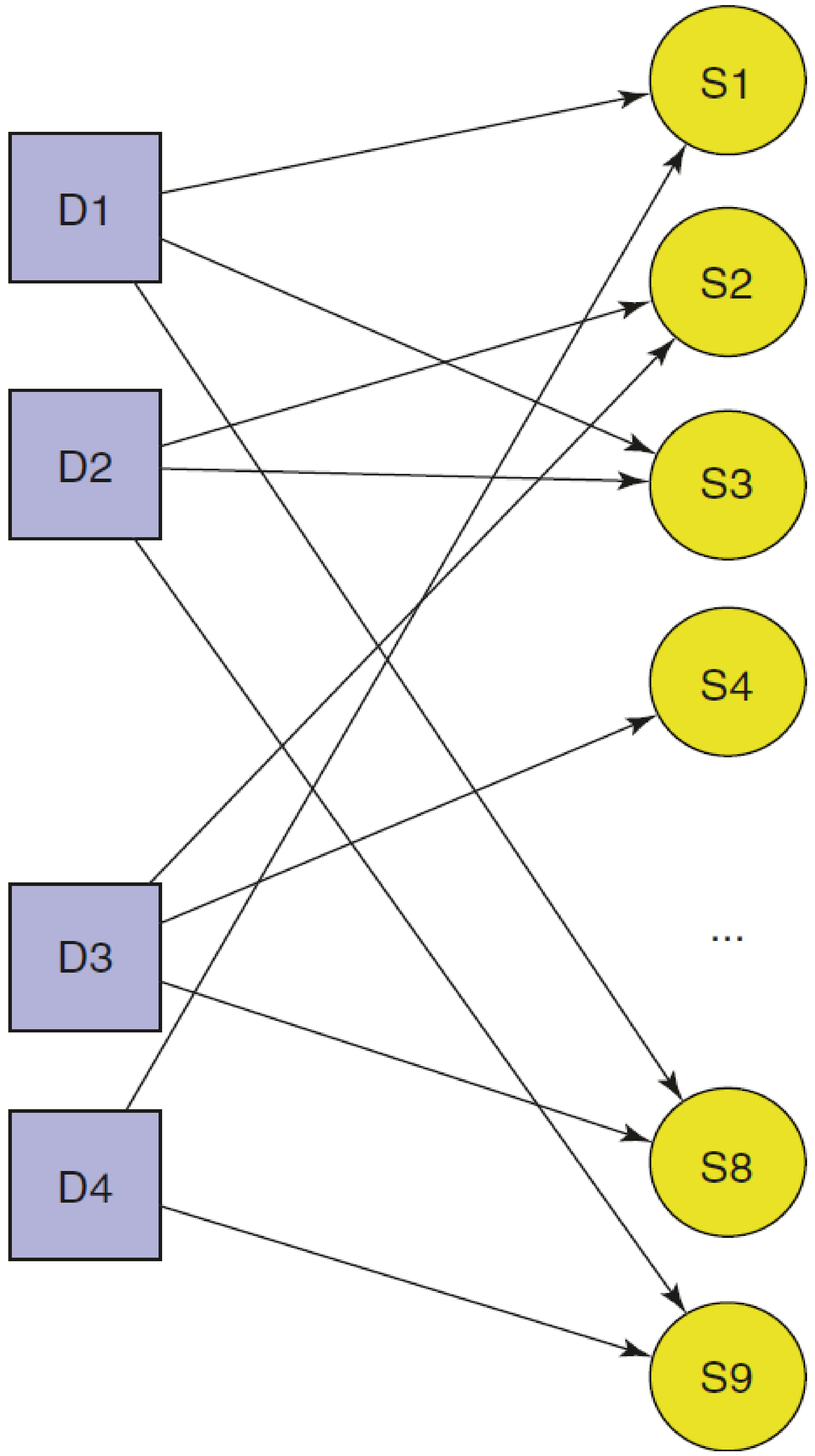
$$\ldots..$$

$$(1 – P(d_k)P(S/d_k))$$

where d_i is whether D_i is present or absent; this model also assumes that there is a “leak” term, namely that the finding might occur with some small probability even if all its causes are absent

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bipartite Bayesian Network



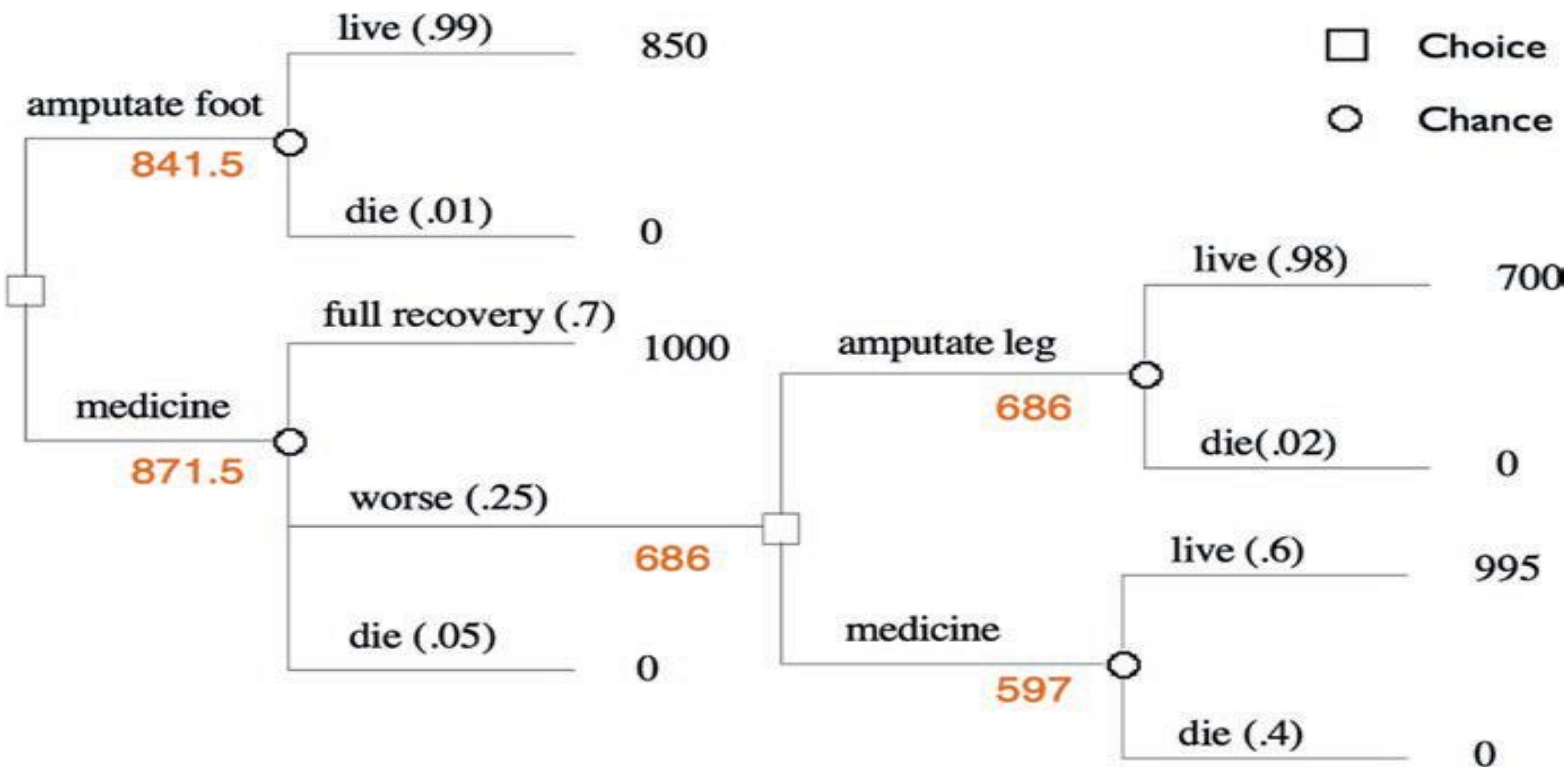
- The diseases may occur simultaneously but are probabilistically independent of each other.
- The symptoms depend only on the diseases and are conditionally independent of each other

