Ontologies: Ancient and Modern

Professor Nigel Shadbolt

School of Electronics and Computer Science

University of Southampton

The work of many people…

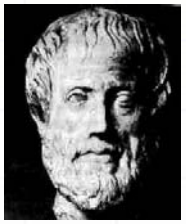
- Harith Alani
- Steve Harris
- Nick Gibbins
- Yannis Kalfoglou
- David Dupplaw
- Bo Hu
- Paul Lewis
- Srinandan Dashamapatra
- Hugh Glaser
- Les Carr
- David de Roure
- Wendy Hall
- Mike Brady
- David Hawkes
- Yorick Wilks
- :
- :

Structure

- A little history
- Ontologies and Knowledge Engineering
- Ontologies in the age of the WWW
- Ontologies in AKT
- Enduring problems and challenges
- Future progress

Ontologies – Realist Stance

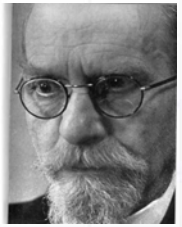
- We engage with a reality directly
  - Reality consists of pre existing objects with attributes
  - Our engagement may be via reflection, perception or language
- Philosophical exponents
  - Aristotle
  - Leibnitz
  - the early Wittgenstein
  - :



- Language and logic pictures the world
- Seen as a way of accounting for common understanding
- Promises a language for science

Constructivist Stance

- There is no simple mapping into external objects and their attributes in the world
- We construct objects and their attributes
  - This construction may be via intention and perception, it may be culturally and species specific
- Philosophical exponents
  - Husserl
  - Heidegger
  - Later Wittgenstein
  - :



- Language as games, complex procedures, contextualised functions that construct a view of the world

Ontologies - Current Context

- The large metaphysical questions remain
  - What is the essence of being and being in the world
- Our science and technology is moving questions that were originally only philosophical in character into practical contexts
  - Akin to what happened with natural philosophy from the 17th century – chemistry, physics and biology
- As our science and technology evolves new philosophical possibilities emerge
  - Particularly when we look at *knowledge* and *semantic* based processing
  - We will return to this…

Knowledge Engineering: Evolution

general-purpose search engines (GPS)	first-generation rule-based systems (MYCIN, XCON)	emergence of structured methods (early KADS)	mature methodologies (CommonKADS)

1965 1975 1985 1995

=> from art to discipline =>

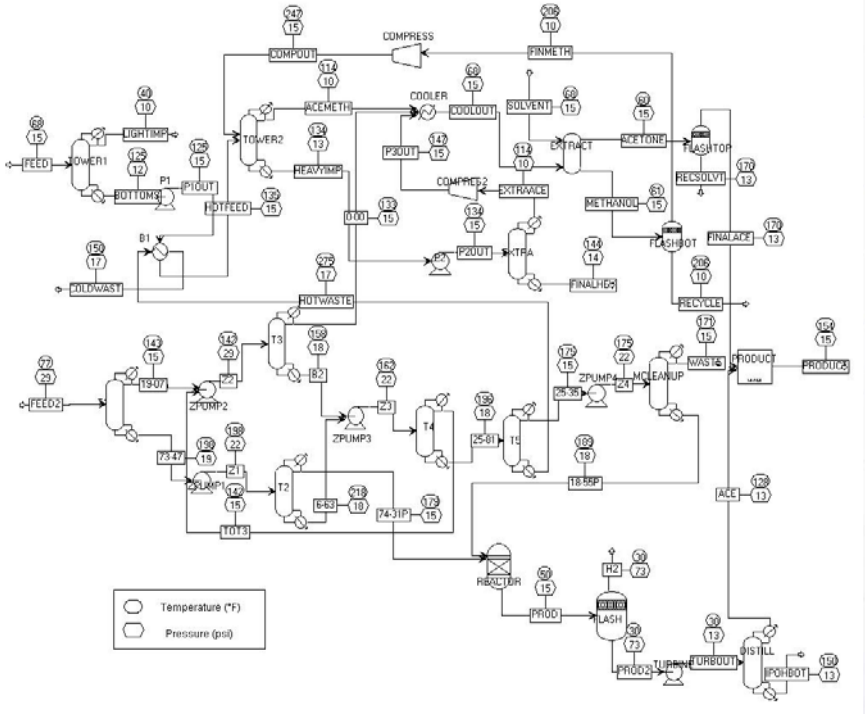
Knowledge Engineering: Principles

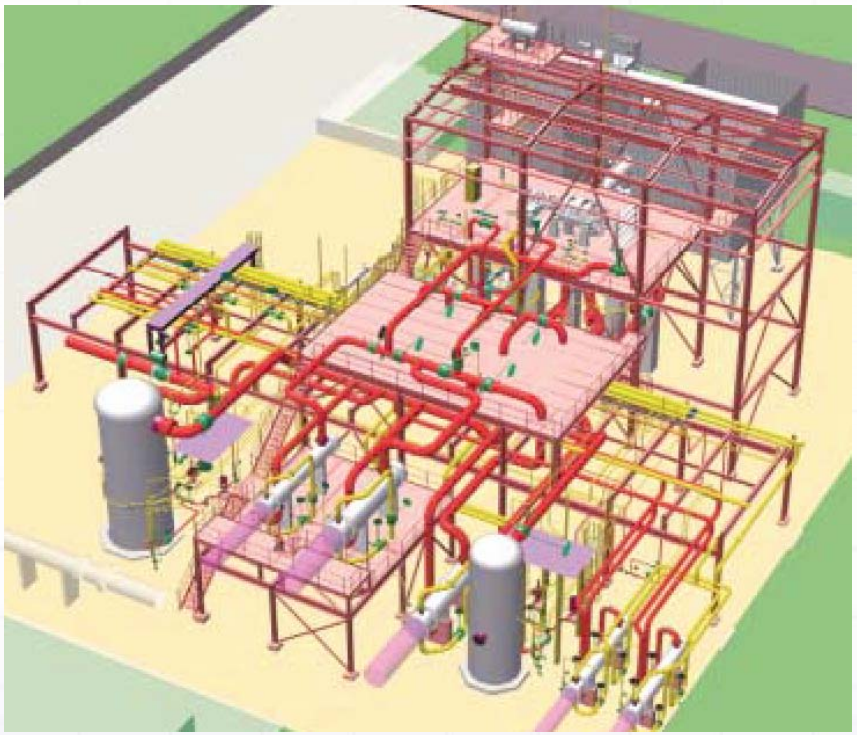
- Knowledge engineering is not about transfer but about modelling aspects of human knowledge
- The knowledge level principle: first concentrate on the conceptual structure of knowledge and leave the programming details for later
- Knowledge has a stable internal structure that can be analysed by distinguishing specific knowledge types and roles

