

# Modeling and Reuse of Business Critical Knowledge

*11 December 1998*



***Tarragon Consulting Corporation***

*"We help you ask the questions and get the answers."*

1563 Solano Avenue, Berkeley, CA 94707

tel: 510.526.7991 fax: 510.524.5622 info@tcc.com www.tcc.com

# Knowledge Management

---

- ◆ KM is fundamentally about the *organization and interpretation* of unstructured, business critical information
- ◆ KM improves the efficiency and effectiveness of business processes that generate and use large amounts of unstructured information
- ◆ The market for KM software products and services is estimated at \$1.8B today growing to \$6.6B by 2002 (Ovum Inc.)

# Tarragon Products and Services

---

- ◆ Tarragon provides KM tools and services that target specific functions/processes in selected industry segments
- ◆ Tarragon helps organizations to develop web-based “Knowledge Portals” that provide integrated, task-oriented access to the corporate information space
- ◆ Tarragon’s portal architecture consists of a series of configurable knowledge application modules that can be integrated with existing information retrieval, document management, and database systems

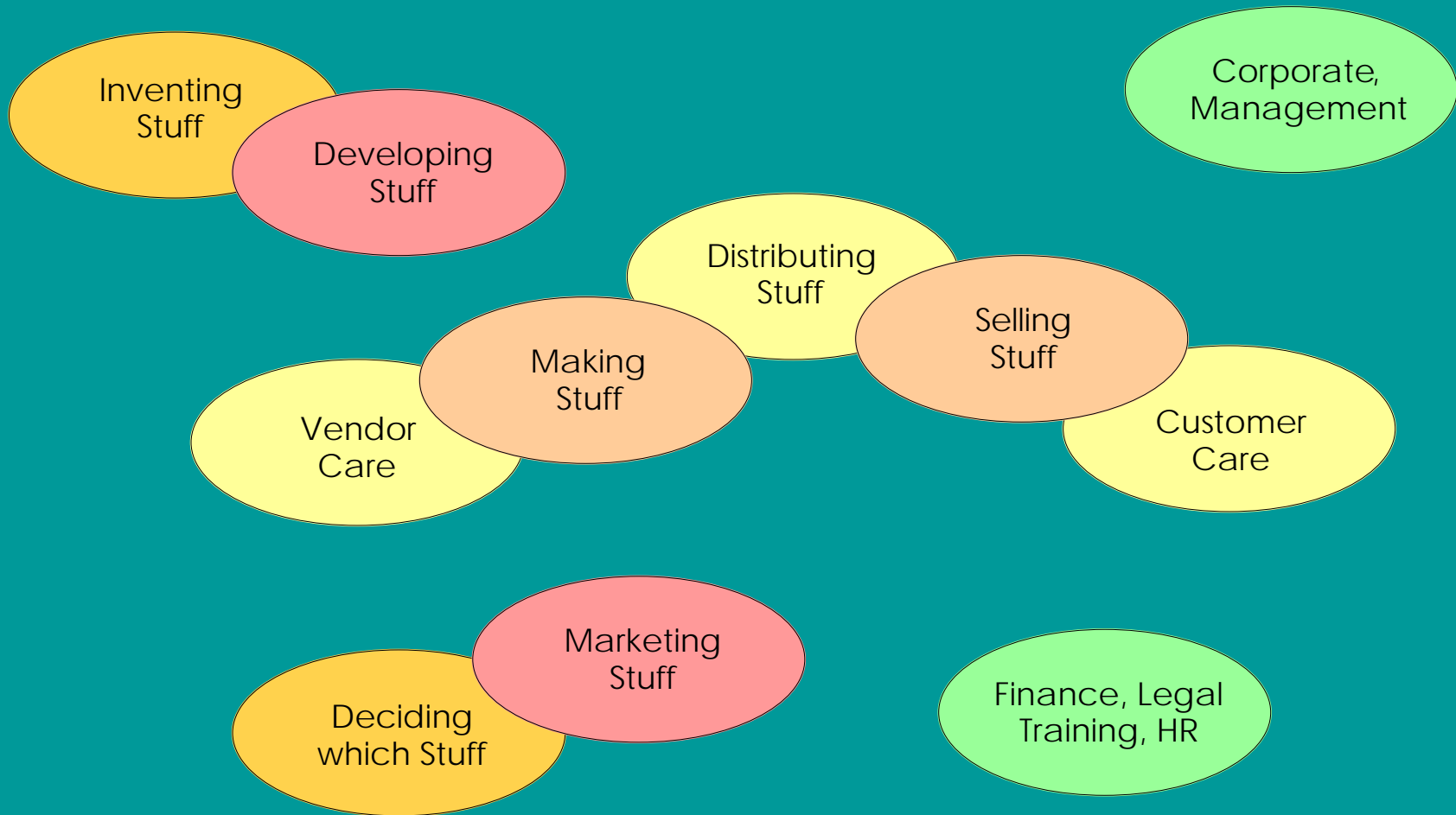
# Selected Clients

---

- ◆ Bristol-Myers Squibb
- ◆ Salomon Smith Barney
- ◆ Information Access
- ◆ Inktomi
- ◆ US Department of Defense

