Representation of Scientific Knowledge vis-à-vis Scientific Progress

Jayant Ganguly,

Regional Science Centre, Calicut

!

Controversial ideas, opinions, hypotheses, and theories that are often important to forming, evaluating, and modifying scientific explanations.

Ideally, our ability to represent semantics computationally should not be reduced to the lowest common denominator upon which we can all agree.

Disagreement may identify topics ripe for breakthrough.

- William Pike & Mark Gahegan

Outline

1. Scientific Knowledge (SK)
2. Knowledge Representation in General
3. SKR incorporating Situatedness - Codex
4. Representation of Probabilistic SK
5. Recent Trends – Abstraction & SK

1. Scientific Knowledge (SK)

On Scientific Knowledge - *crisis*

The scientific community's ability to generate new information – ever more detailed observations, about more diverse phenomena – often seems to outpace its ability to turn these measurements into useful knowledge.

What insight was discovered and then forgotten, or discovered but never communicated?

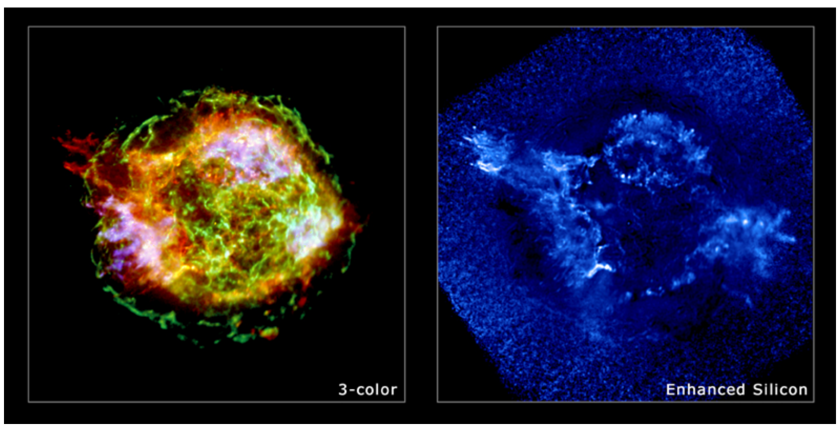
The problem is not that there is no wisdom contained in the digital artifacts of modern science, nor that contemporary science is at a standstill for its inability to make sense of increasingly complex descriptions of the world – quite the opposite, and ***that* is the problem**.

How do we make efficient and effective use of that knowledge?

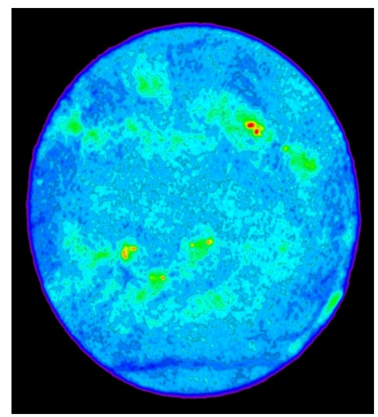
On Scientific Knowledge… *Diverse View points*

Even within a single discipline, say *Astrophysics* , the variety of information types and analytical methods brought to bear on a problem can complicate assessing commensurability between researchers' approaches.

The clearest picture of a problem might only be painted when diverse points of view are integrated into an explanation broader than any one alone could provide.



Cassiopeia A



Sun

On Scientific Knowledge…

A multi-wavelength picture of Sun , for instance, says something about the image at a particular wavelength that it depicts, although what it says to an individual researcher is either locked in the data, locked in the researcher's head, or described elsewhere in natural language text.

In any case, it is not easily accessible to others who want to know how or why to use this information (say, to devise a new theory), or whether it went into any existing theories.

For domains where meaning depends, in part, on the **subjective perspectives of its inquirers**, a restricted view of what constitutes a concept does not do justice to the complexity of human knowledge structures.

On Scientific Knowledge…

The information science literature is rife with efforts to represent human “concepts” computationally, but the prevailing view of a concept in much research is as a **category label** useful for integrating *heterogeneous data sources*.

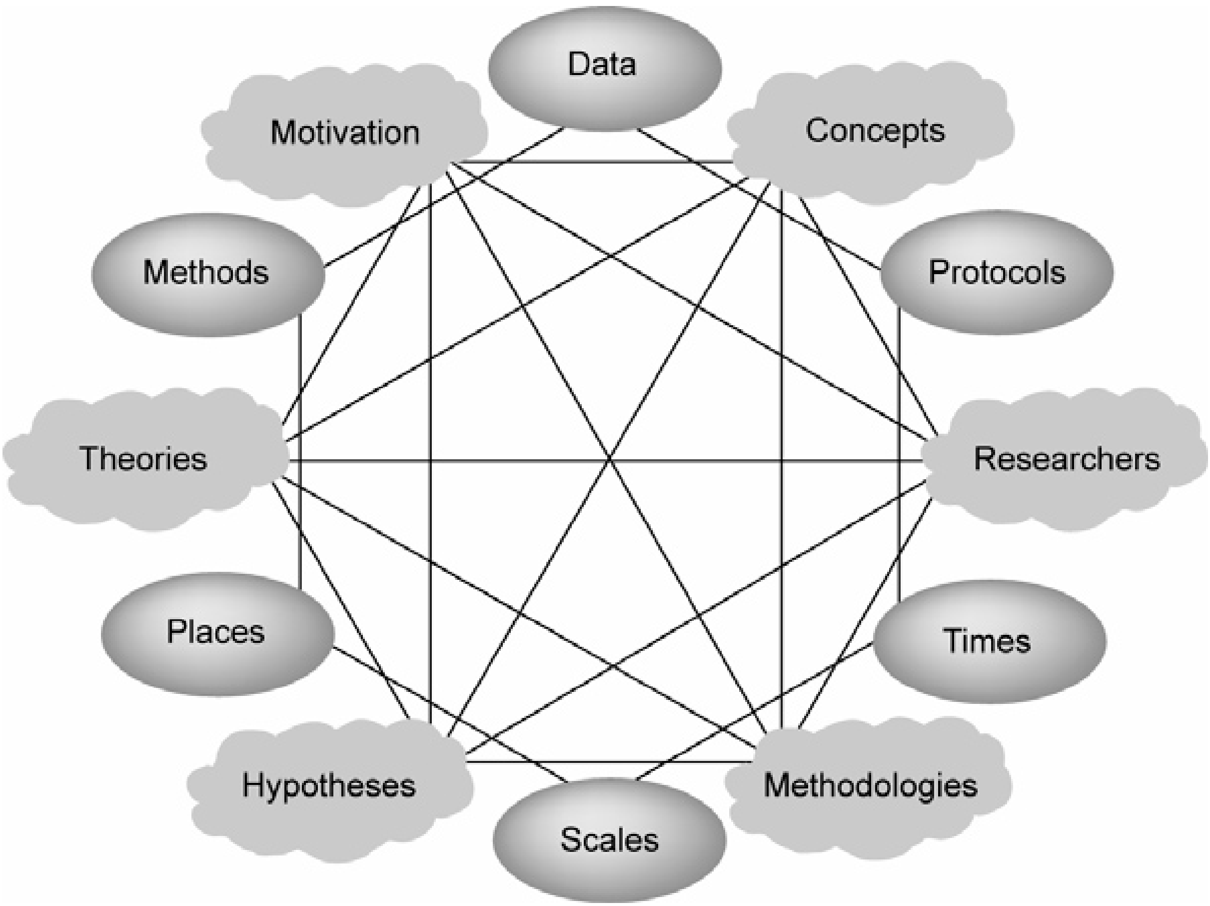
Computational data *contains* knowledge, to be sure, and it is used to create and apply knowledge, but that knowledge is not yet represented well.

As a result,

information integration tasks are often data-centric;
semantics are important to the extent that they support data interoperability,

But the human knowledge and practices that guided the collection or use of that data remain implicit somewhere in the data’s syntax or schema.

On Scientific Knowledge…



[1]

Nexus of constructs concerning the development and application of scientific knowledge. Some (ellipses) are often made explicit in scientific reports or metadata, while others (clouds) are not; the latter, however, are crucial to understanding, communicating, and reusing scientific knowledge.

On Scientific Knowledge… *Science as process*

Presently people think of capturing, storing, and communicating scientific knowledge by treating science foremost as a process.

Knowledge is constructed and applied during this process as observations are collected and manipulated, hypotheses generated and tested, and results transmitted and built upon.

Here, concepts rather than datasets are the primitive elements of scientific inquiry.