Ontologies in Knowledge Engineering

- A variety of tools developed to support the acquisition and modelling of knowledge structures
- Many of the patterns developed could be viewed as abstract conceptual structures – ontologies were there throughout and became more prominent
- There were explicit ontologies for modelling domain classes and their relationships
- There were claims and counter claims about how task neutral such conceptual structures could be

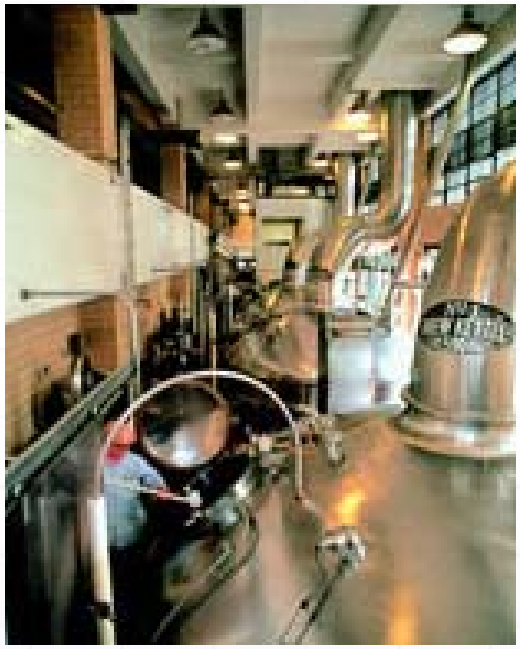
Constraint and Frame Oriented Knowledge-Based System

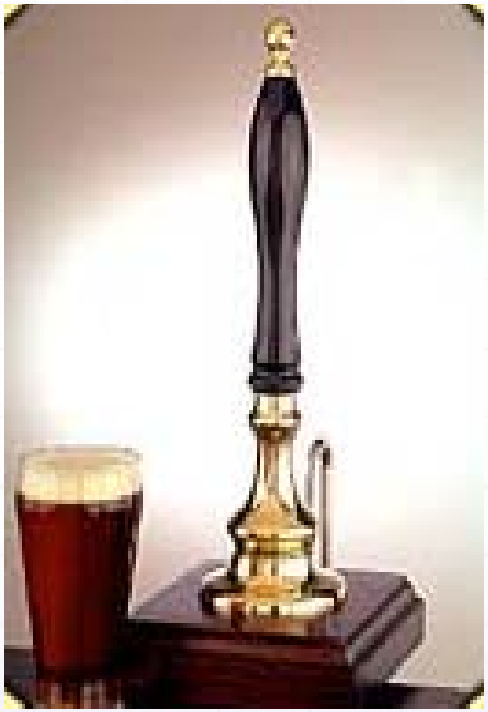




McBrien, A.M., Madden, J and Shadbolt, N.R. (1989). Artificial Intelligence Methods in Process Plant Layout. *Proceedings of the 2nd International Conference on Industrial and Engineering Applications of AI and Expert Systems*, pp364-373, ACM Press

Perceptually Oriented Knowledge-Based System





Bull, H.T, Lorrimer-Roberts, M.J., Pulford, C.I., Shadbolt, N.R., Smith, W. and Sunderland, P. (1995) Knowledge Engineering in the Brewing Industry. *Ferment* vol.8(1) pp.49-54.

And then the Semantic Web

- Fundamentally changed the way we thought about KA and knowledge management
- Suggested a different way in which knowledge intensive components could be deployed
- Also brought together a community unencumbered by close attention either to AI or Knowledge Engineering
- New funding opportunities…

Advanced Knowledge Technologies IRC

AKT started Sept 00, 6 years, £8.8 Meg, EPSRC

www.aktors.org

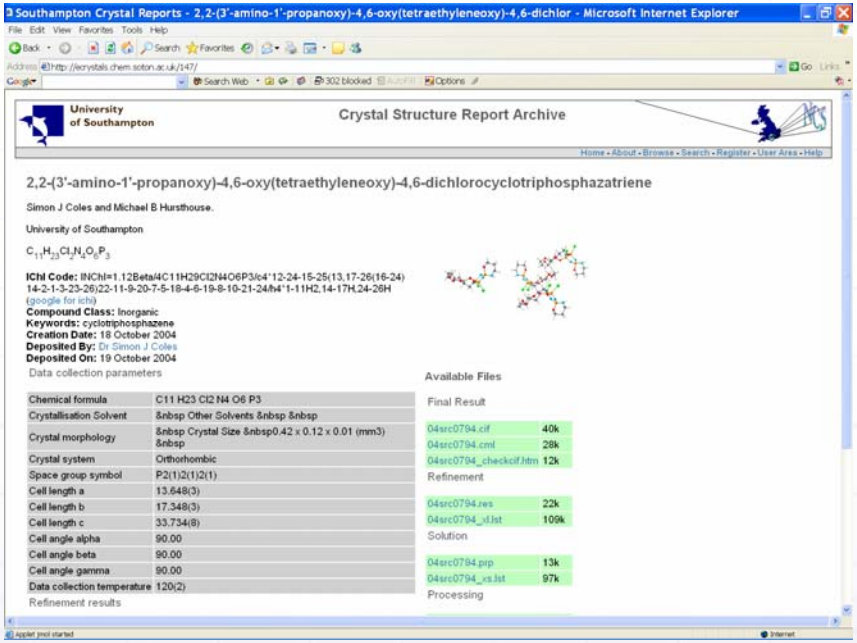
Around 65 investigators and research staff