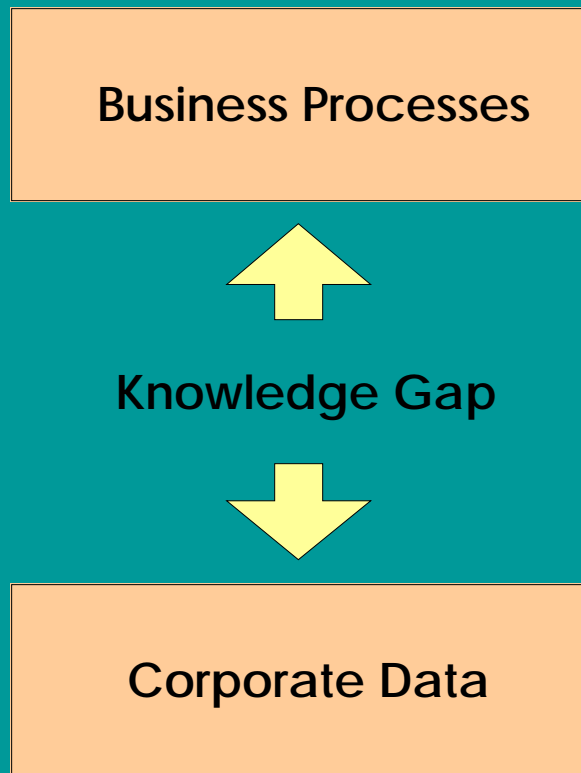
# Business Process Chains

---



# The Knowledge Gap

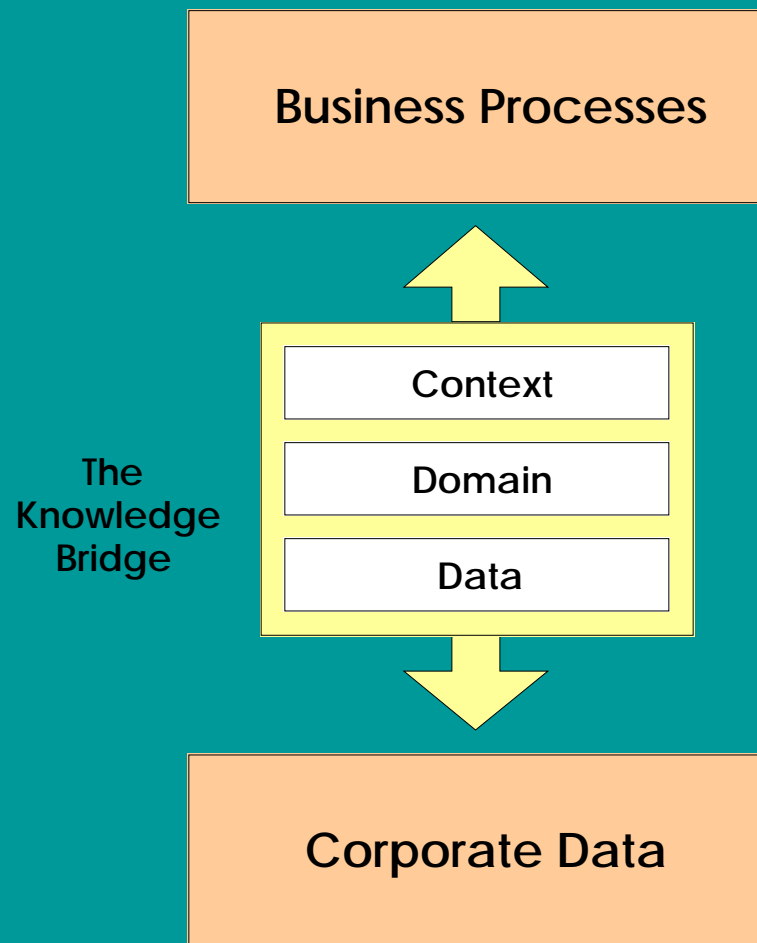
---



- Corporations are amassing large amounts of (unstructured) data
- Mere accumulation of data is not knowledge
- Knowledge emerges from the effective organization and interpretation of that data
- Knowledge drives business processes and supports the decision making tasks they entail

# Knowledge Bridges

---



- Knowledge bridges are a mechanism for managing data to impact business processes
- Knowledge bridges themselves require knowledge to provide effective organization and interpretation of the data

# Three Flavors of Knowledge

---

## ◆ Knowledge of context

- knowing why we are looking and what constitutes an answer
- context models capture the situated information need

## ◆ Knowledge of domain

- knowing what to look for and how to categorize responses
- domain models capture objects and associations of interest in the domain

## ◆ Knowledge of data

- knowing how to look and what defines a response
- data models capture the characteristics of the “objects” in the underlying data-space—meta-data, structure and content

# Domain Models

---

- ◆ Domain models represent the objects and associations of interest in the business domain
- ◆ We use a standard O-O representational framework for these models
- ◆ We draw from a core Tarragon ontology but typically customize for each application
- ◆ Main effort is
  - defining client specific objects and associations
  - instantiating the classes
  - documenting the models