On Scientific Knowledge… Recent thoughts

This approach emphasizes interoperability of ideas, not simply data;
it recognizes that the knowledge these ideas embody is by turns a shared and contested conceptualization, the result of collaboration, negotiation, and manipulation by teams of researchers.

Whereas modern ontology is very much concerned with Aristotelian classification (a logic of *terms*), the new trend is moving toward knowledge representations as logics of *inquiry* and *interpretation*.

By devising a system for capturing individual perspectives on a problem,
concepts can be represented as ***cooperatively constructed***, ***experientially grounded***, and ***semantically interoperable resources capable of reflecting their evolution***, ***in multiple contexts, over time.***

Ultimately, the scientific record can be made more useful to collaborators across space and over time, as the audit trail that is captured can result in more robust explanations.

Classical views

Scientific Knowledge – Key features

- Consists of facts, procedures, judgment rules, highly disseminated
- Helps us to solve problems in particular scientific domain
- Helps us to make predictions
- Helps us to analyse, reason out

Scientific Knowledge – Types

- **Perceptual Knowledge** – *physical properties*
- **Concepts and Relationships** – *scientific laws*
- **Strategic Knowledge** – *how to set about a problem*

Classical views

Scientific Knowledge – What to represent?

- **Objects** – Aspects about things of interest
- **Events** – actions that occurs
- **Performance** – how to do things / its behaviors
- **Meta-knowledge** – Knowledge about what we know
- **Facts** – World realities for representations

Scientific Reasoning

- **Formal Reasoning** ***{logic based and production rule based representation}***
  - Syntactic manipulation of data structures to deduce new ones
- **Procedural Reasoning*****{frame based and semantic network based systems}***
  - Involves specialised routines or procedures for answering questions and solving problems
- **Reasoning by analogy**
  - Extrapolation (or induction) of new facts from existing facts
- **Generalization and Abstraction**
  - General reasoning process for human beings
  - Difficult to formalize yet
- **Meta level Reasoning**
  - Knowledge about extent of one's knowledge in solving one's problem

2. Knowledge Representation in General

Properties of Good Representation

- Make the **important objects** and **relations** explicit – **can see what going on at a glance**
- Expose **natural constraints – can express the way one object or relation influence another**
- Brings **objects** and **relations together**
- Suppress **irrelevant details**
- Transparent
- Complete
- Concise
- Fast – store & retrieve information rapidly
- Computable

Knowledge Representation Categories

- **Logical** Representation scheme
  - Procedural & Predicate Calculus
- **Procedural** Representation scheme
  - Rule based Expert system {if…then…}
- **Network** Representation scheme
  - Semantic Networks { node – relationship arcs}
  - Conceptual graph
    - Finite , connected, bipartite; doesn’t use labeled arcs
    - Conceptual Dependency (et. al. Roger Schank) ***{ 11 primitive acts & 4 primitive categories}***
      - Two sentences that are identical in meaning, regardless of languages have one representation
      - Any information in a sentence that is implicit must be made explicit in representing the meaning of that sentence
- **Structured** Representation scheme
  - Frames ***(Marvin Minsky)***
      - Collection of attributes / slots and associated values describing real world values
  - Scripts ***(Schank and Abelson)*** ***{ Entry conditions, Results, Props, Roles, Scenes}***

On Knowledge Representation

There are two broad approaches to the problem of knowledge representation-

Top – Down & Bottom Up

The ontological approach is characterized by a top-down, authoritative encyclopedia. Ontological tools focus mainly on enabling sharable underlying representations of knowledge and less on interfaces and supporting infrastructure to let collaborators construct this knowledge together.

The alternative approach emphasizes the bottom-up, discursive nature of knowledge. This approach acknowledges the perspectives of collaborating inquirers (*rather than an imposed ontology*) in defining concepts relevant to a community.

The cooperative approach is evident in computer-mediated communication methods such as the Delphi method (Turoff and Hiltz, 1996), where the aim is to generate shared understanding (or areas of disagreement) over time.

Tool CSCW

The bottom-up, cooperative approach to knowledge construction is characterized by the tools and methods of **Computer-Supported Cooperative Work** (**CSCW**). CSCW applications for scientific collaboration often take the form of electronic notebooks, organized into hierarchies of chapters and pages (*e.g., Lysakowski and Doyle, 1998; Myers et al., 2001*), in which researchers can enter and search for free-form records (although these notebook are still linear in structure).

3. SKR incorporating Situatedness - Codex

Situatedness..

Magnani (2001) suggests that ***situatedness*** is precisely what makes abduction a useful model for computer-based hypothesis creation – even under conditions of hypothesis failure, it produces useful information.

Situated Representations of Scientific Knowledge

This approach to representing scientific concepts computationally that reflects

(1) the situated processes of science work,
(2) the social construction of knowledge, and
(3) the emergence and evolution of understanding over time.

Here is a model, knowledge is the result of collaboration, negotiation, and manipulation by teams of researchers.

Situated Representations of Scientific Knowledge..

Capturing the situations in which knowledge is created and used helps these collaborators discover areas of agreement and discord, while allowing individual inquirers to maintain different perspectives on the same information.

The capture of provenance information allows historical trails of reasoning to be reconstructed, revealing the process by which **knowledge is adopted, revised, and reused in a community**; as a result, **end users can evaluate the utility and trustworthiness of knowledge representations**.

Using this notion a proof-of-concept system, called **Codex**, based on this situated knowledge model. Codex supports visualization of knowledge structures through **concept mapping**, and enables inference across those structures.

Situated Representations of Scientific Knowledge..
On Ontologies – *incorporating Situatedness*

Ontologies, as they are typically implemented in information systems, are often **hierarchical** and **authoritative**: these ontologies are useful formalizations in circumstances where formalization is called for, such as mapping terms between domains.

But real-world cognition is often more ***fluid, flexible, and context-dependent*** than strict formalizations suit.

Can ontologies properly capture the nuance of human knowledge?

Does any single ontology reflect what is truly relevant to a particular application or domain, or is the ontology more a reflection of its creators' worldview than of neutral or common belief?

In this representation it is propose that knowledge representations for computational environments should reflect the ***situated*** nature of human understanding.

Situated Representations of Scientific Knowledge..

Knowledge Ingredients

Three components that are required to represent knowledge in a more contextualized fashion:
concepts, metadata, and situations.

The *concept* expresses the existence of an abstract category and encompasses everything in its extension. A given concept may have different names in different circumstances while preserving the same underlying meaning (its intension)…..e.g. **entropy**

Situated Representations of Scientific Knowledge..

Concept Representation

In this model representation of concepts is a dimensional variety of probabilistic model. Here concepts are defined through the values (or range of values, as in *Gardenfors (2000))* they occupy along continuous dimensions (*Smith and Medin, 1981*).

A further characteristic of the view of concepts taken here follows from the notion of **perceptual-functional affordances** (*Tversky, 2005*) (***the roles a concept plays or the capabilities it enables***) initially developed to account for visual and spatial properties of an entity.

In a dimensional approach to represent concepts, these concepts come to occupy a multidimensional “concept space” within which we might look for some of the same functional affordances.

Representing concepts’ functional roles in a larger knowledge structure is important to depicting “how” and “why” in scientific reasoning.

Situated Representations of Scientific Knowledge..

Situations

In this model, **each concept** is wrapped in metadata that consists of the attributes that can be recorded regardless of how or why a concept is used:

who created it, using what tools, at what time and place, and so on.

In **the process of inquiry**, concepts are selected based on relevant criteria and linked together into larger structures.

These acts of conceptual manipulation have been described as ***situation*** (*Solomon et al., 1999*), the bringing together of background information and current observations and analyses toward some goal.

Situation is important to knowledge representation because it explicitly reproduces the enactment that is part of selecting and reasoning with a set of concepts (*Barsalou, 2002*).

Lemke (1997) calls situation an "ecology," a term that evokes the dynamic interaction between concepts and thinkers in the process of knowledge construction.