Ontological Lessons Learnt

- The content is primary
  - It needs rich semantic annotation via ontologies
  - Services emerge/designed to exploit the content
- Lightweight ontologies work
  - In support of rapid interoperability
- Ontologies as mediators
  - Aggregation as a key capability
- Ontologies are socio technical
  - Act as declarative agreements on complex social practice

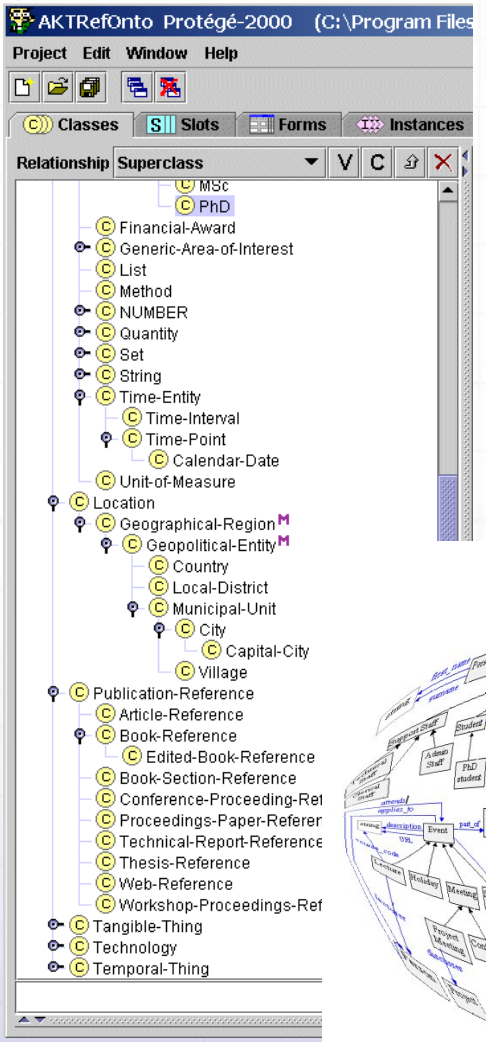
Primacy of content - eCrystal

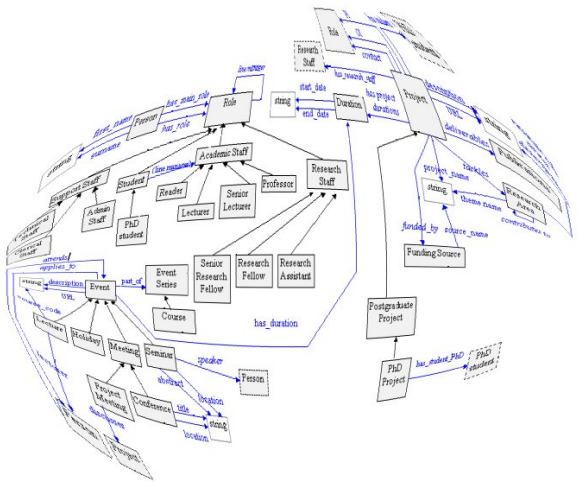
- Simple but powerful use of existing conceptual structures
- Domain markup language
- Close to a realist interpretation of an ontology
- Protégé Requirement
  - Import of simple CML schema



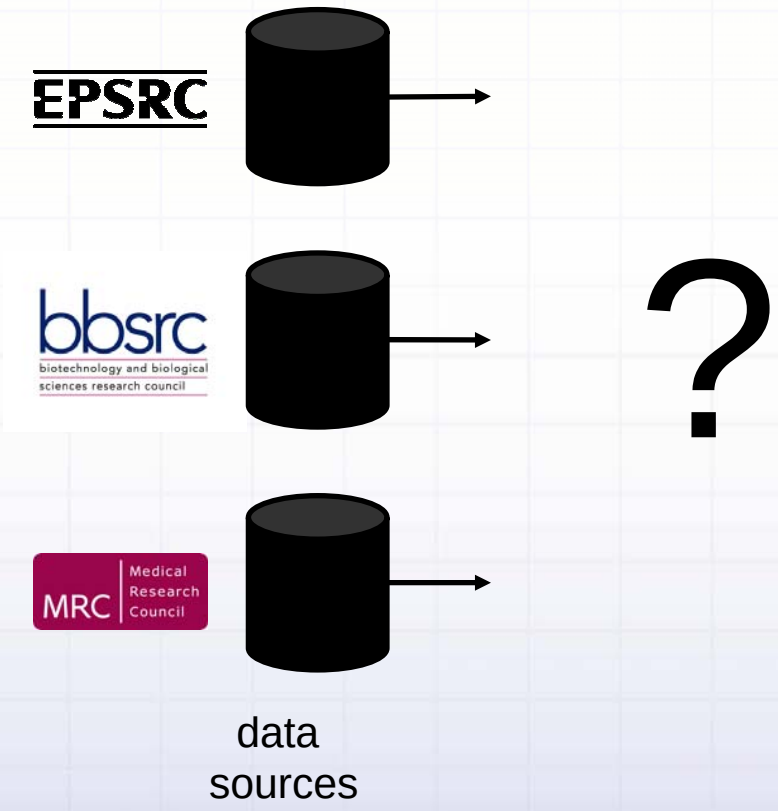
The AKT Ontology

- Designed as a learning case for AKT
- Adopted for our own Semantic Web experiments including CS AKTive
- Uses a number of Upper Ontology fragments
- Reused in many contexts