Decision Analysis and Influence Diagrams

- **Principle of rationality**: the right action to take is the one with the best expected outcome
- **Decision analysis**: assign numerical values to various outcomes and probabilities to the effects of various actions conditioned on what ails the patient
- **Decision trees** (used in simple cases) that contain:
  - **Choice nodes** representing the choices facing the clinician, that lead to
  - **Chance nodes** representing the probabilistic outcomes of the chosen actions
  - **Value nodes**, at the leaves of such a decision tree, showing the value of that outcome

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Example: Decision analysis for how to treat an elderly man with gangrenous foot



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Probabilities and Values

- Ascertaining probabilities is a difficult task because they should reflect the case under consideration
- Ascertaining numerical values to various outcomes is even harder because these should reflect the views of the patient
  - For example, on a scale of 0 (death) to 1000 (full health) what is the value of living with an amputated leg?

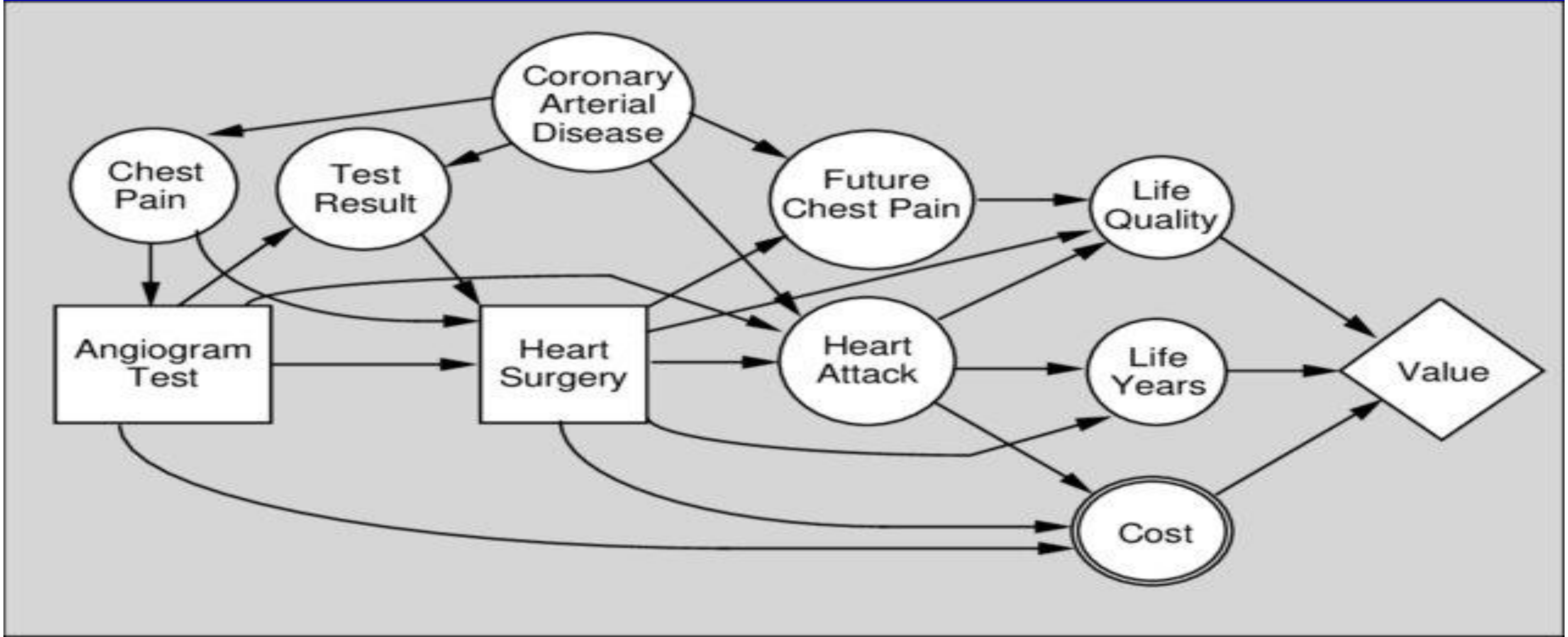
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Influence Diagrams

- Influence diagrams are **directed graphs** with nodes representing events and decisions and arrows between them representing the (probabilistic) dependencies.
  - An arrow denotes an **influence**
  - An arrow from A to B means that knowing A would directly affect our belief or expectation about the value of B
  - An influence expresses knowledge about relevance
- They combine decision analysis with Bayesian networks for the potentially more complex probabilistic relations among choices, chances and decisions, thus providing a much more compact representation of complex decision problems and avoiding having to specify the order of decisions as in a decision tree.