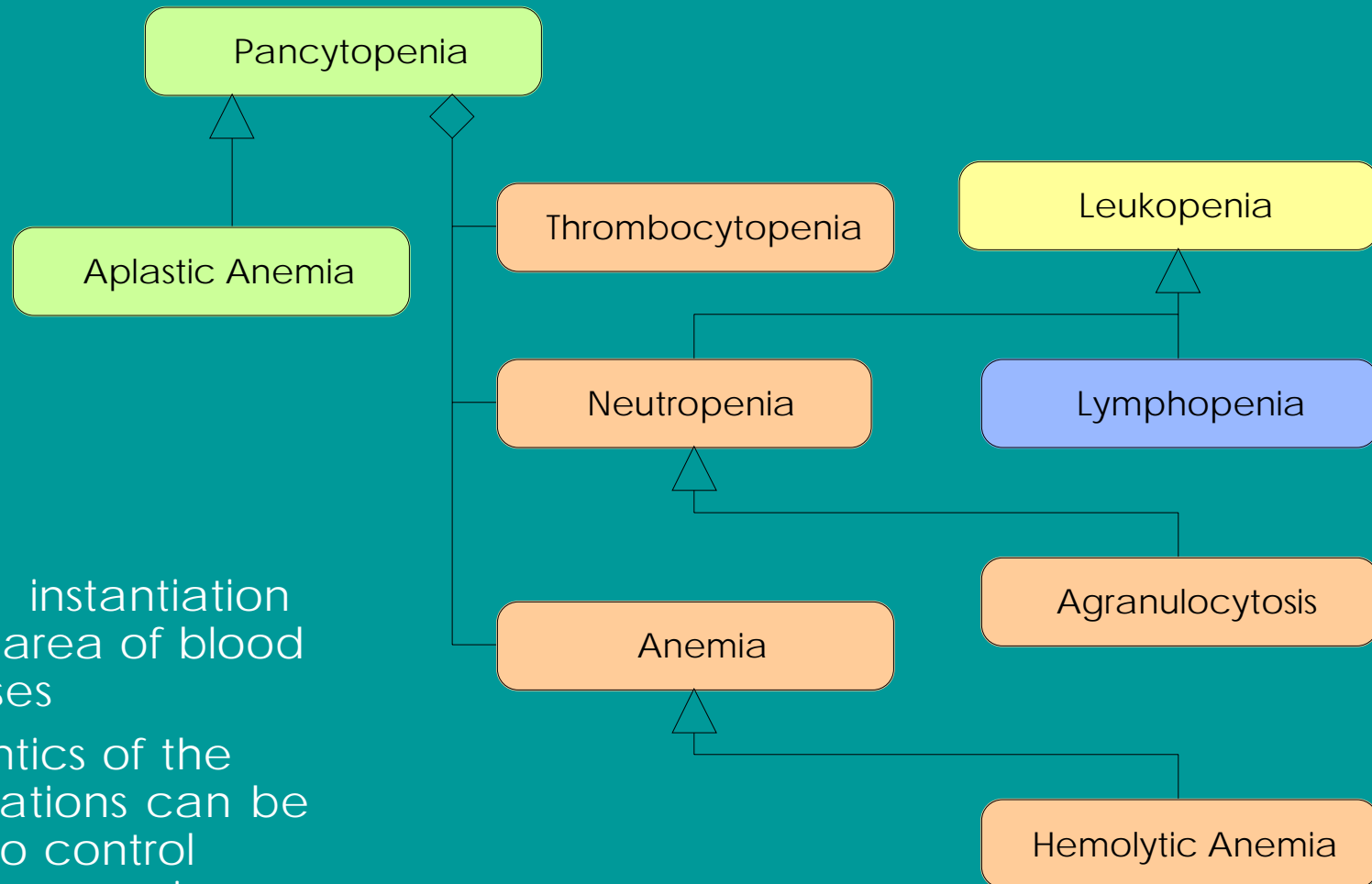
# Objects and Associations

---



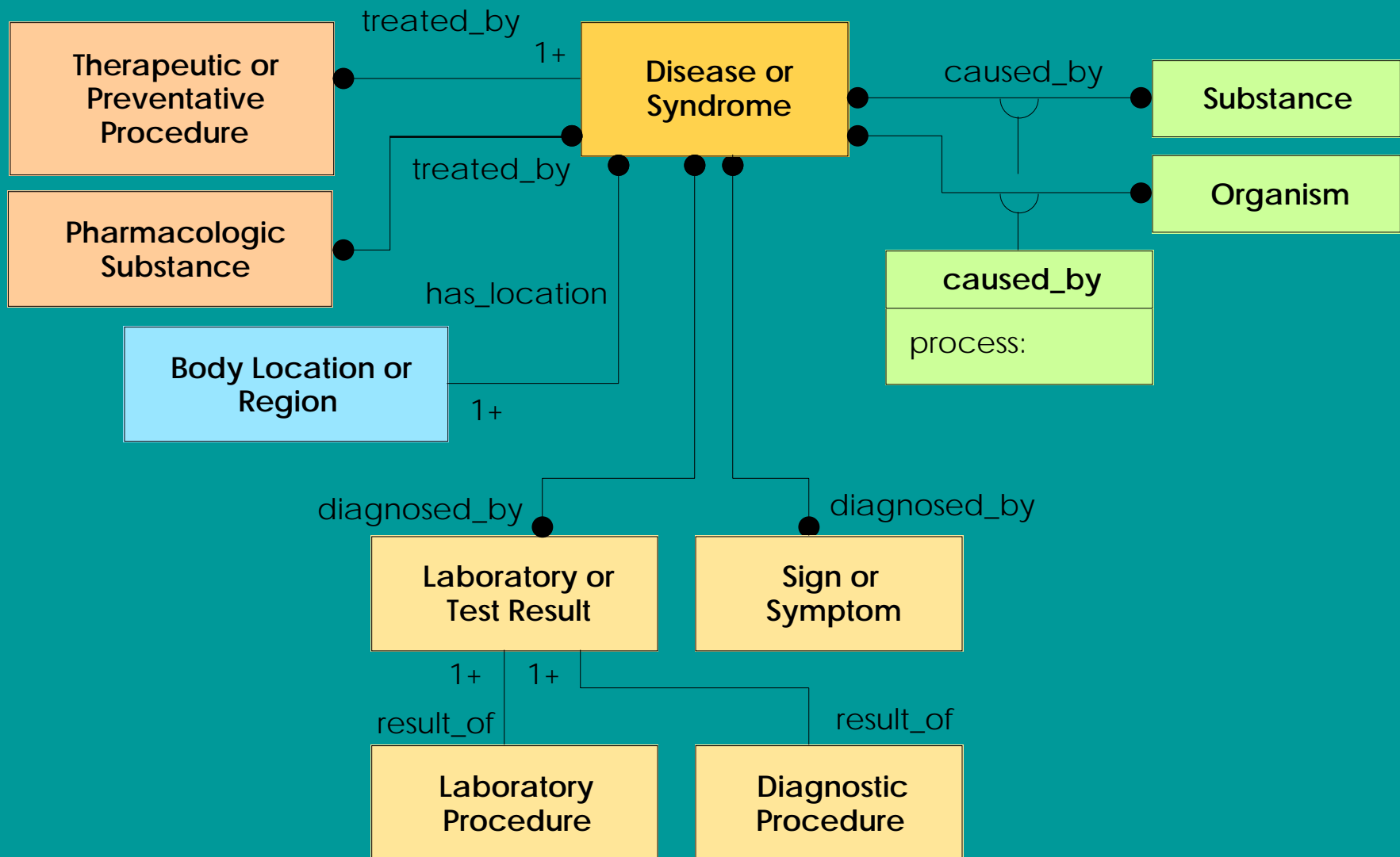
- Diseases are in hierarchical relationship to one another
- Can be aggregates of other diseases
- May co-occur with other diseases
  
- Tarragon medical ontology based on a public domain specification (UMLS)

# Model Instantiation



- Partial instantiation in the area of blood diseases
- Semantics of the associations can be used to control query expansion and results analysis