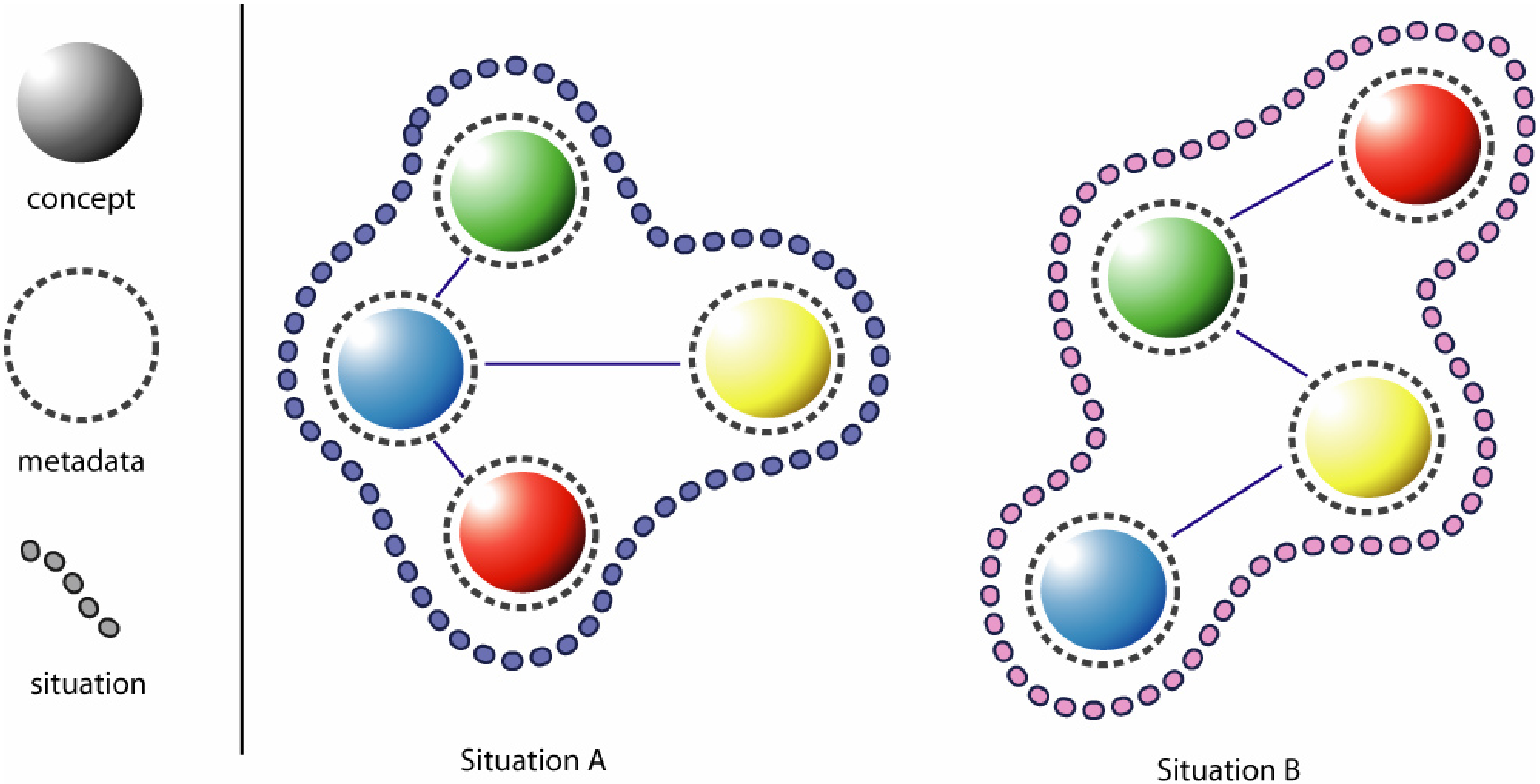
Situated Representations of Scientific Knowledge..

Situations cont..

Situation, then, encompasses the coordinated activity that is directed toward some goal.

A given concept can be reused in different circumstances, but there will be some information we want it to carry with it regardless of circumstance (***called metadata***), and some that will be unique to the role it plays in a particular case (***called situation***).

To denote the particular choice of concepts, metadata, and situations that a particular thinker (or community of thinkers) uses to describe a process, problem, or phenomenon, we can use the term ***perspective***.

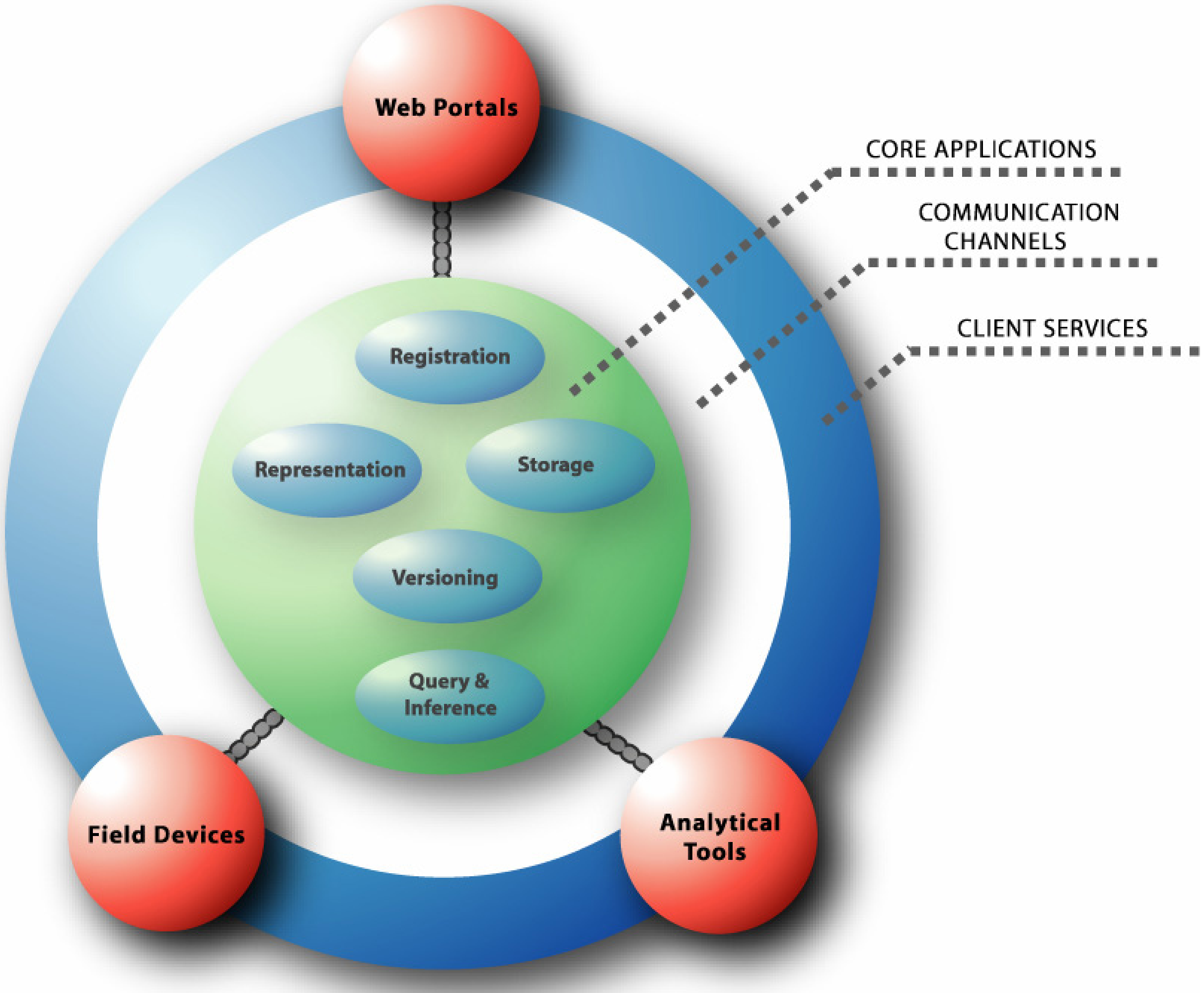


[1]