Example influence diagram for a patient with heart disease



Reinforcement Learning

- A method that helps a decision maker to **choose the best course of action** under any modeled circumstance
- An action is typically modeled to have
  - An **immediate reward**, e.g., the patient's elevated heart rate decreases, and
  - A **long-term reward**, e.g., the patient survives the hospital stay
- Potential action evaluation:
  - Combines the immediate award with the discounted expected sequence of immediate and long-term rewards anticipated from the possible future states resulting from the action (obtained in the manner of a decision tree or influence diagram)
- Given a large database of past treatments of other patients, it is possible to estimate the relevant expected rewards, but not for actions that were rarely if ever taken in the past; sometimes though it may be helpful to try a different, less explored therapy – simulate first using retrospective data, and not experiment with real patients; a randomized clinical trial can be viewed as a step in such an exploration strategy
- RL is gaining popularity as a way to exploit data on complex sequences of past decisions

Causality

- Very important in medicine; qualitative, quantitative or hybrid approaches
- CADUCEUS, CASNET and ABEL are amongst the pioneering medical knowledge-based systems that exhibit elaborate causal models
- If we had a complete understanding of how the human body works, we could build mechanistic models that could predict the response to various conditions and treatments with precision
- Notable attempt to build such a deep model is Guyton’s cardiovascular model
  - The Guyton-Coleman implementation of this model led to new insights into the relationship between cardiac output, blood pressure and control of sodium
  - NASA used the model to predict the effects of weightlessness on the circulatory system of astronauts as a safety check as they prepared for space travel
  - The model continues to be developed and is now called Digital Human, including about 5000 variables covering renal, respiratory, endocrine, neural and metabolic physiology

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Knowledge Acquisition: the use of ontologies

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

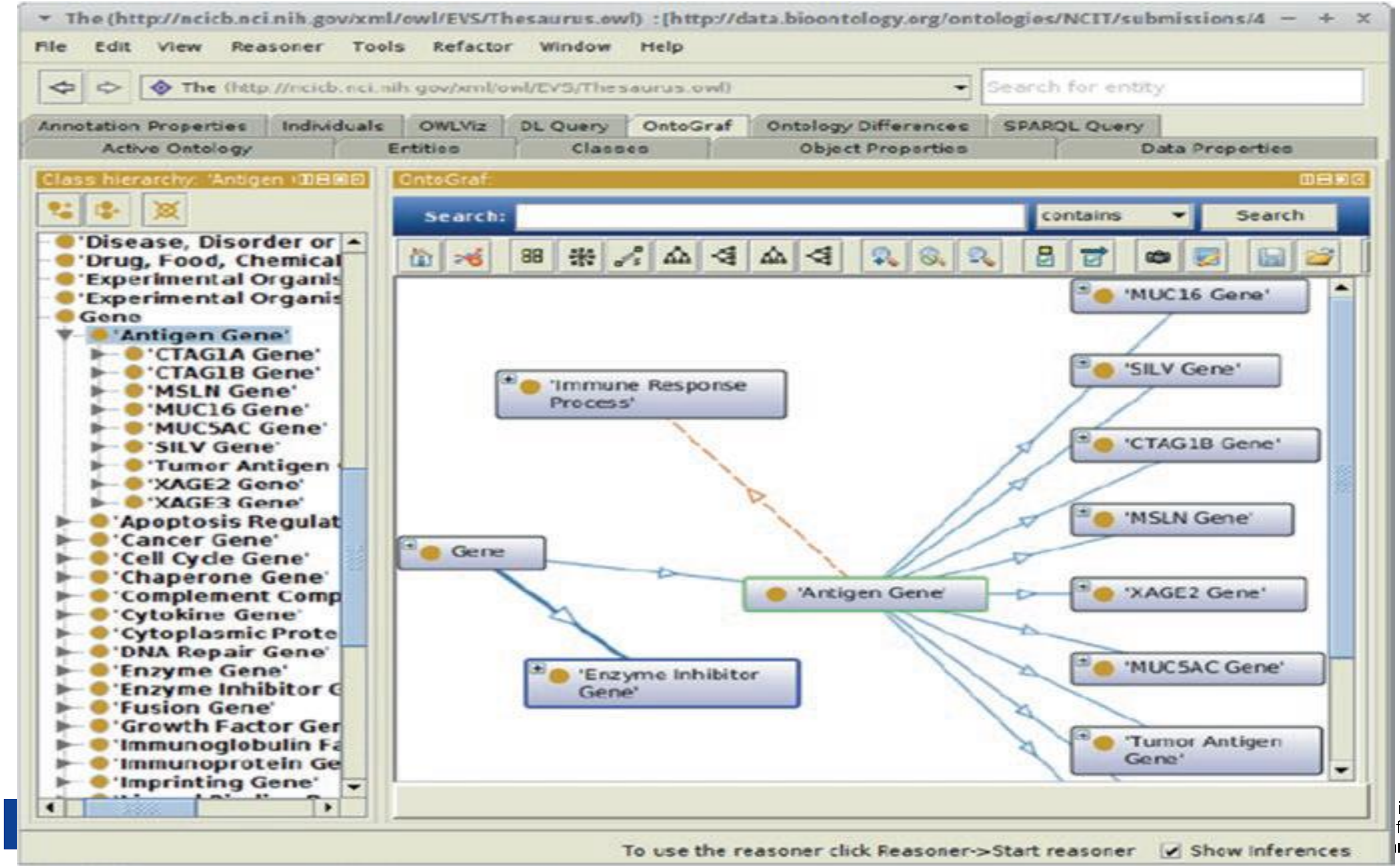
Knowledge Acquisition

- The well-known bottleneck of knowledge-based systems striving to incorporate knowledge models of specialized domains
- Tailor-made knowledge acquisition systems (e.g., TEIRESIAS for MYCIN) tried to semi-automate the process, focusing though on the representation level
- Knowledge engineering methodologies, like CommonKADS, proposed ways of combating the complexity of building knowledge-based systems focusing on conceptual modeling, reusability and the adoption of software engineering elements
- More recent approaches involve
  - Taxonomic organizations of concepts into **ontologies** that describe each concept in terms of its super- and sub-categories and its attributes and constraints on them, adopting a frame-like view; classification and inheritance are the main inference tasks
  - **Knowledge graphs**, which can be constructed manually or via unsupervised methods that exploit the co-occurrence of terms in sentences, paragraphs or articles

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Protégé: Ontology construction tool



The Protégé system showing relationships among gene concepts from the National Cancer Institute (NCI) Thesaurus (gene ontology)

is run under the context of Action
financed by the EU CEF Telecom
r. INEA/CEF/ICT/A2020/2267423

Creating ontologies for new domains

- Protégé guides the user in defining a hierarchy of concepts (called classes) in the domain, including the specification of defaults and constraints on properties of the classes
  - What is the value of a property? Single-valued or multi-valued? Are its values simple or other classes?
  - Is a new class a subclass of an existing class, or does it contain instances of the superclass with specific properties?
- A description logic is used to define class properties, trading off between expressiveness and computational tractability of doing inference with it

- Important use of ontologies: integrate the concepts used in different clinical systems – an early use of this technology was the GALEN project, using the GRAIL description logic

The Unified Medical Language System (UMLS)

- Pioneered by the U.S. National Library of Medicine
- Created a Metathesaurus, instead of trying to build a comprehensive ontology
- Combines concepts and taxonomic and other relationships from over 200 different terminological systems
- Machine and manual curation has assured that the nearly 13 million terms from 25 languages (though mostly English) map to over 4.4 million concepts, to help coordinate information among the terminologies
- Reliably mapping many terms to a single concept is a major aid to clinical natural language processing
  - E.g., “acute myocardial infarct”, “AMI” and “heart attack” represent the same concept
- UMLS also provides some linguistic tools for lemmatization and assigns a semantic category to each concept from amongst a set of 189 such categories

UMLS overall goals and assumptions as publicized by the NLM UMLS Team

"The Unified Medical Language System (UMLS) project is … designed to facilitate the retrieval and integration of information from many machine-readable information sources, including descriptions of the biomedical literature, clinical records, factual databanks, and medical knowledge bases. The UMLS project is not an attempt to impose either a single standard vocabulary, a single standard record format, or a single medical knowledge base on the biomedical community. The UMLS approach assumes that diversity will continue to exist and therefore seeks to provide products that can compensate for differences in the vocabularies or coding schemes used in different systems, as well as for differences in the terminology employed by system users."