# Domain Model for Medicine



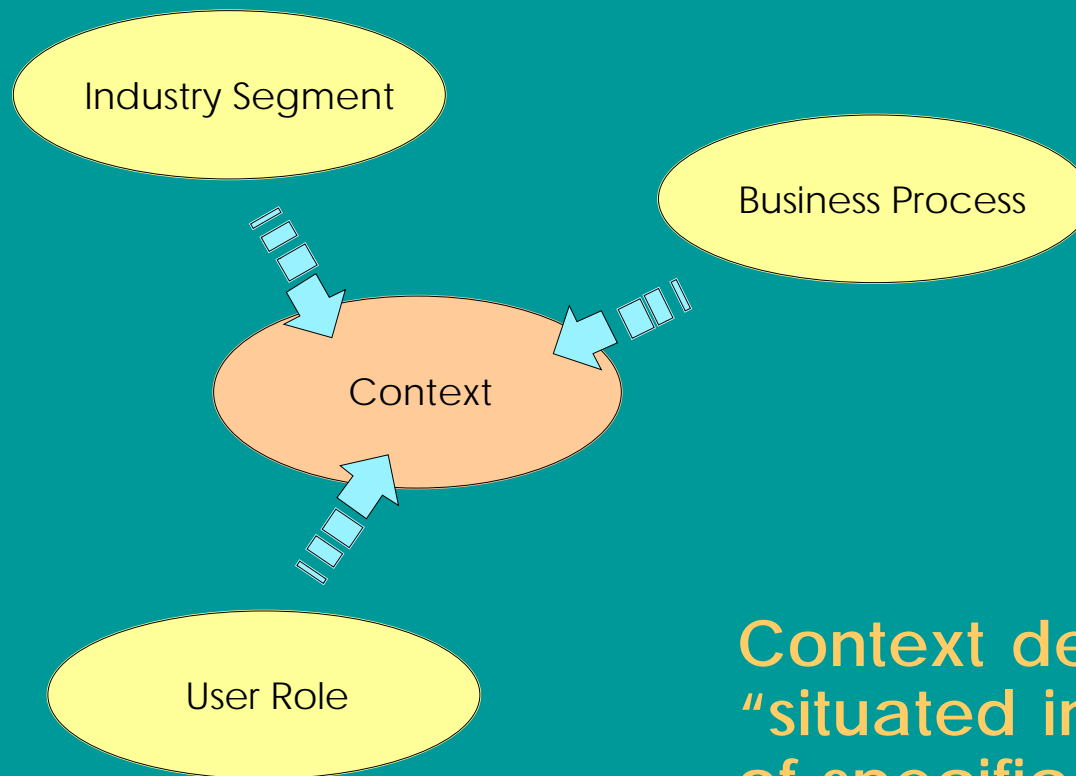
# Modeling Practice Notes

---

- ◆ **Class vs. attribute decision usually resolved on the basis of whether the context models need to reference the entities explicitly**
- ◆ **Typically only model attributes that are necessary for**
  - information access
  - internal administration of the models
- ◆ **Class methods usually not implemented as methods *per se*, but as functional specifications**