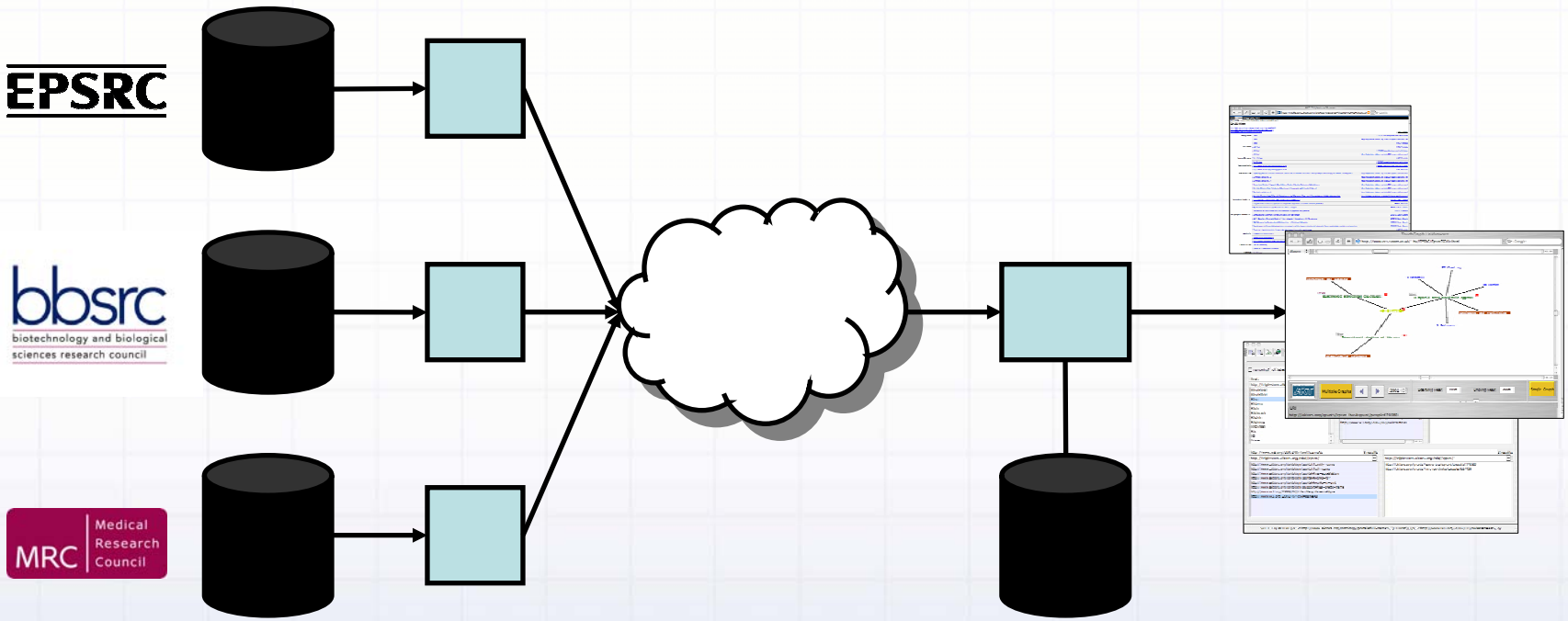




Mediation and Aggregation: UK Research Councils



A Proposed Solution



data sources

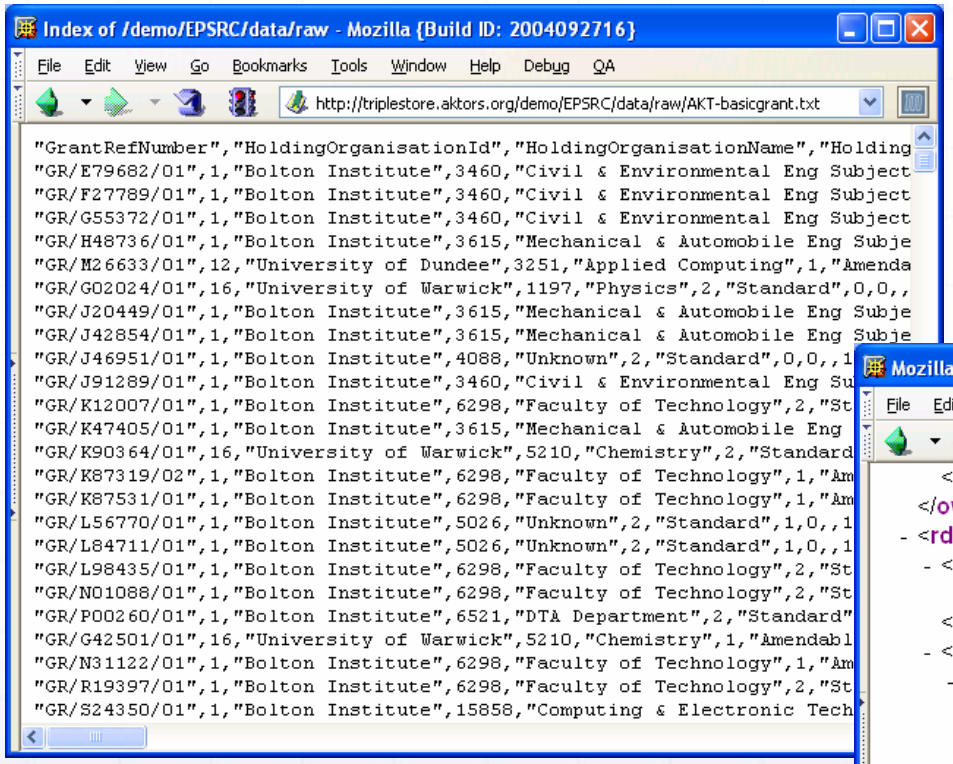
gatherers

Ontology

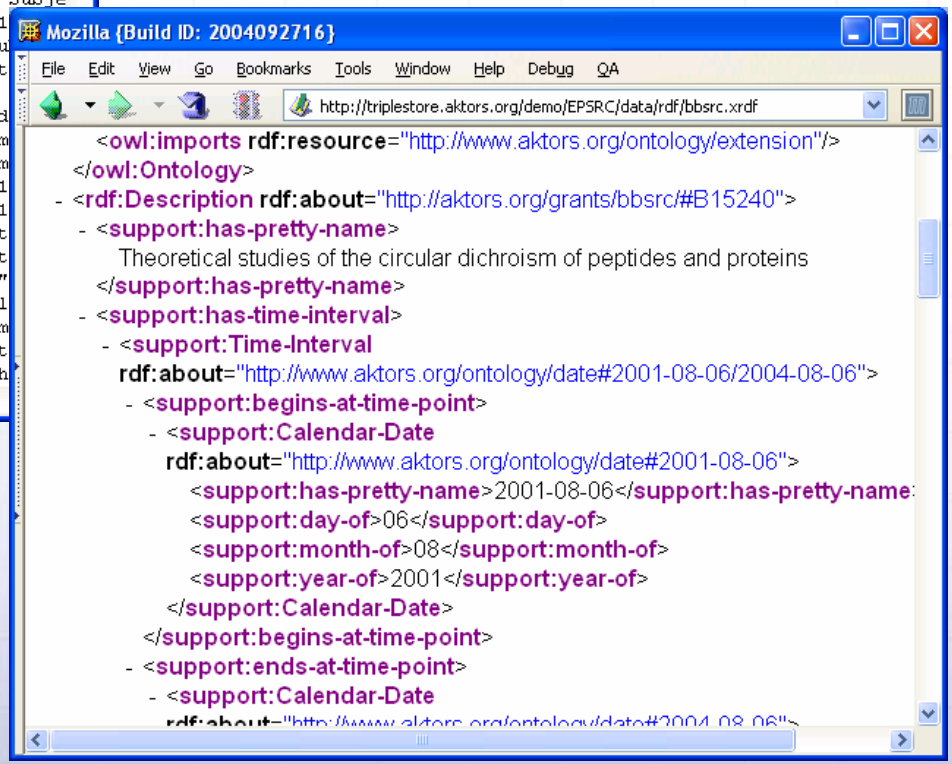
knowledge repository (triplestore)

applications

Mediation and Aggregation: UK Research Councils



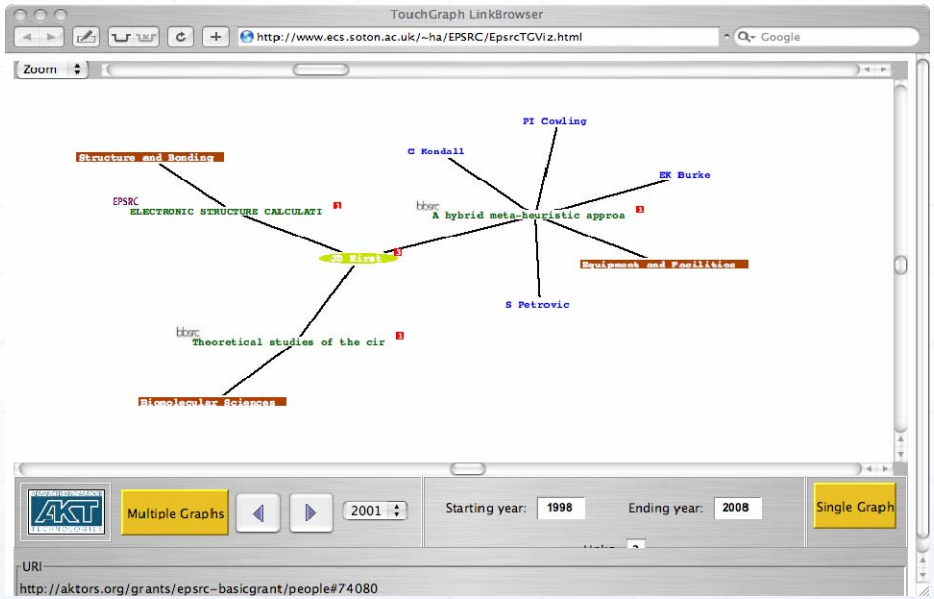
Raw CSV data
Heterogeneous tables



Processed RDF information
Uniform format for files

An Application Service

- Relatively simple could yield real information integration and interoperability benefits
- Reuse was real but again lightweight
- Ontology winnowing would be very useful
- Protégé Requirement
  - Stats packages for ontologies – how to map back from implemented ontologies to the statistics of use