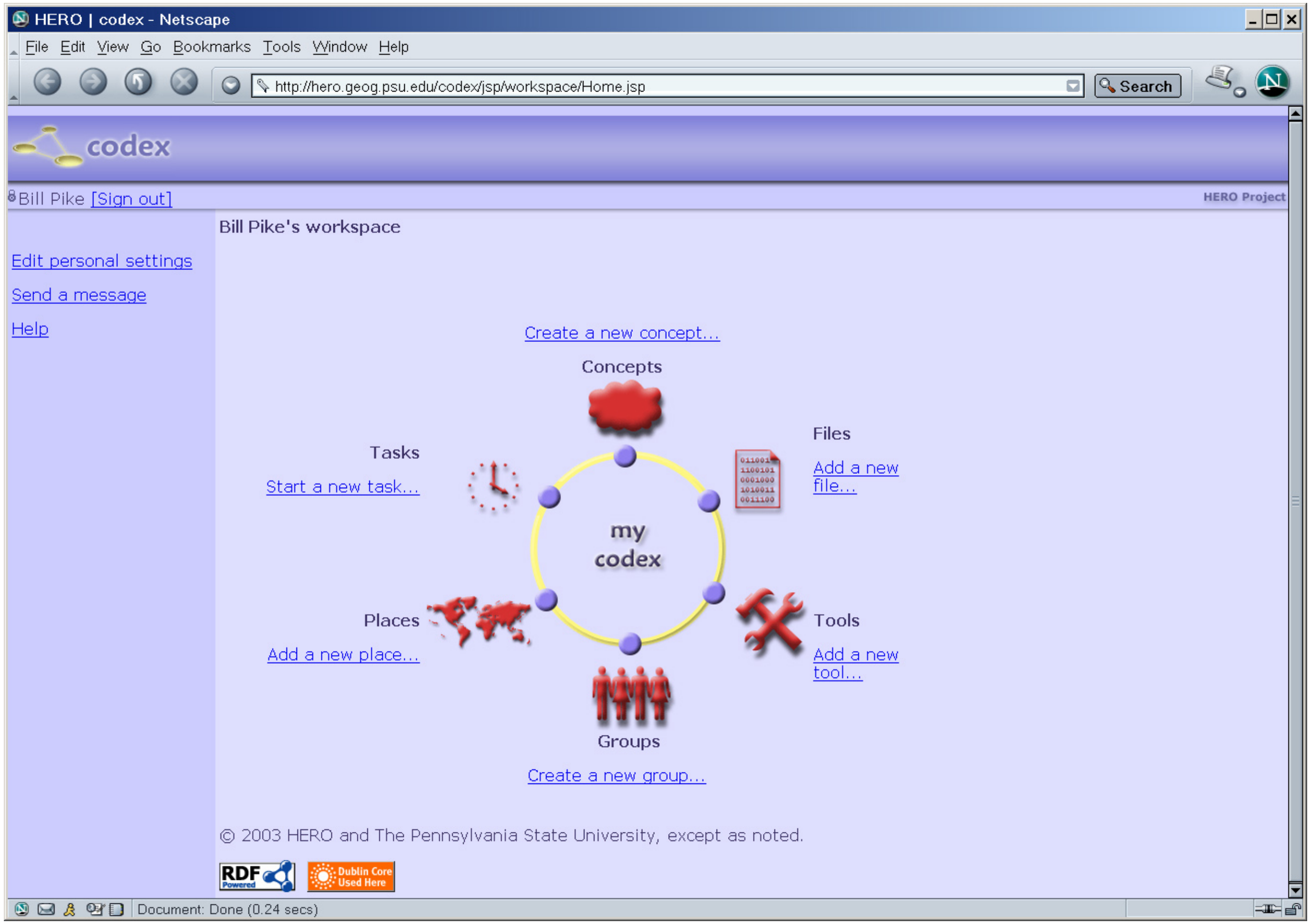
Metadata in this model describe the circumstances surrounding the creation or use of an individual resource or concept (a node in a conceptual network); situations describe the circumstances of larger knowledge structures arising from the different ways these nodes can be connected.

Implementing a Situated Knowledge Model

Codex Architecture



Codex – Home page



Codex Characteristics

- while Codex allows data files to be stored and linked together, data are described foremost by the **human concepts they signify.**
- Codex also builds on online scientific workbenches (e.g., Stevens *et al.*, 2003) that emphasize **data integration for automated analysis**;
- Codex treats problem-solving as an issue of human consideration and interpretation.
- Codex is at once a CSCW tool that enables rich semantic descriptions, and a semantic markup platform that relaxes the constraints of common ontological approaches.

Codex Resource Categories

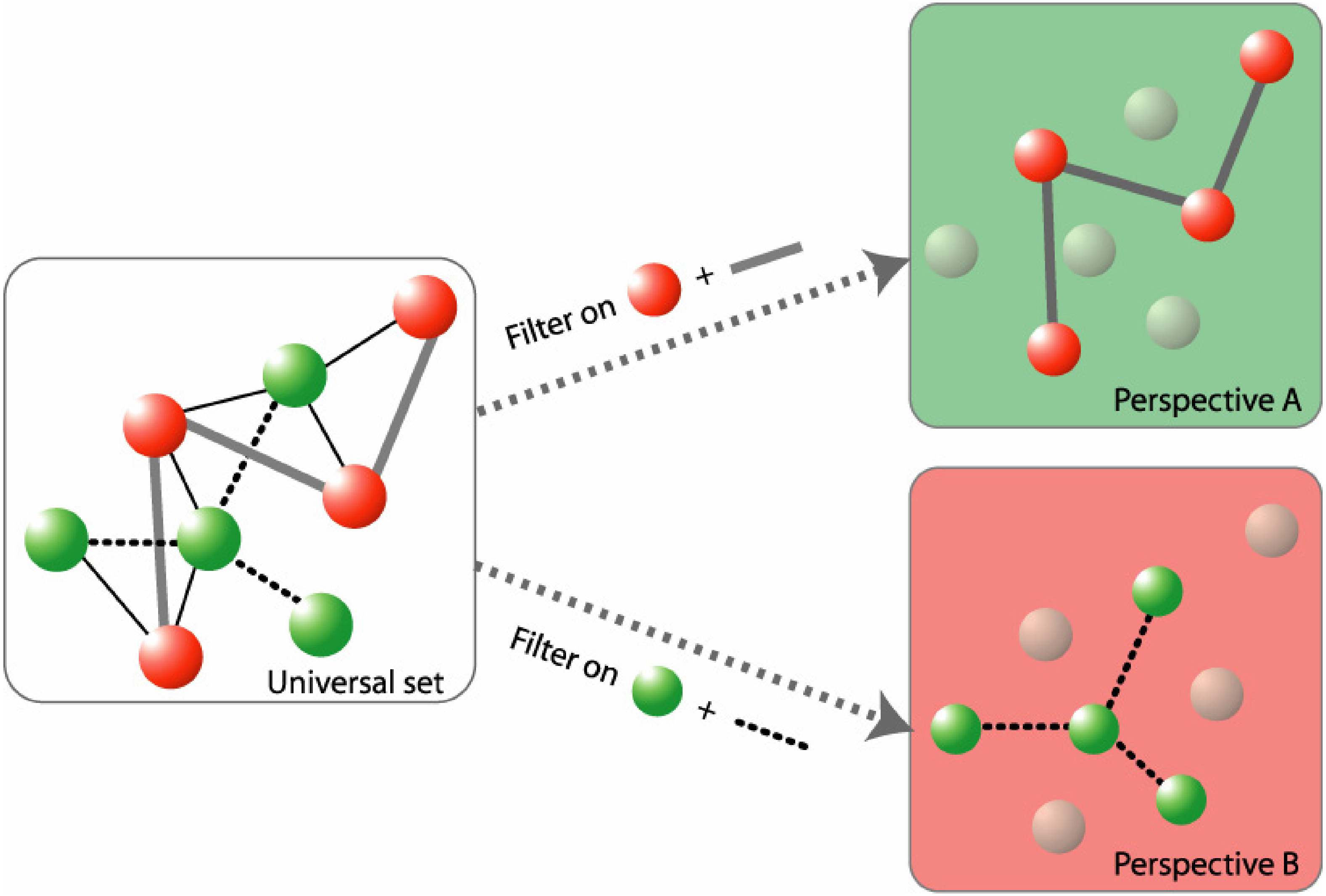
- **People.** The individuals and groups who create or apply resources accessed through the Portal. Each person maintains a profile that can communicate elements of his or her background and expertise.
- **Concepts.** Descriptions of abstract ideas, such as “Tsunami” or “Supernova”.
- **Files.** Binary data that express something about a concept. Files could include *spreadsheets, text documents, images, audio clips, maps, or other data* formats (quantitative or qualitative) that connect observations or measurements to the cognitive structures represented by concepts.
- **Tools.** The methods used to analyze data and to construct instantiations of concepts (categories) from data. Tools could include GIS operations, visualization methods, predictive models, interviewing instruments, or statistical tests.
- **Places.** Geography is fundamental to integrative research, and places help researchers define the locations and scales under study, whether described as bounding polygons or as place names. Place also helps to account for differences in epistemology between researchers.
- **Tasks.** People, concepts, files, tools, and places are linked together through tasks that might describe a workflow process, an experimental procedure, or a problem-solving approach