# Context Models

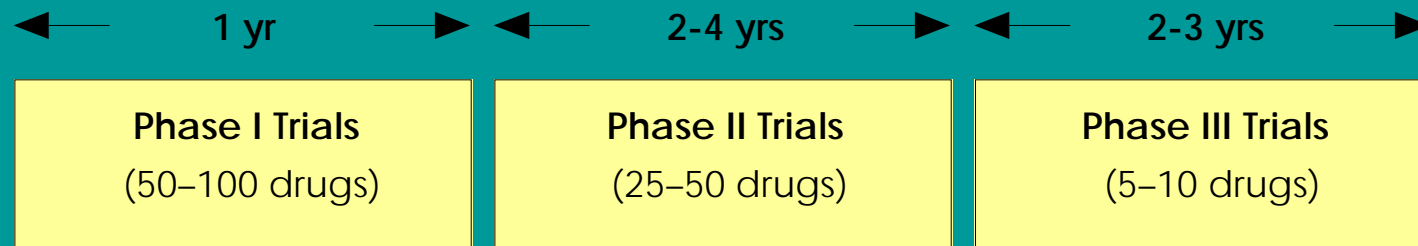
---



Context determines the “situated information need” of specific classes of users in specific task/function environments

# Clinical Development Model

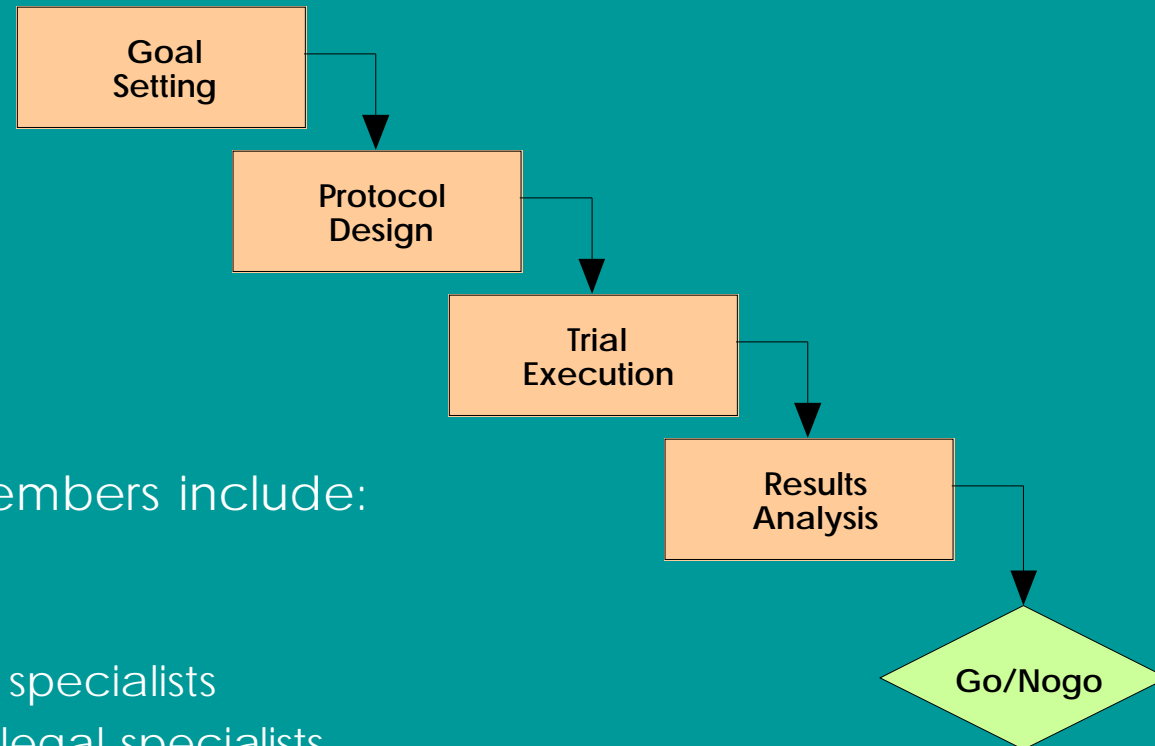
---



- 10-25 studies/drug
  - healthy volunteers
  - dosage ranges
  - initial toxicity
  - PK/PD characteristics
- 20-50 studies/drug
  - target population
  - dosage regimen
  - PK/PD parameters
- 2-5 studies/drug
  - diseased patients
  - NDA submission
- Average cost of clinical development is \$125M/drug
  - Average time of clinical development is 7.2 years
  - Each failed Phase III study delays approval at least 6 months and costs approx \$5M
  - Need to identify ineffective or unsafe drugs early in the development process