However, because the content of UMLS comes from many separately developed databases, taxonomic inconsistencies do arise, and relations other than the taxonomic are sparse

The CYC project

- **CYC** is a long-term AI project that aims to assemble a comprehensive ontology and knowledge base that spans the basic concepts and rules about how the world works, also capturing common sense knowledge.
- Manual curation of large knowledge bases is phenomenally costly and subject to gaps because it is difficult for people to think of everything needed
- CYC's knowledge base has been developing for over 30 years, and among many applications are some in healthcare
- CYC's knowledge is represented in a logical form and contains 10,000 predicates over millions of concepts, encoding 25 million assertions in higher-order logic
- It also includes several specialized inference engines

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Intelligent Data Analysis in Medicine

- ❑ Data Abstraction
- ❑ Data Mining

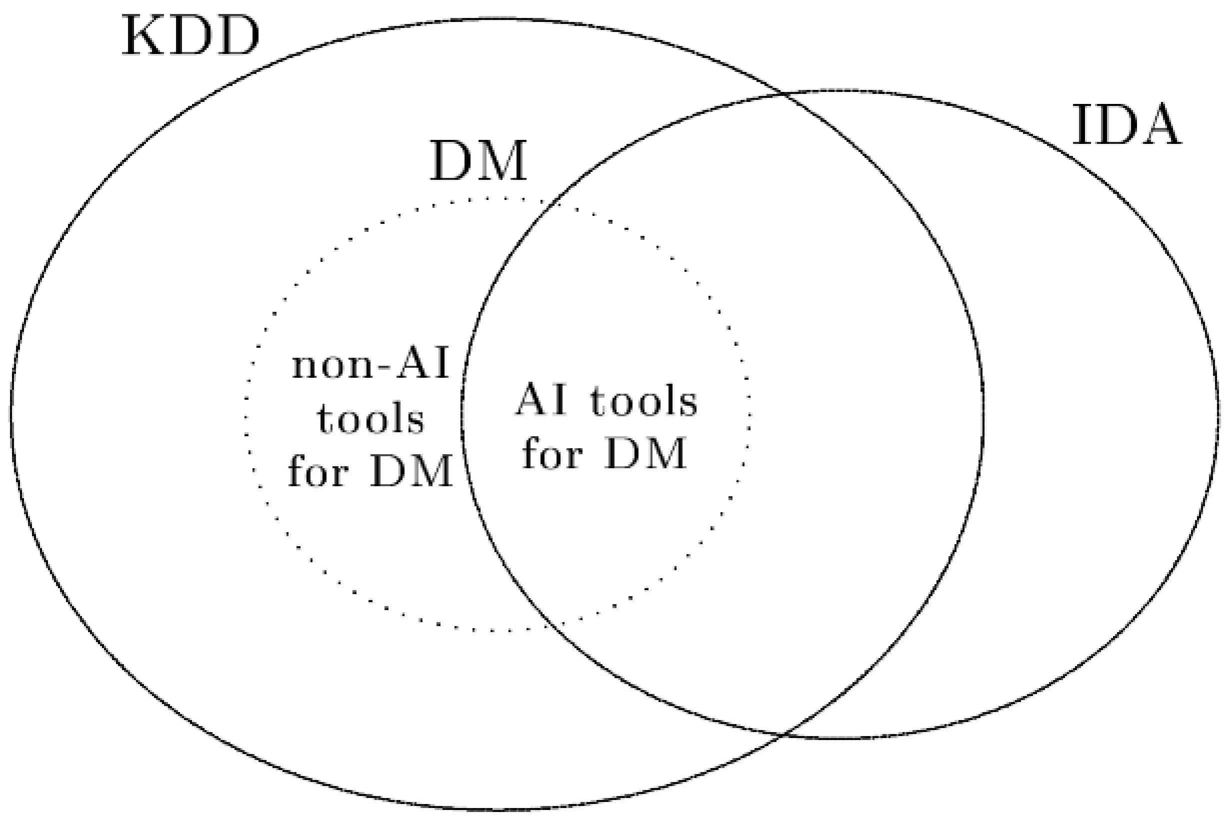
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

IDA, KDD and DM

Intelligent Data Analysis (IDA) encompasses statistical, pattern recognition, machine learning, data abstraction and visualization tools to support the analysis of data and discovery of principles that are encoded within the data.

Knowledge Discovery in Databases (KDD) is a process consisting of the following steps: understanding the domain, forming the dataset and cleaning the data, extracting the regularities hidden in the data thus formulating knowledge in the form of patterns, rules, etc. [this step is usually referred to as **Data Mining** (DM)], postprocessing of discovered knowledge and exploitation of results.



IDA and KDD have in common the topic of investigation, which is data analysis, and they share many common methods; however, IDA uses AI methods and tools while KDD employs both AI and non-AI methods. Moreover, KDD is typically concerned with the extraction of knowledge from very large datasets, whereas in IDA the datasets are either large or moderately sized.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The role of IDA systems in a clinical setting

- Their role is that of an intelligent assistant that tries to **bridge the gap between data gathering and data comprehension,** in order to enable the physician to perform his task more efficiently and effectively; the physician must have at his disposal the right information at the right time.
- Nowadays is **possible to store large volumes of data from diverse sources** on electronic media; data could be on a single case (e.g., one patient) or multiple cases.
- **Raw data are of little direct use**; their sheer volume and/or very specific level makes impossible their operationalization in problem solving.
- But such data can be converted to a mine of information wealth if the real gems of **information** are **extracted from the data** by computationally intelligent means.
- Useful, operational information/knowledge is expressed at the **right level of abstraction** and made readily available **to support the decision making**.