Modeling Knowledge in Codex

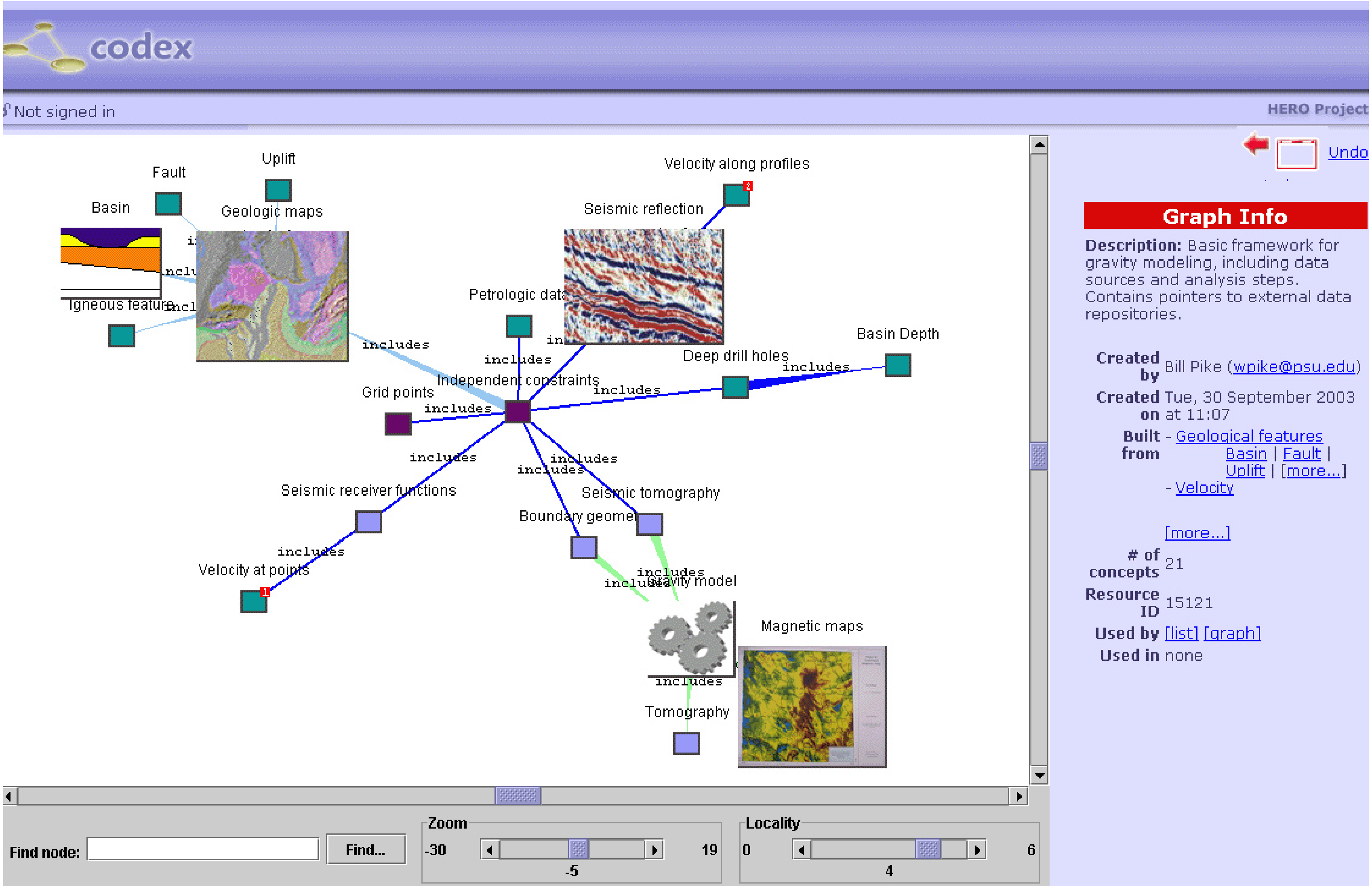
The Concept (capital C) is the **universal set** in Codex; every resource and set of resources that can be described using Codex is either a member of the class of Concepts or a member of a proper subset.

The use of Concept as a universal quantifier also places Codex's knowledge model in explicit opposition to contemporary style.

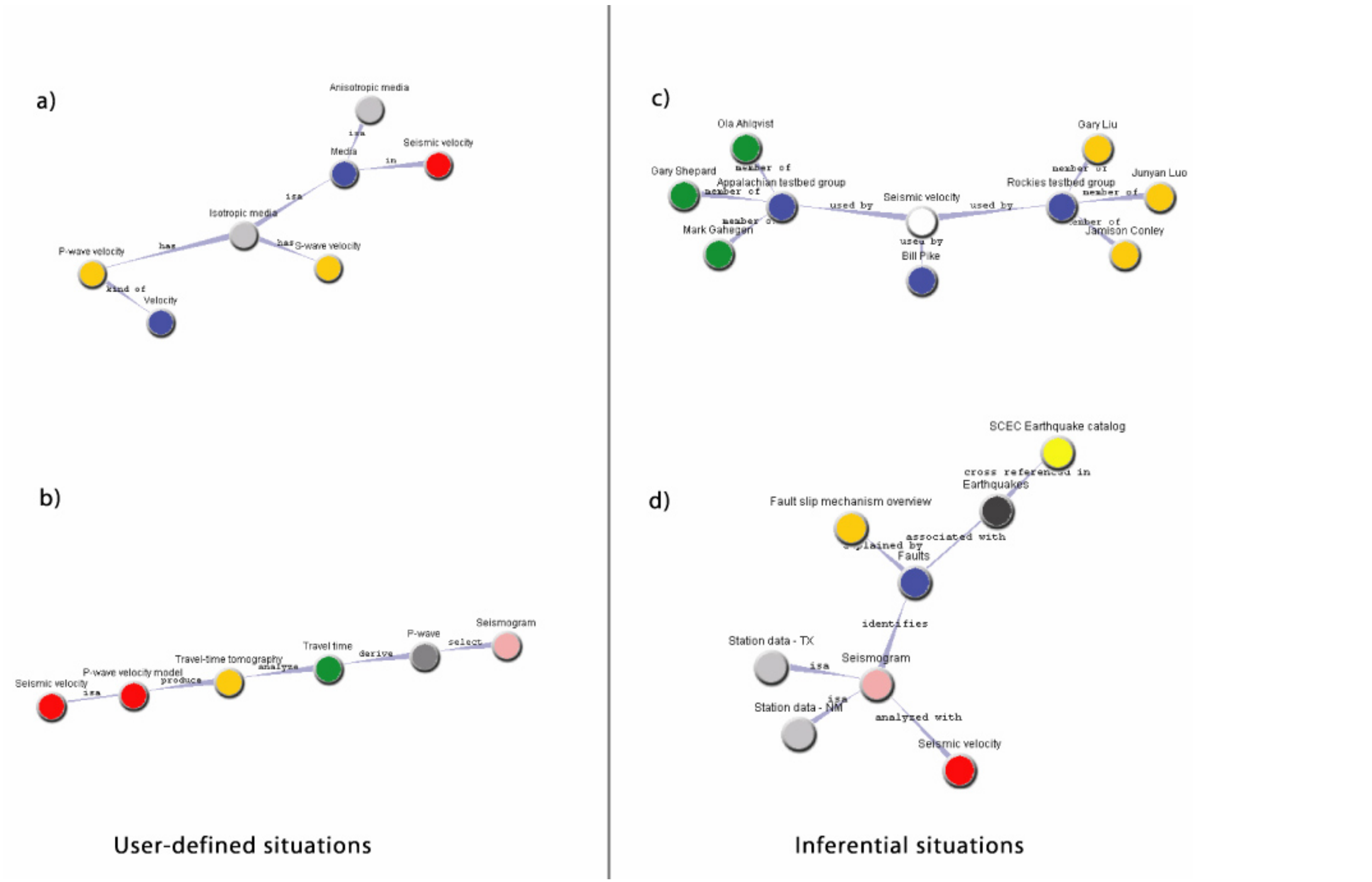
In Codex: a Concept *C* is the set of properties {*P1…Pn*} that characterize it. Each *P* is another Concept typecast.