# Clinical Trial Sub-Tasks

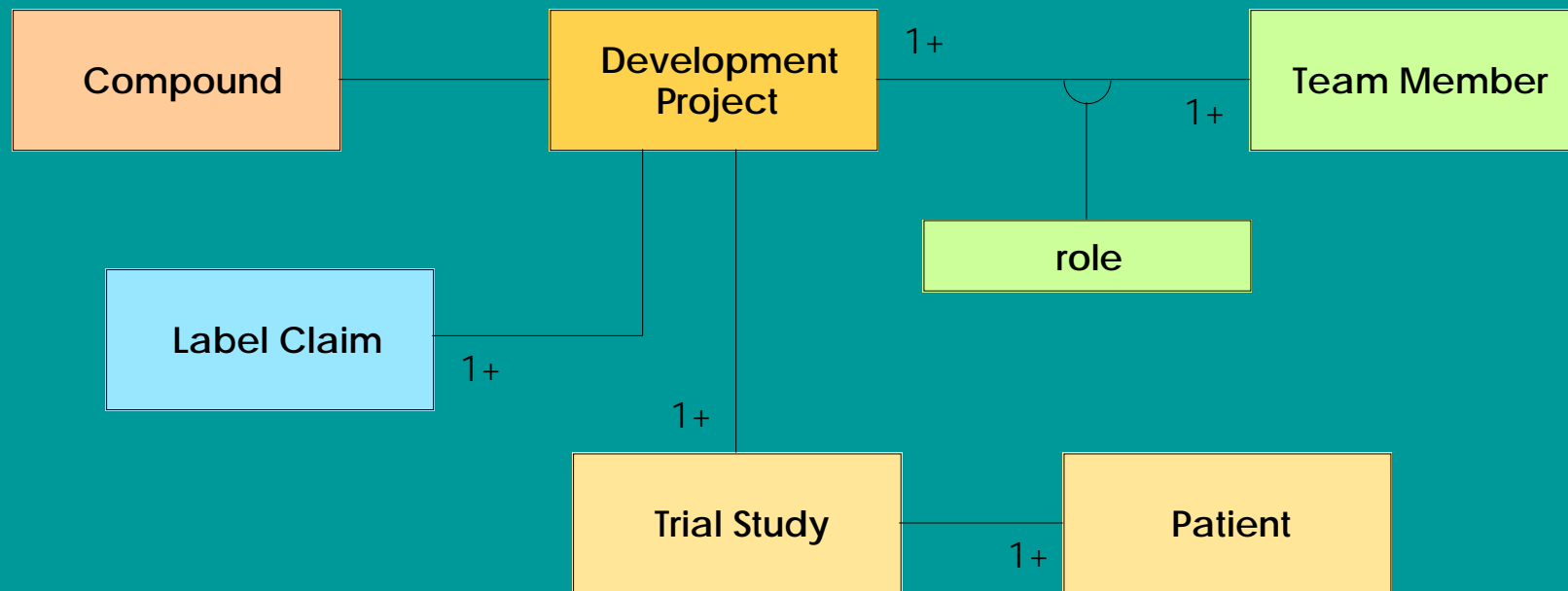
---



- Project team members include:
  - » clinicians
  - » scientists
  - » manufacturing specialists
  - » regulatory and legal specialists
  - » project and corporate management
- Each has different responsibilities and information needs
- The relative contribution of each user group changes as drugs move from Phase I to Phase III trials

# Process Context

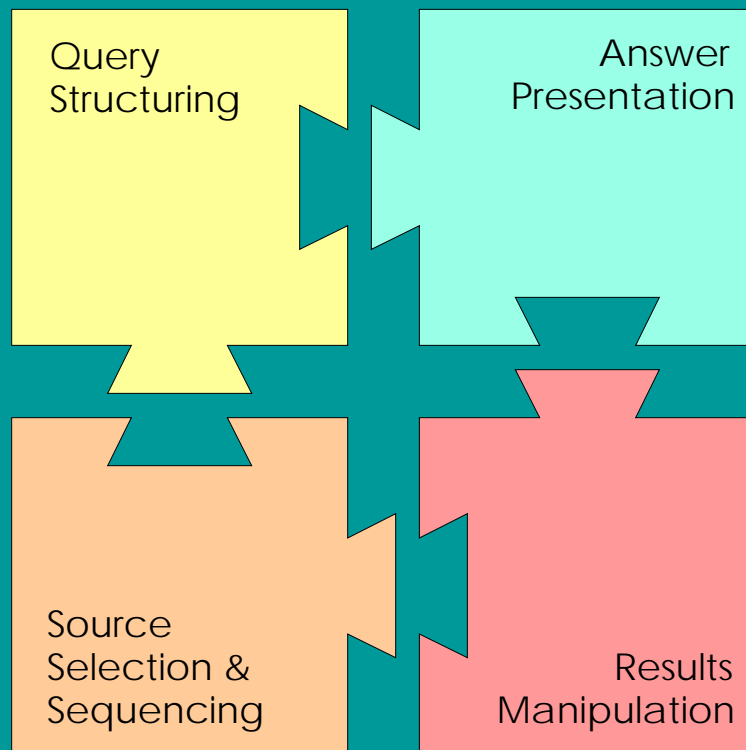
---



- In addition to flow and task models, we also often models that represent the context in which the flows and tasks operate
- These can also be represented as object models and used to organize and interpret data

# Context Model Components

---



- Specifications of information need
  - » query formulation and modality
  - » content and meta-data search criteria
- Strategies for information finding
  - » source selection and search plans
  - » cost and time constraints
- Definitions of answer structure and format
  - » data transformation and merging
  - » presentation style and modality

# Data Models

---

## ◆ Two views of the data/documents

- as structured objects in their own right
- as sources of “evidence” that can support assertions over the domain models