The globality of data and information calls for …

- The provision of standards in terminology, vocabularies, and formats to support multilingualism and sharing of data
- Standards for the abstraction and visualization of data
- Standards for interfaces between different sources of data
- Integration of heterogeneous types of data, including images and signals
- Standards for electronic patient records
- Reusability of data, knowledge, and tools

Co-financed by the European Union
Connecting Europe Facility

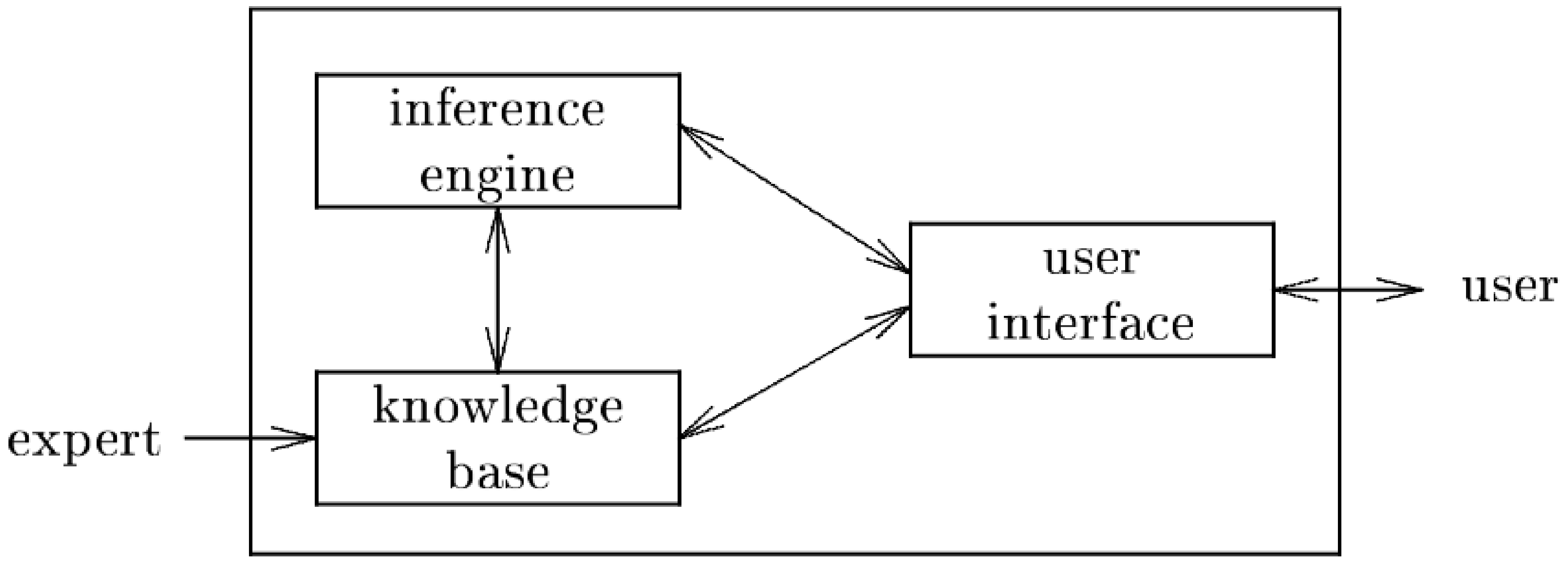
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Data Abstraction and Data Mining

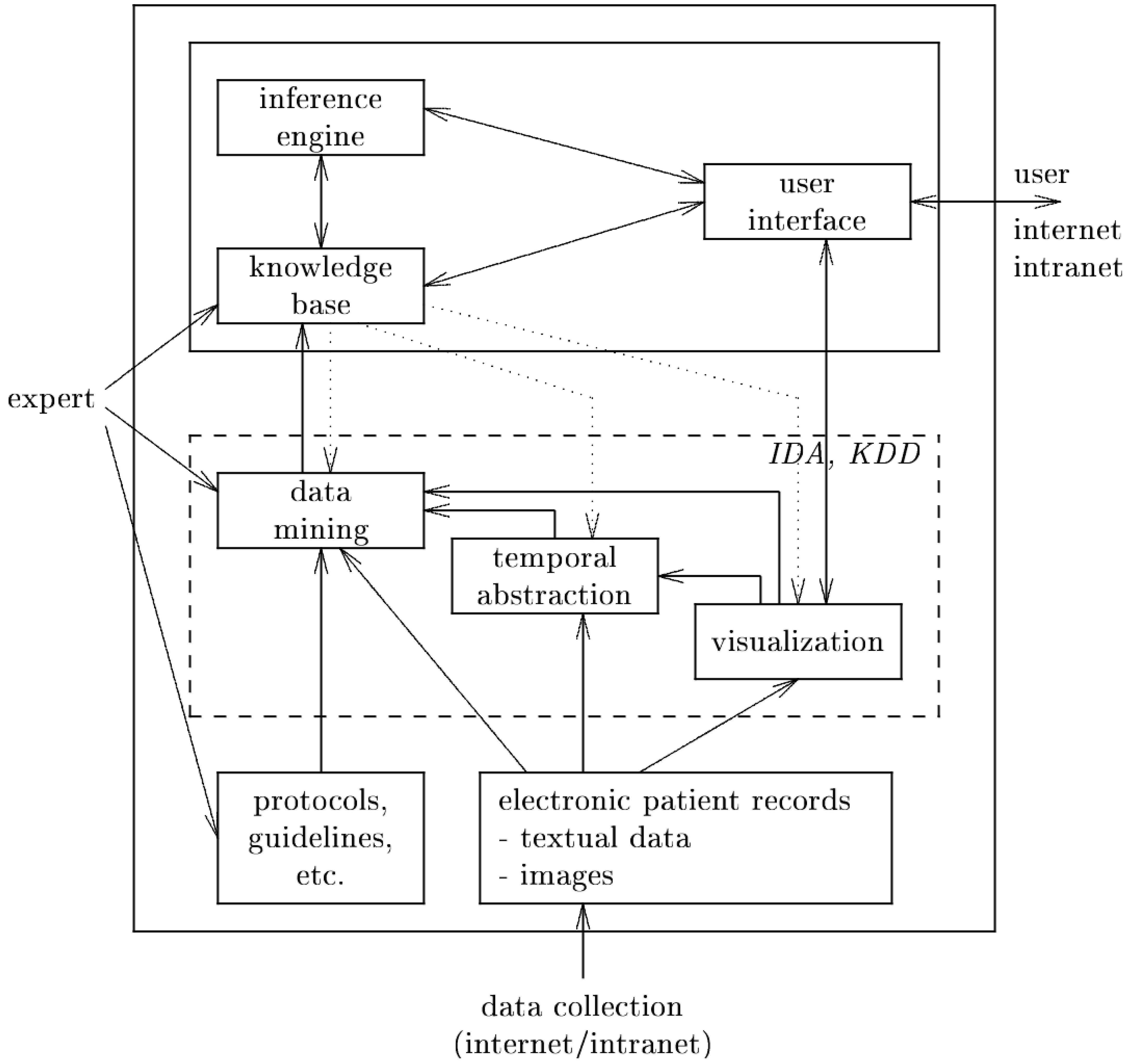
- IDA methods applied to supporting decision making in medicine can be classified into two main categories: data abstraction and data mining.
- **Data abstraction** is concerned with the intelligent interpretation of patient data in a context-sensitive manner and the presentation of such interpretations in a visual or symbolic form, where the temporal dimension in the representation and intelligent interpretation of patient data is of primary importance.
- **Data mining** is concerned with the analysis and extraction (discovery) of medical knowledge from data, aimed at supporting diagnostic, screening, prognostic, monitoring, therapy support, or overall patient management tasks.
- Most DM methods belong to **machine learning** and most data abstraction methods perform **temporal abstraction**.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Knowledge versus Data