Perspectives ***filter a complex information space according to particular situations. Perspectives A and B preferentially select different types of resources and relations from the universal set of all Codex resources***

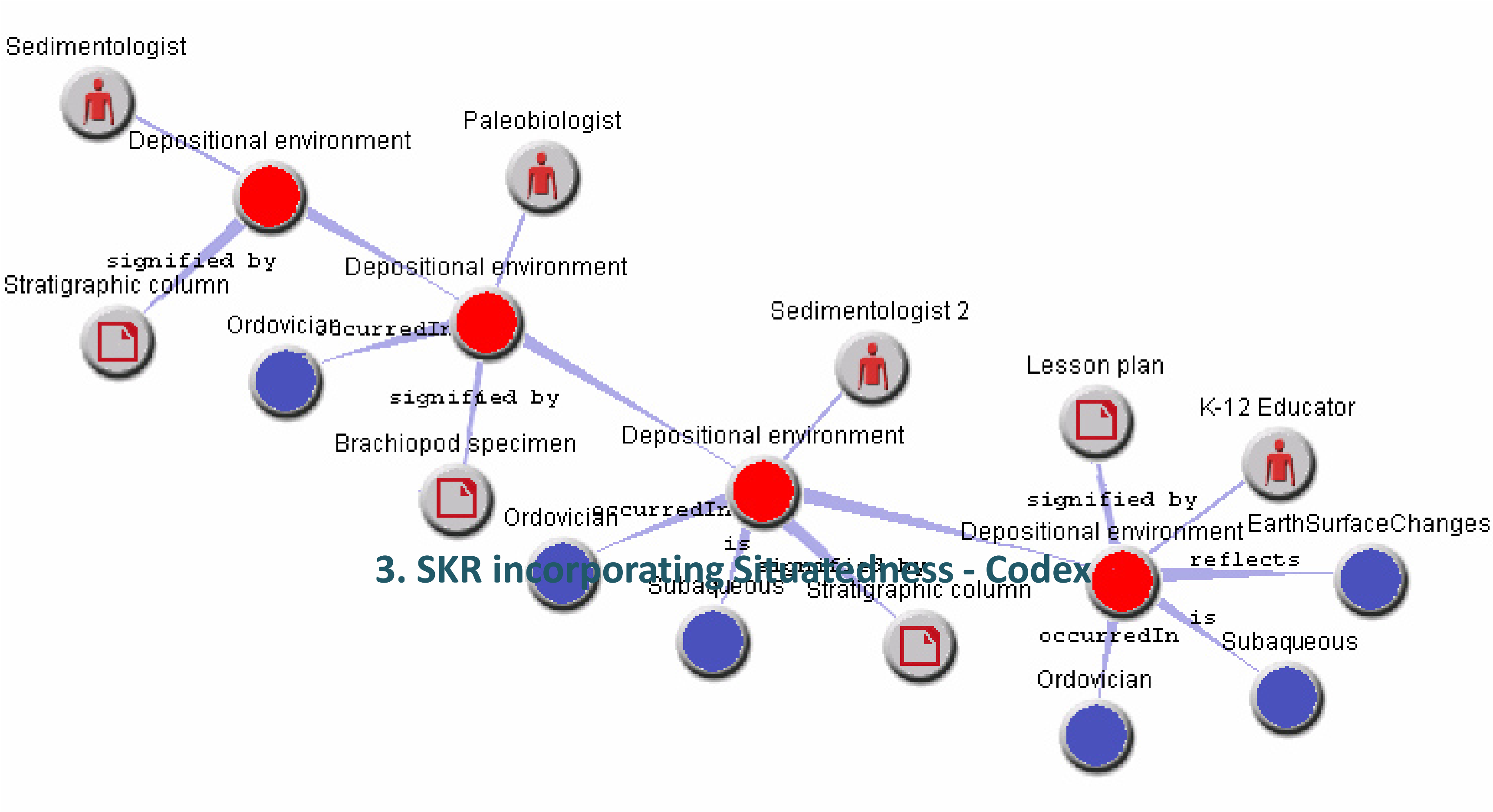


Codex concept map client



Four perspectives on a "Seismic velocity" concept (red node). a) Intensional concept structure. b) A task that describes how seismic velocity can be measured. c) A social network built around users of the concept. d) Data resources that have been used to describe seismic velocity

3. SKR incorporating Situatedness - Codex



Evolution of "Depositional environment" concept through use by researchers in different communities of practice, progressing from upper left to lower right.

4. Representation of Probabilistic Scientific Knowledge

Representation of Probabilistic Scientific Knowledge

Scientific knowledge is inherently uncertain: experimental observations may be corrupted by noise, and no matter how many times a theory has been tested there is still the possibility that new experimental observations will refute it — as famously happened to Newtonian mechanics.

Probability theory has from its conception been utilized to represent this uncertainty in scientific knowledge.

However the role of probability theory has proved controversial, with for example the great philosopher of science Karl Popper arguing that probabilities cannot be applied to scientific theories on the grounds that an infinite number of theories can explain any scientific data, therefore their a priori probabilities are zero.

Representation of Probabilistic Scientific Knowledge

Bayesian Approach

Presently, a Bayesian approach to the use of probabilities in science is widely accepted.

In Bayesian reasoning a priori probability estimates for hypotheses are updated through observation of additional evidence.

The Bayesian approach is arguably the only rational method for updating beliefs .

The conventional knowledge representations in bio-medicine are insufficient to support probabilistic reasoning. The best available representation, in our view, is the **Evidence Code Ontology (ECO)**

ECO enables the recording of evidence that supports scientific statements, e.g. **experimental evidence, sequence similarity, curator inference;** and also by what method the evidence was obtained, e.g. through computational combinatorial analysis, inference from background knowledge.

This information enables researchers to qualitatively evaluate the degree of uncertainty of scientific statements.

HELO (HypothEsis and Law Ontology)

- The HELO ontology was originally designed to support development of Robot Scientists, these are physically implemented laboratory automation systems that exploit techniques from the field of artificial intelligence to execute cycles of scientific experimentation.
- A probability that a research statement is true may vary greatly depending on the source of the statement.
- HELO aims to provide a framework for the recording of probabilities that research statements are true, and for probabilistic reasoning with such statements.

HELO Classes

- The HELO representation of research statements is based on the representation of research hypothesis as PREDICATE(entity i, entity j) defined in an ontology LABORS, where predicate is a relation and entity is a class or instance defined in a domain ontology.
- HELO enables one to formulate complex research statements, where basic (atomic) statements like PREDICATE(entity i, entity j) are combined by logical operators
- Entities that form research statements may be replaced by more generic entities (parent classes) and/or be specialized by their properties
- Specific environmental factors could be replaced with general terms

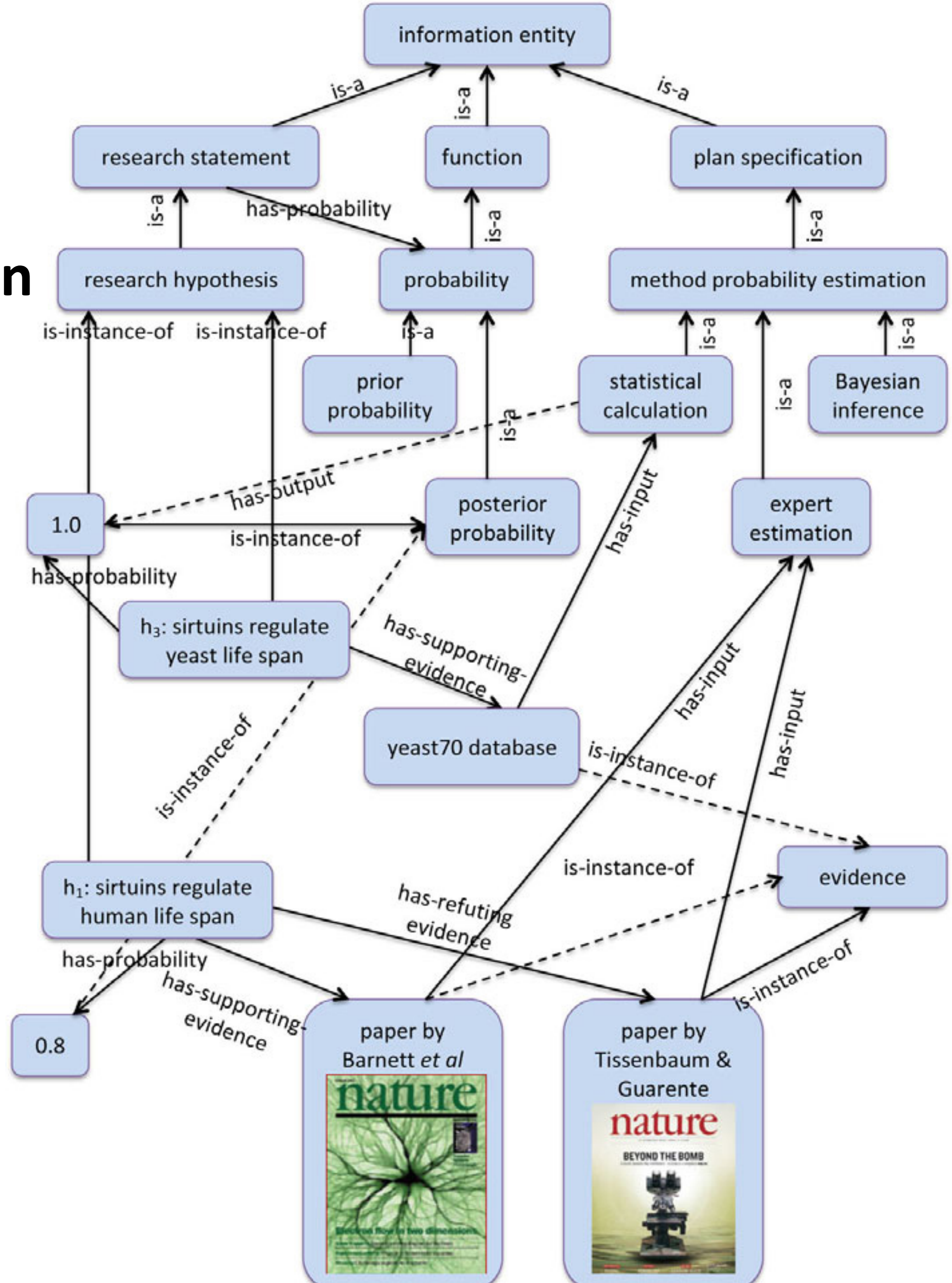
HELO cont..