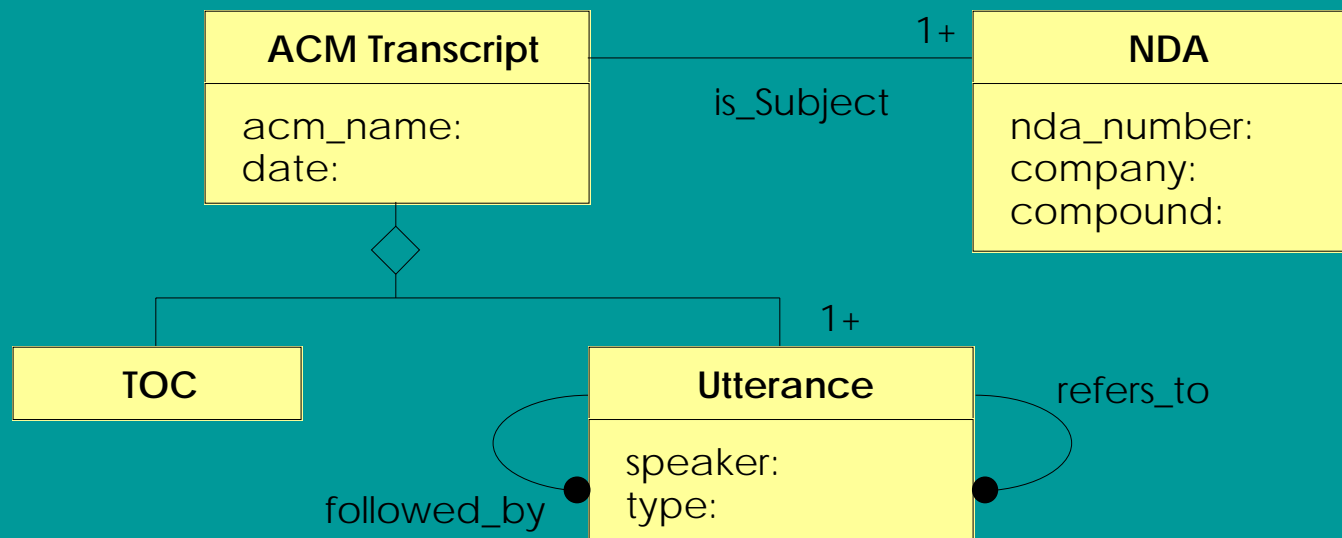
## ◆ In the structured object view we focus on

- data/documents as aggregates of sub-components that are of interest
- connecting these data/document models to other object classes in the domain

## ◆ In the evidential view we focus on

- data/document “features” that can be used in search and retrieval
- defining efficient ways of extracting these features given the data and system environment

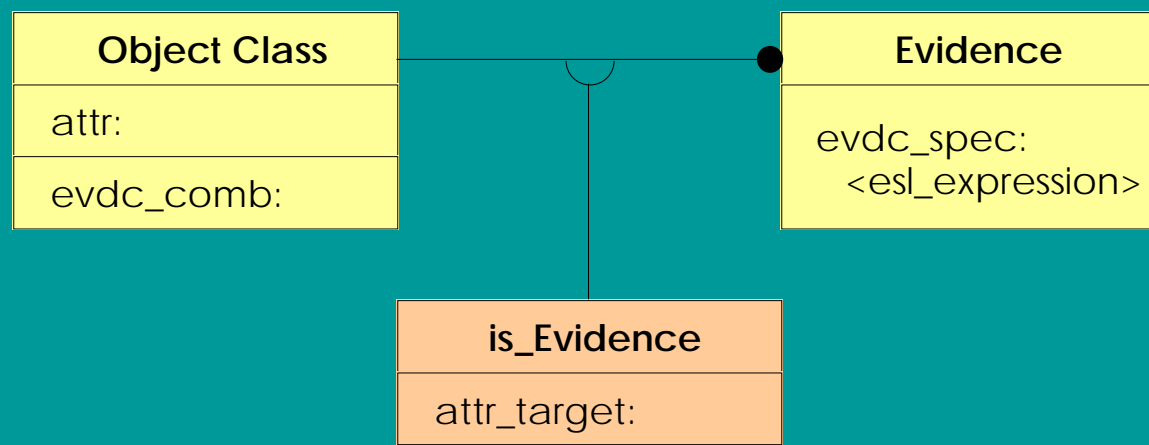
# Documents as Structured Objects



- Advisory Committee Meetings are held by the FDA as part of the final drug approval process
- The transcripts of the meeting are valuable sources of information about the likely concerns of the committee members

# Documents as Evidence

---



- All domain objects can be associated with zero or many Evidence objects
- Evidence objects specify the evidential support we will consider in making assertions about the object class
- Evidence specification is done using Tarragon's proprietary Evidence Specification Language (ESL)