The classical architecture of an expert system



A decision support schema in the age of data: arrows denote normal information flow, and dotted arrows represent information flow in processes involving iteration and loops between the different steps of the IDA process

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Data Abstraction Methods

- **Support specific knowledge-based problem-solving activities** (data interpretation, diagnosis, prognosis, monitoring, etc.) by extracting useful abstractions from the raw, mostly numeric data.
- **Temporal data abstraction** methods represent an important subgroup where the processed data are temporal.
- The derivation of abstractions is often done in a **context-sensitive** and/or **distributed** manner and it applies to discrete and continuous supplies of data.
- The abstraction can be performed over a single case (e.g., a single patient) or over a collection of cases.
- The data abstraction methods are **knowledge-driven** (both general and specialist knowledge)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Data Mining Methods

- Extract knowledge, preferably in a **meaningful and understandable symbolic** form.
- Most frequently applied methods are **supervised symbolic machine learning methods**, e.g., effective tools for inductive learning exist that can be used to generate understandable diagnostic and prognostic rules.
- Other methods include symbolic clustering, discovery of concept hierarchies, qualitative model discovery and learning of probabilistic causal networks.
- Sub-symbolic learning (e.g., nearest-neighbor method, Bayesian classifier, and (non-symbolic) clustering) and case-based reasoning methods can also be classified in the DM category.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Simple atemporal data abstractions

- **Qualitative abstractions**, where a numeric expression is mapped to a qualitative expression; such abstractions are based on simple associational knowledge, e.g., (“a temperature of at least 39 degrees C”, “fever”).
- **Generalization abstraction**, where an instance is mapped to (one of) its classes; such abstractions are based on strict or tangled concept taxonomies.
- **Definitional abstraction**, where a datum from one conceptual category is mapped to a datum in another conceptual category that happens to be its definitional counterpart in the other context; the resulting concept must be more abstract than the originating concept, e.g., it refers to something more easily observable. The knowledge driving such abstractions consists of simple associations between concepts across different categories.
- In an atemporal situation everything is assumed to refer to “now”: $\boldsymbol{holds(P,D)} \rightarrow \boldsymbol{holds(P,abs(D))}$

Why time is important … particularly in medicine

- Time is intrinsic to many problem domains where **dynamic situations** arise – temporal reasoning is largely **commonsense reasoning**
- In medicine:
  - Disease processes evolve in time
  - Patient records give the history of patients
  - Therapeutic actions, like all actions, are indescribable without considering time
- The modelling of time enables a more accurate formation of potential solutions:
  - The presence of an abnormality may not be diagnostically significant as such, but its specific pattern of appearance is
  - The expected picture of a disease is different, depending on the state of its evolution
- These call for **temporal abstractions**

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Temporal data abstraction

- **Temporal data abstraction** is a fundamental intermediate reasoning process for the intelligent interpretation of temporal data in support of tasks such as diagnosis, monitoring, etc.; unlike atemporal abstraction where a single datum is abstracted, here sets of data, or more accurately time series of data, can be abstracted to single data.
- **Background domain knowledge** can be effectively utilized in the context of temporal data abstraction, e.g., persistence semantics of concepts
- **Commonsense reasoning** involves the intuitive handling of multiple time granularities and temporal relations such as before, overlaps, and disjoint.
- Patient data can be considered as **temporal objects**, where a temporal object is an integral association between an item of information and a time.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Types of temporal data abstractions

- **Merge abstraction**, where a collection of data, sharing the same **concatenable property** and whose temporal aspects collectively form a (possibly overlapping) chain (at some time granularity) are abstracted to a single datum with the given property whose temporal aspect is the maximal time interval spanning the original data; also known as **state abstraction**.
- **Persistence abstraction**, where again the aim is to derive maximal intervals spanning the extent of some property; here, though, there could be just one datum on that property, and hence the difficulty is in filling the gaps by ‘seeing’ both backward and forward in time from the specific, discrete recording of the given property.
  - **Default persistence rule**: some property is assumed to persist indefinitely until some event (e.g., a therapy) is known to have taken place and this terminates the persistence of the property.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Types of temporal data abstractions

- **Trend abstraction**, where the aim is to derive the significant changes and the rates of change in the progression of some parameter; it entails merge and persistence abstraction in order to derive the extents where there is no change in the value of the given parameter.
- **Periodic abstraction**, where repetitive occurrences with some regularity in the pattern of repetition are derived, e.g., headache every morning for a week of increasing severity:
  - repetition element, e.g., headache – it can be of any order of complexity, e.g. it could itself be a periodic abstraction, or a trend abstraction
  - repetition pattern, e.g., every morning for a week
  - progression pattern, e.g., increasing severity
- The data abstraction types can be combined in a multitude of ways, yielding **complex abstractions.**