HELO defines a hierarchy of **research stat**ements: research hypothesis, hypotheses set (a collection of hypotheses with a total probability 1, it usually combines research hypotheses, negative hypotheses, and alternative hypotheses, assumption, conclusion, scientific law (models and generic rules, including Bayes rule), theorem (including Bayes theorem).

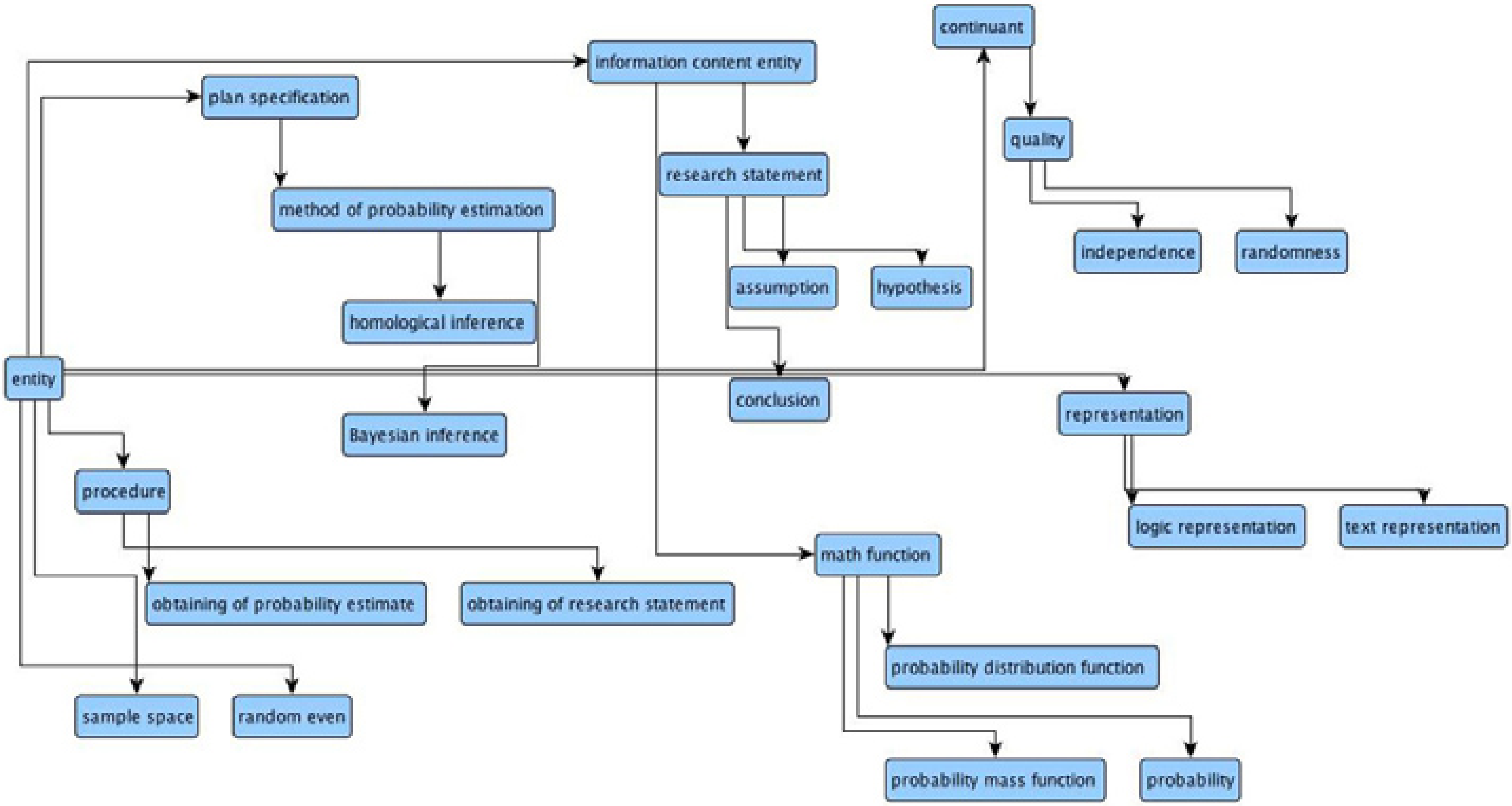
Research laws may be represented as **production rules** (statement i, statement j), where statements correspond to hypotheses, evidence, conclusions.

HELO is designed to consistently accommodate scientific hypotheses and laws collected from different sources: interviews with scientists, web pages, research papers, databases, program codes. Any research statement in HELO has an associated probability of being true

HELO Representation



HELO Ontology



5. Recent Trends – Abstraction & SK

Abstractions and scientific knowledge representation - Valentin Bazhanov, EPISTEMOLOGIA, 2013

Abstractions play a crucial role in scientific knowledge representation.

The author analyzes the nature and mechanisms of functioning of some scientific abstractions in the scientific knowledge representations as well as limitations that they placed upon the result of scientific knowledge acquisition. Abstraction is the process (and result) of limitation of certain kinds of differences.

The crucial problem of the emergence of abstraction is the problem of eliminating extraneous premises.

Abstraction enables us to overcome the '**entropy of experience**' and to represent knowledge as ***ordered sets of judgments*** which possess particular sense within the '**interval of abstraction**'.

'Interval of abstraction' may be interpreted as the **measure of informative capacity of abstraction** with its model of applicability, and not only the degree of its distraction properties.

References

[1] William Pike and Mark Gahegan **; Beyond Ontologies: Toward Situated Representations of Scientific Knowledge; 2012;**

[2] Larisa N. Soldatova*, Andrey Rzhetsky, Kurt De Grave, Ross D King ; **Representation of Probabilistic scientific knowledge;** Bio-Ontologies 2012

[3] Valentin Bazhanov; **Abstractions and scientific knowledge representation** *(Astrazioni erappresentazione della conoscenza scientifica);*EPISTEMOLOGIA

[4] Bowrow, D. G. and A . Collins (Eds.), **Representation & Understanding**, Academic Press, NY, 1975

[5] Patterson, D. W., **Introduction to AI & Expert system**, PHI, 1998

Thank you

Best wishes to CLC on it’s Silver Jubilee

A Restaurant Script

Script: RESTAURANT			
Track:	Coffee Shop	Entry cond.:	S hungry
Props:	Tables		S has money
	Menu		
	F=Food	Results:	S has less money
	Check		O has more money
	Money		S is not hungry
			S is pleased (optional)
Roles	S=Customer		
	W=Waiter		
	C=Cook		
	M=Cashier		
	O=Owner		