# Evidence Specification Language

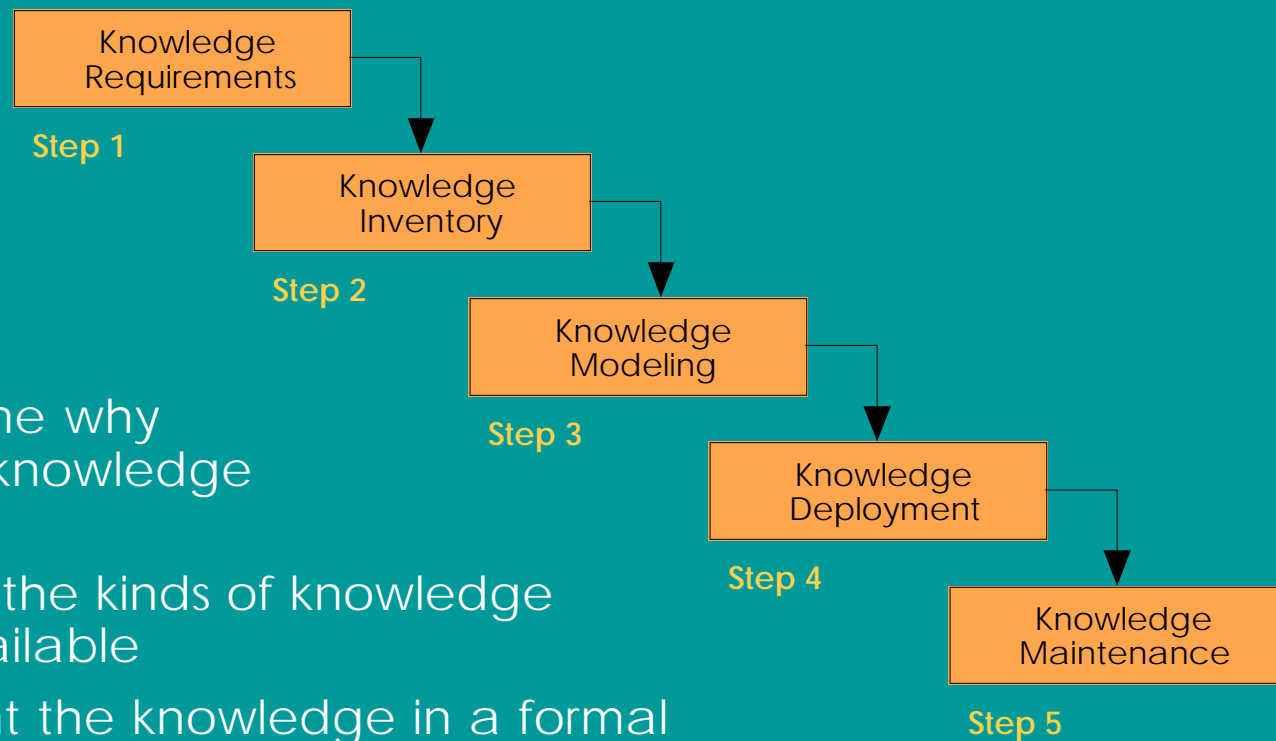
---

- ◆ ESL expressions define what constitutes evidence
- ◆ ESL contains primitives that specify
  - basic tokens
  - syntactic structures
  - collocations
  - special syntax
  - connectives

```
                                $token(tarragon)
                                $literal(Z39.50-1992)
$colloc:6:0($literal(KGB), $token:N(agent))
        $phrase $np    $compName  $prodName
                                $and   $or    $not
```

# Tarragon Methodology

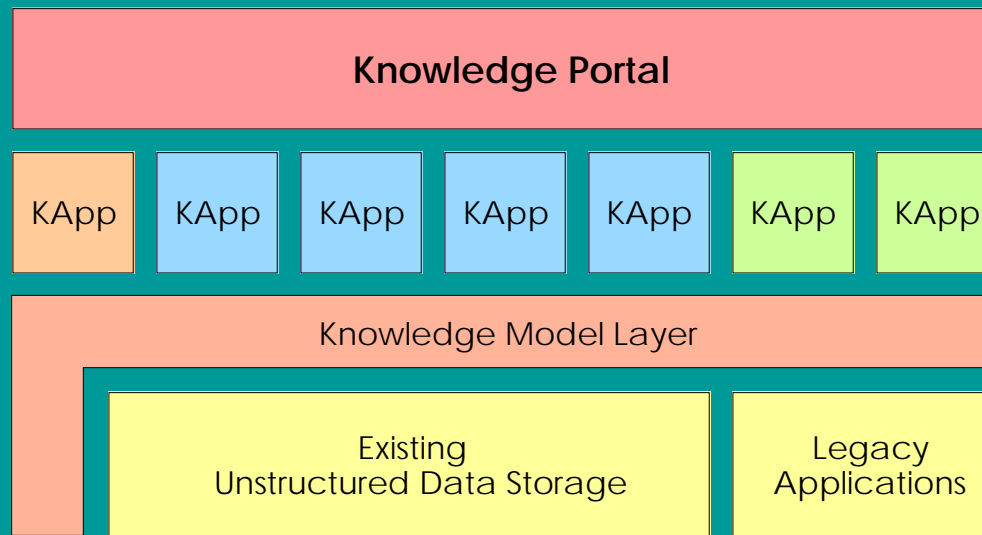
---



- We determine why and where knowledge is needed
- We catalog the kinds of knowledge that are available
- We represent the knowledge in a formal model
- We integrate the knowledge into the systems environment
- We provide training and tools to simplify maintenance

# Knowledge Portal Architecture

---



- Portal architecture supports integration with industry standard data storage—Verity S97 IS, Documentum EDMS, Oracle8i/ConText
- KApps are configurable by end-user organizations
- KApps can be “mixed-and-matched” to provide necessary functionality within the portal

# Summary

---

- ◆ Corporations invest significant amounts of time and money in the collection of large amounts of (unstructured) data
- ◆ Corporations do not realize the true value of this data because it is not *organized* and *interpreted* to meet the information needs of their critical business process
- ◆ Corporations need to deploy Knowledge Portals to provide integrated, task-oriented access to this business critical data