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

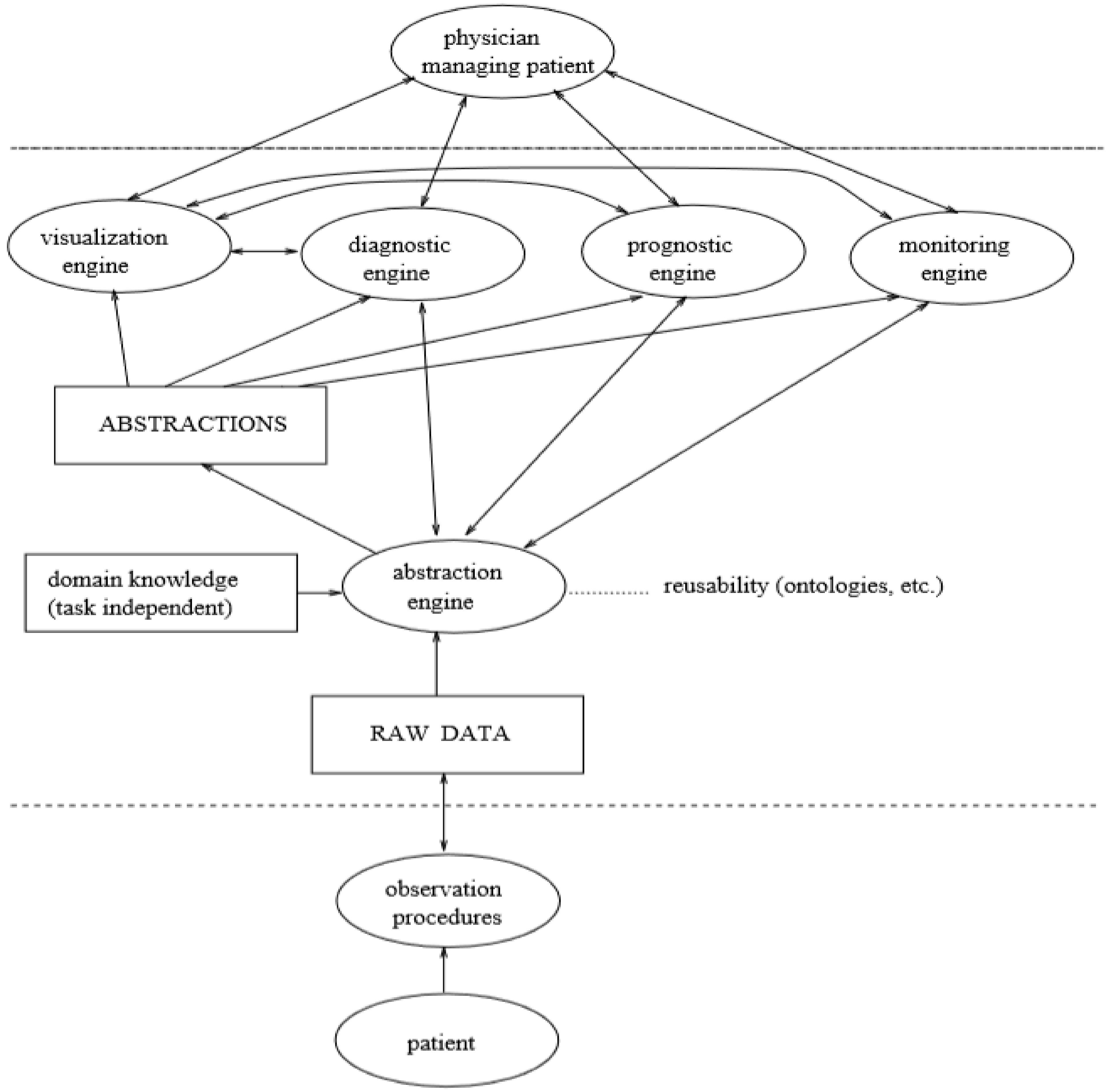
Modes of deployment of data abstraction

- **Directed or goal-driven,** i.e., the problem-solving system (in exploring its hypothesis space) predicts various abstractions that the data abstraction process is required to corroborate against the raw patient data.
- **Nondirected or event-driven**, e.g., in a monitoring system; the aim is to comprehensively interpret all the data covered by a (moving) time window.
- Nondirected data abstraction can be used in a **stand-alone fashion** where the derived abstractions should be presented to the user in a visual form; **visualization** is also of relevance when a data abstraction process is not used in a stand-alone fashion as it is a good way of justifying the reasoning.
- **Truth maintenance** is of relevance to any inference system: as raw data may be received out of temporal sequence, abstractions referring to the present may need to be modified, or abstractions referring to the past are revoked by new data.

Co-financed by the European Union
Connecting Europe Facility

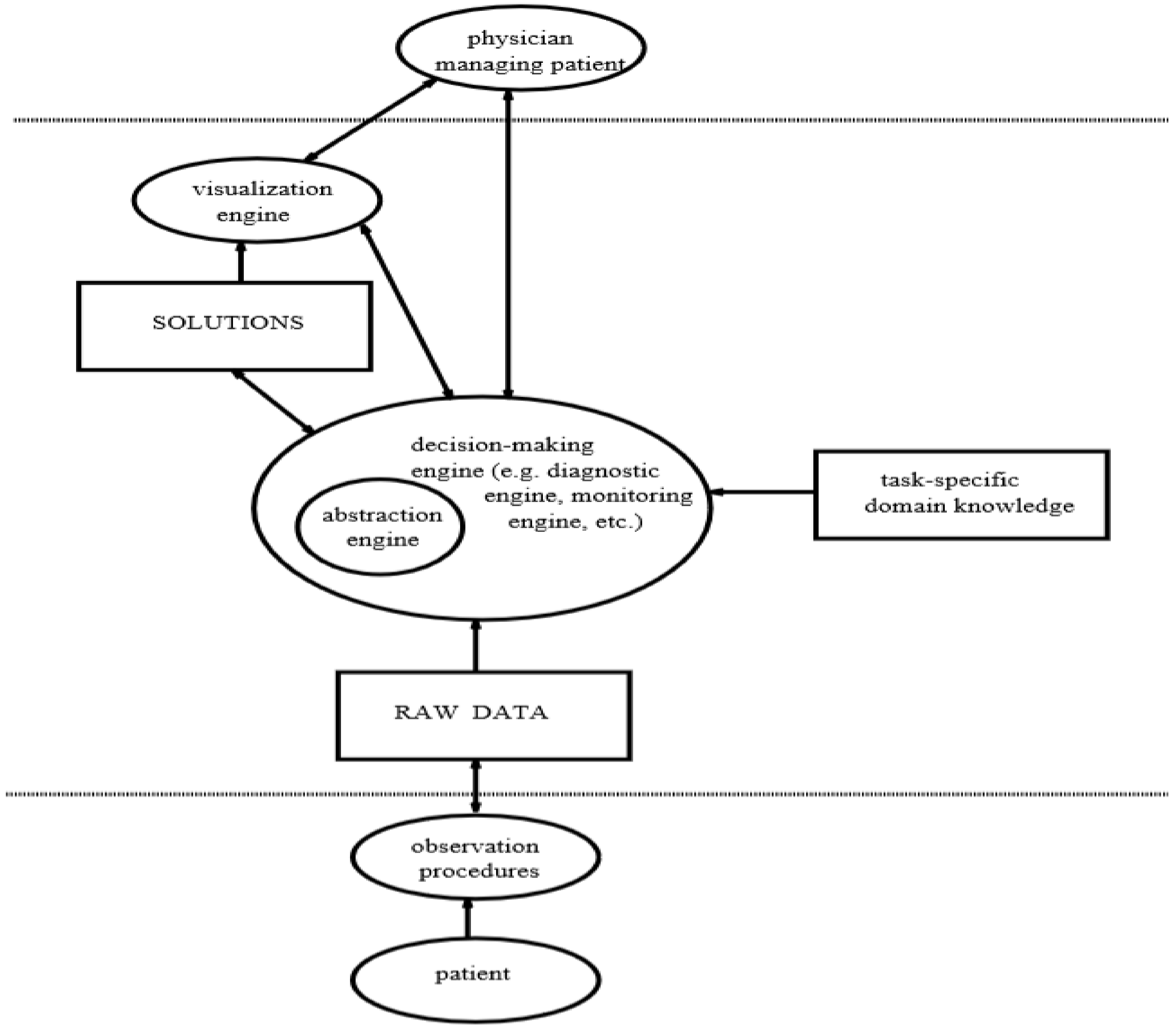
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Integration of data abstraction into a problem-solving system



Data abstraction as a loosely coupled process

Integration of data abstraction into a problem-solving system



Data abstraction as a task-dependent process

Data Mining through symbolic classification methods

- **Rule Induction** – given a set of classified examples, a rule induction system constructs a set of rules: IF Conditions THEN Conclusion