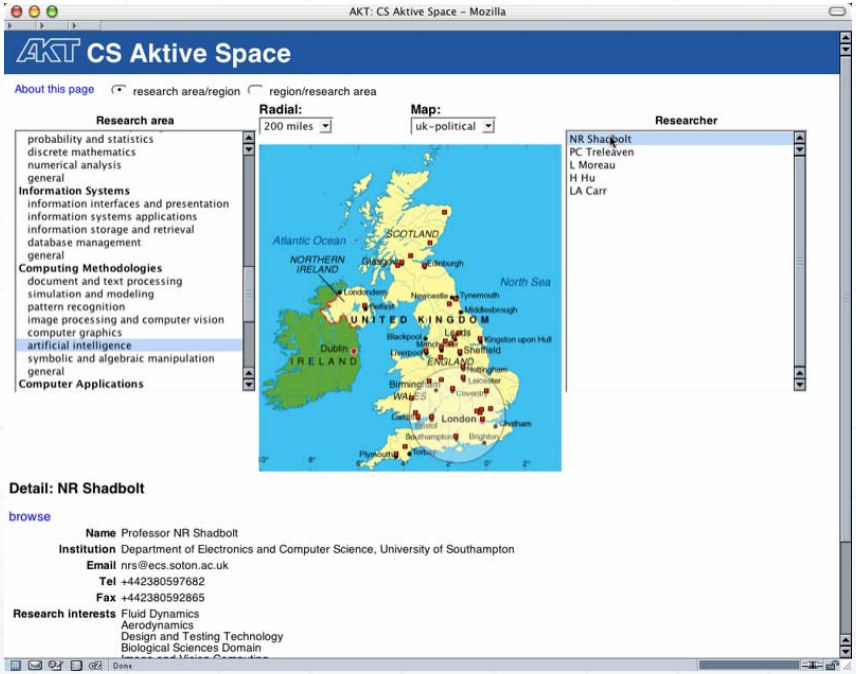


Mediation and Aggregation: CS AKTive Space

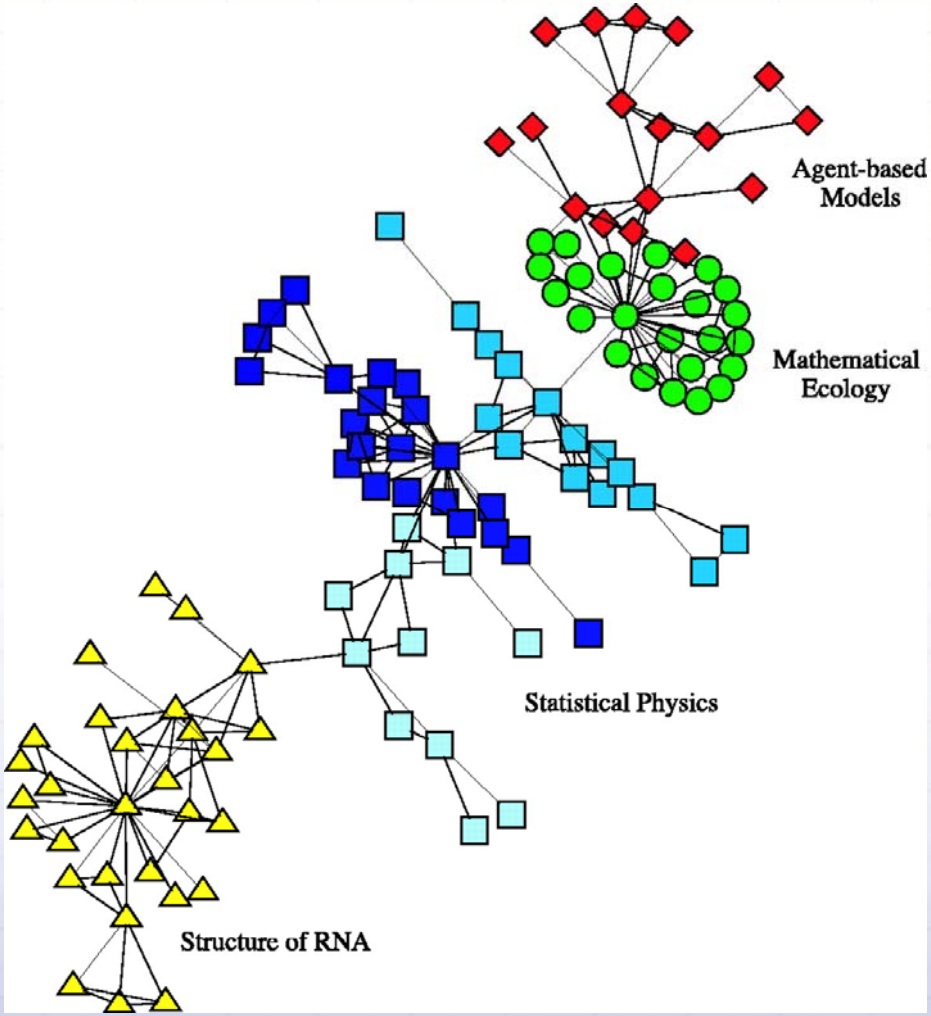
- 24/7 update of content
- Content continually harvested and acquired against community agreed ontology
- Easy access to information gestalts - who, what, where
- Hot spots
  - Institutions
  - Individuals
  - Topics
- Impact of research
  - citation services etc
  - funding levels
  - Changes and deltas
- Dynamic Communities of Practice…