Frames

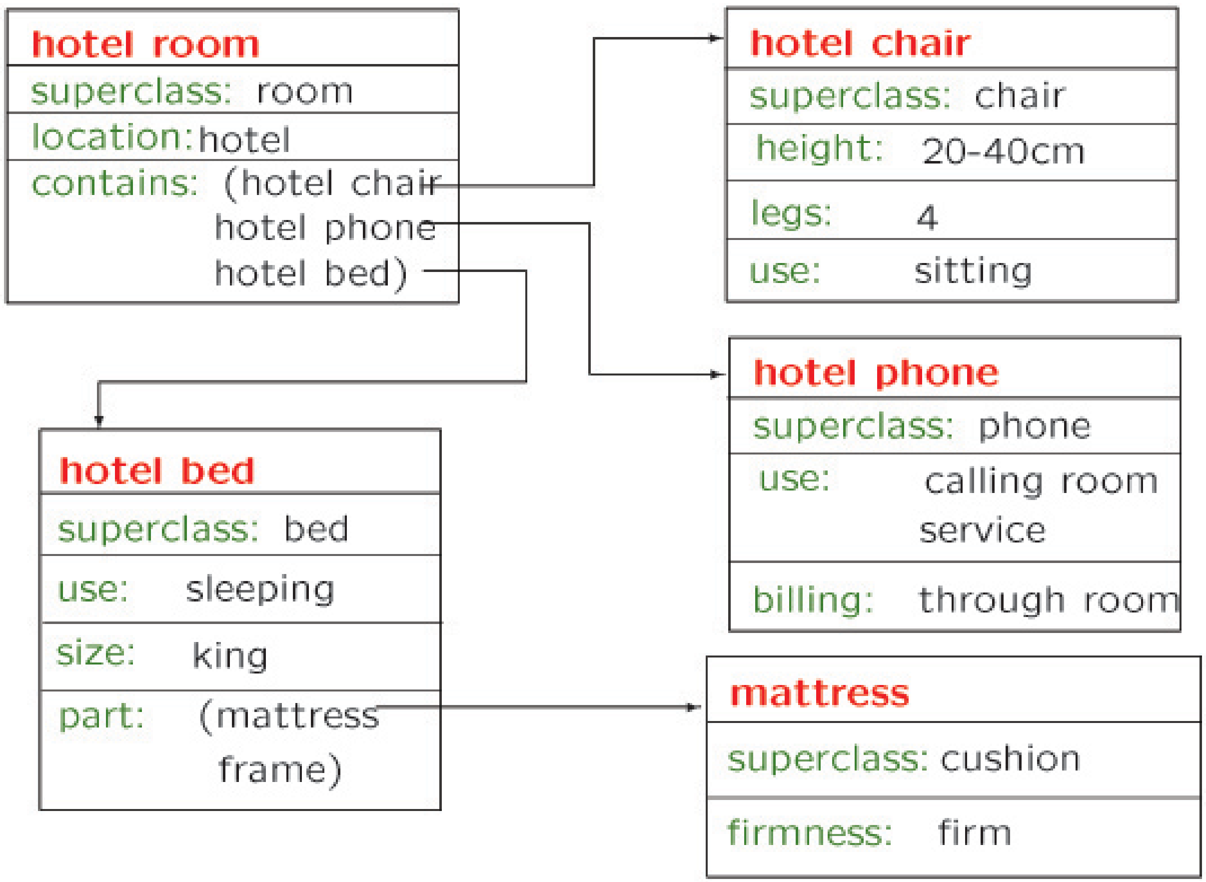
Frames: A frame consists of a collection of **slots** which can be **filled** by values or pointers to other frames.

Meaning of "child's birthday party" poorly approximated by definition like "a party assembled to celebrate a birthday" with "party" defined as "people assembled for a celebration".

Children know more plus default assignments:

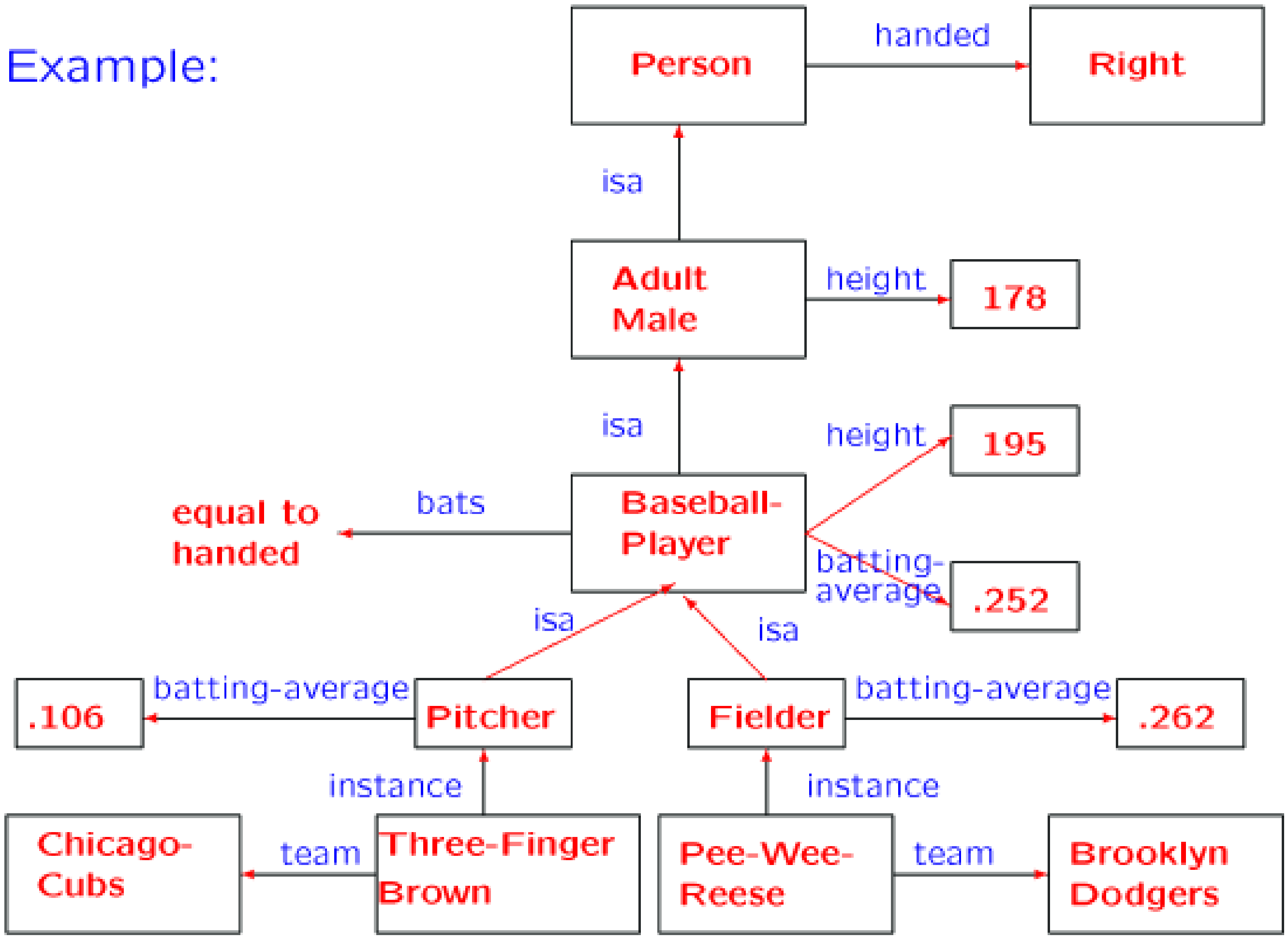
Child's Birthday Party	
Dress:	Sunday-Best
Present:	Must please host. Must be bought and gift-wrapped.
Games:	Hide and seek. Pin tail on donkey.
Decor:	Balloons. Favours. Crepe-paper
Party-meal:	Cake. Ice-cream. Soda. Hotdogs
Cake:	Candles. Blow-out. Wish. Sing Birthday Song.
Ice-cream:	Standard three-flavour

Part of the Frame Description of a Hotel Room



Knowledge Representation in Semantic Nets

Example:



Procedural Attachment

rectangle	
superclass:	polygon
(x, y)-Position:	(0cm,0cm)
length:	5cm
breadth:	2cm
area:	procedure(z) $\text{length}(z) \cdot \text{breadth}(z)$
circumference:	procedure(z) $2 \cdot (\text{length}(z) + \text{breadth}(z))$

square	
superclass:	rectangle
(x, y)-Position:	(0cm,2cm)
length:	5cm
breadth:	procedure(z) $\text{length}(z)$

Instead of writing explicit values, the values of the slots area, circumference (as well as breadth in the case of **square**) are calculated by need. Consider update of length from 5cm to 6cm!

Discourse Representation Theory (DRT)
- Kamp

- Representation of natural language discourses
- Consists of two parts
  - Set of discourse markers, used to represent objects in the discourse
  - Conditions on this objects

Database & Knowledge Base

A database of facts, for example, is sometimes called a “knowledge base.”

But does this accumulation of facts reflect understanding (that is, are the experiences and reasoning facilities capable of making use of this knowledge present)?

Or are the facts meant solely to facilitate the recollection or creation of knowledge by their user?

Only in the former case could this computational information, as it is stored, properly be called knowledge.

Recently efforts are made to incorporate aspects of understanding into the representational medium itself, e.g.. *Situatedness*. Thus, the representation explicitly preserves the sense of utility that creates knowledge out of information.