Example rule induced by CN2 in the domain of early diagnosis of rheumatic diseases

IF Sex = male

AND Age > 46

AND Number_of_painful-joints > 3

AND Skin_manifestations = psoriasis

THEN Diagnosis = Crystal_induced_synovitis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Data Mining through symbolic classification methods

- **Rough Sets** – If-then rules can also be induced by using the theory of rough sets which are concerned with the **analysis of classificatory properties of data** aimed at approximations of concepts.
- The basic concept is an **indiscernibility relation**: two objects x and y are indiscernible based on the available attribute subset B if they have the same values of attributes B; the set of objects indiscernible from x using attributes B forms an **equivalence class**.
- A main task is to find minimal subsets of attributes that preserve the indiscernibility relation; this is called **reduct** computation.
- **Decision rules are generated from reducts** by reading off the values of the attributes in each reduct.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Data Mining through symbolic classification methods

- **Association Rules** – given a set of transactions, where each transaction is a set of items (i.e., literals of the form *Attribute* = *value*), an association rule is an expression of the form $X \rightarrow Y$, where X and Y are sets of items; the intuitive meaning of such a rule is that transactions in a database that contain X tend to contain Y.
- *Example: "80%" of patients with pneumonia also have high fever. 10% of all transactions contain both these items; 80% is the confidence of the rule and 10% its support.*
- **Confidence** of the rule is the ratio of the number of records having true values for all items in X and Y to the number of records having true values for all items in X.
- **Support** of the rule is the ratio of the number of records having true values for all items in X and Y to the number of all records in the database.
- Association rule learners use minimum support and minimum confidence constraints.

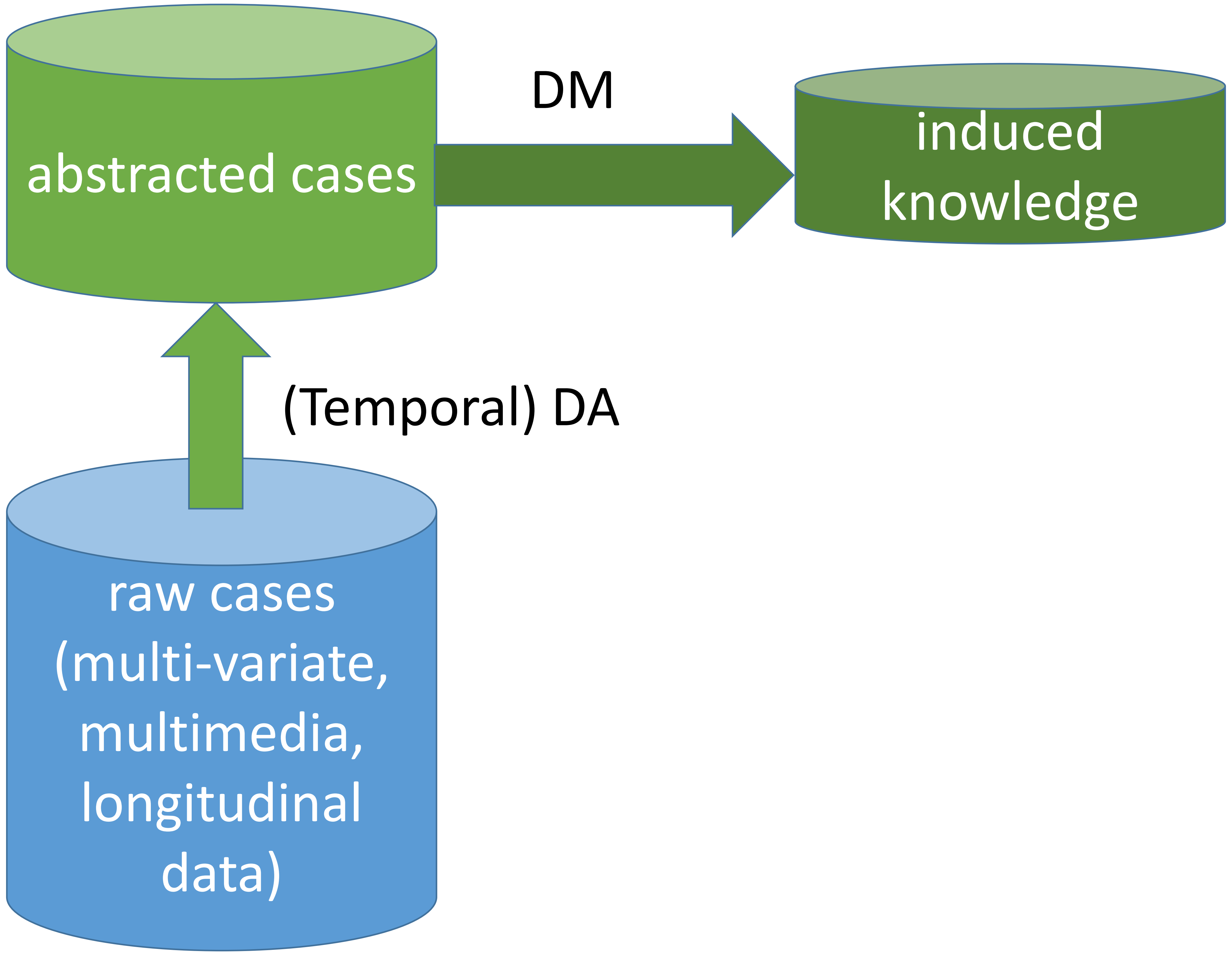
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Other symbolic classification methods

- **Learning of classification and regression trees** – Systems for top-down induction of decision trees generate a decision tree from a given set of attribute-value tuples; each of the interior nodes of the tree is labeled by an attribute, and branches that lead from the node are labeled by the values of the attribute. The tree construction process is heuristically guided by choosing the most informative attribute at each step, aimed at minimizing the expected number of tests needed for classification.
- **Inductive Logic Programming (ILP)** – ILP systems learn relational concept descriptions from relational data, in the form of Prolog clauses.
- **Discovery of concept hierarchies** – Decompose a classification dataset to equivalent but smaller, more manageable, and potentially easier to comprehend datasets. Function decomposition is such a method, which besides the discovery of appropriate datasets it arranges them into a concept hierarchy.
- **Constructive induction** – An ability of the system to derive and use new attributes in the process of learning.

Data abstraction for knowledge discovery



If the same complex abstraction, such as a nested periodic occurrence, is associated with a significant number of patients from a representative sample, it makes a strong candidate for being a significant piece of knowledge; sharing a complex abstraction is a strong similarity, whereas sharing a concrete datum is a weak similarity, if at all.

Summary

- The challenges of AIM – the change of focus from knowledge-intensive to data-intensive applications
- Influential Knowledge-Based Systems in Medicine
- Probabilistic models – Bayesian models, decision analysis, influence diagrams
- Ontology systems
- Intelligent data analysis in medicine

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423