Mediation and Aggregation: CS AKTive Space

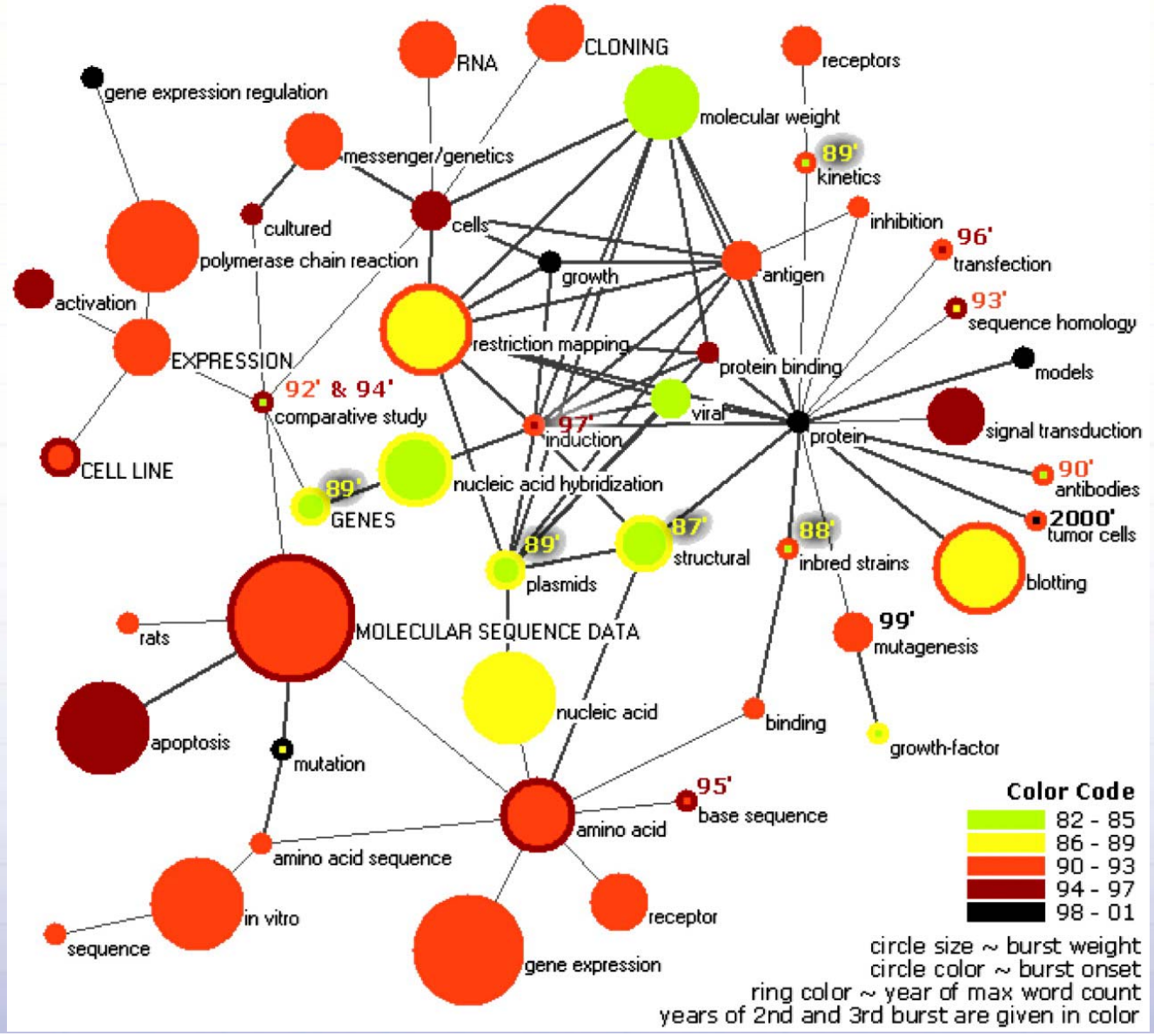
- Content harvested and published from multiple Heterogeneous Sources
- Higher Education directories
- 2001 RAE submissions
- UK EPSRC project database (all grants awarded by EPSRC in the past decade)
- Detailed data on personnel, projects and publications harvested for:
  - all AKT partners
  - all 5 or 5* CS departments in the UK
  - Automatic NL mining: Armadillo
- Additional resources
  - All UK administrative areas (from ISO3166-2)
  - All UK settlements listed in the UN LOCODE service
  - (and they're all integrated via the AKT reference ontology)
- Protégé Requirement
  - Support between a frame and DL oriented perspective



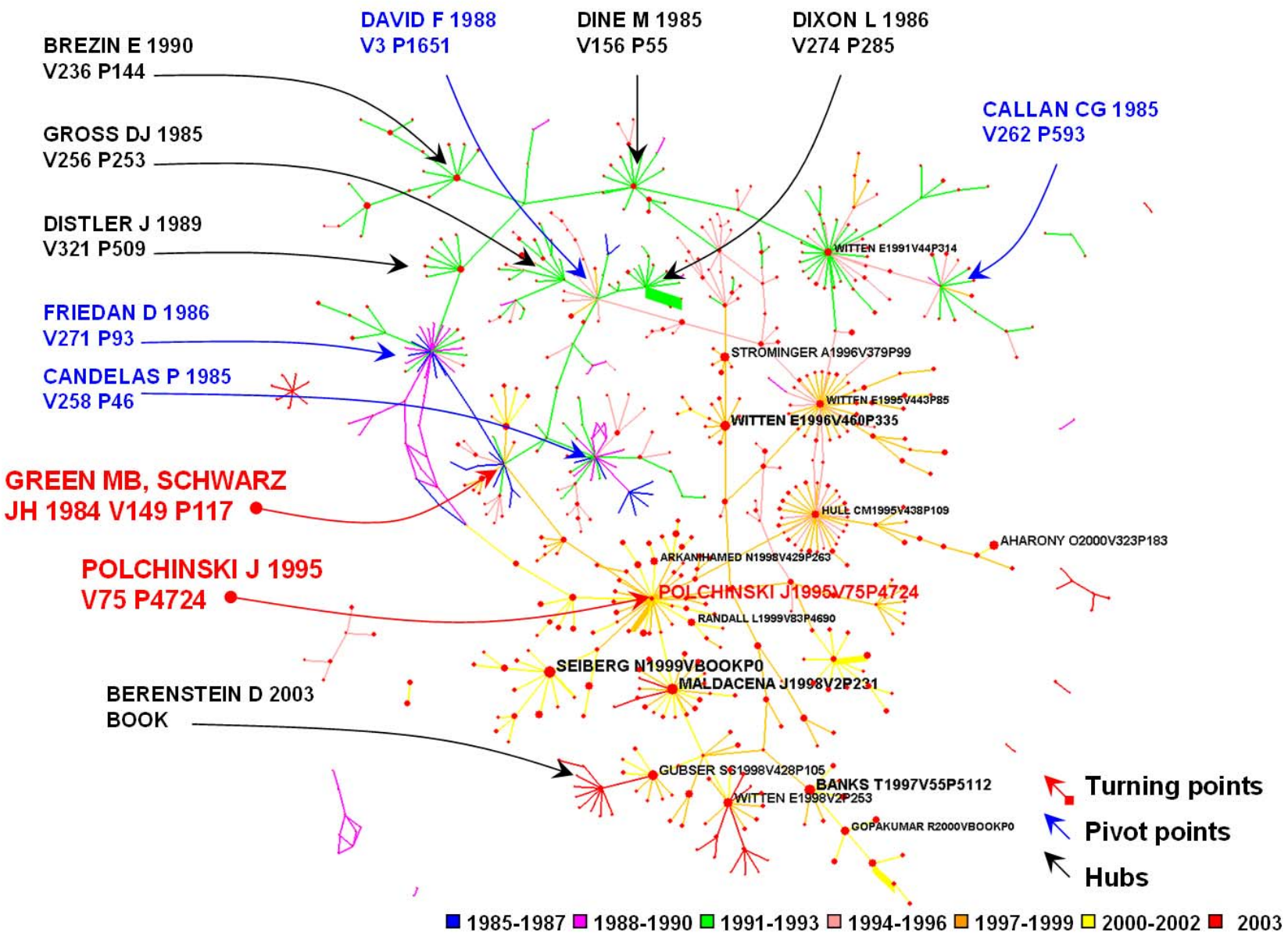
Extending the model – knowledge mapping: author mapping



Extending the model – knowledge mapping: topic bursts



Extending the model – knowledge mapping: pathfinder

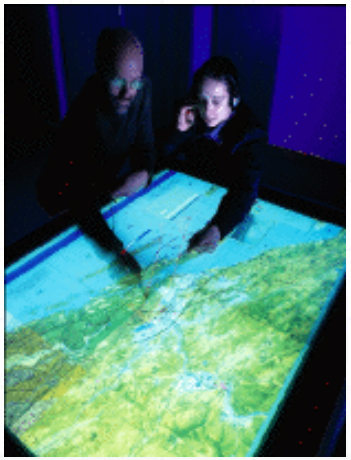


DTC Project: OOTW

- improved situational awareness in the coordination, planning and deployment of humanitarian aid operations
- integrating operationally-relevant information
- discovery and exploitation of novel information sources

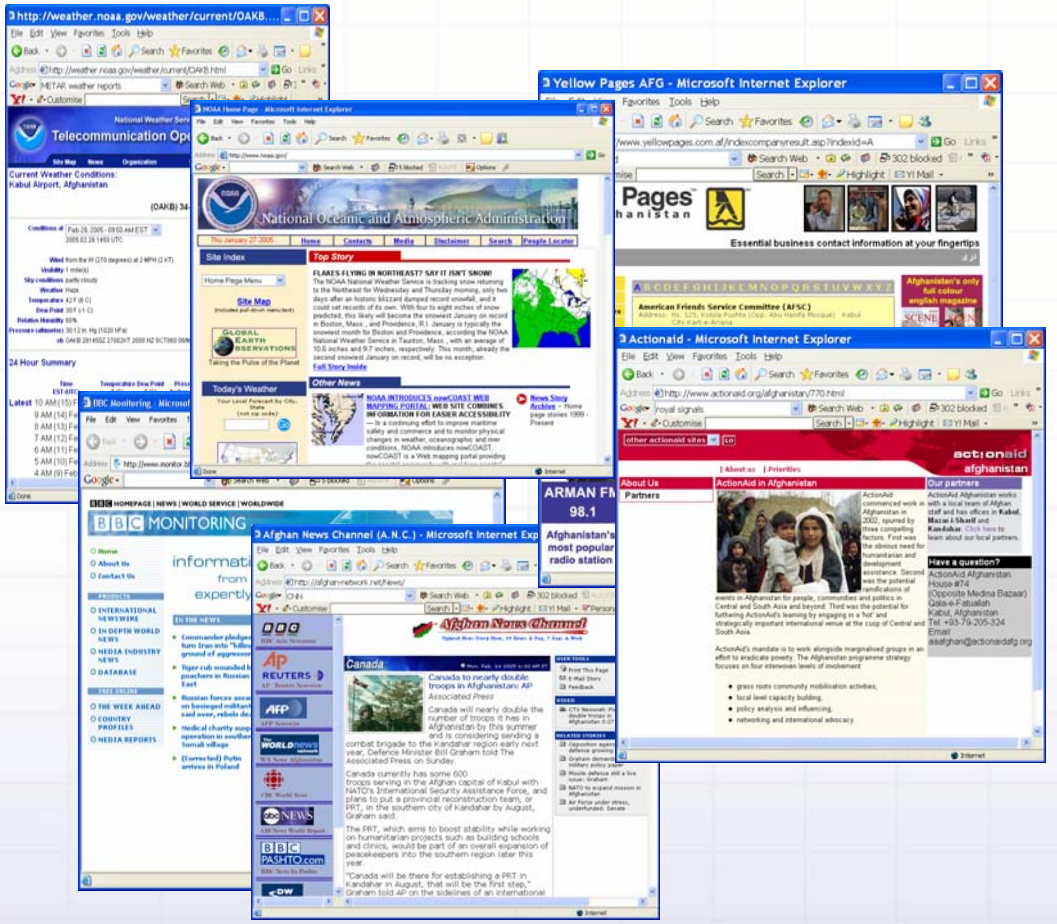
Capability Requirements

- Event notification
- Facilitation of agent communication networks
- Coordination, planning and deployment of humanitarian aid efforts
- Collaboration of military and humanitarian aid operatives
- Semantically-enriched decision support



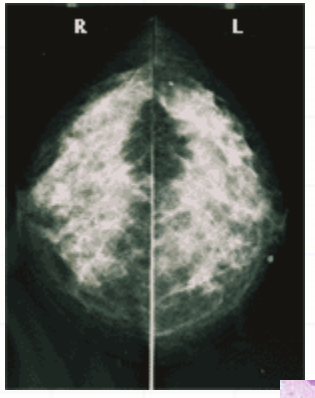
Information Resources

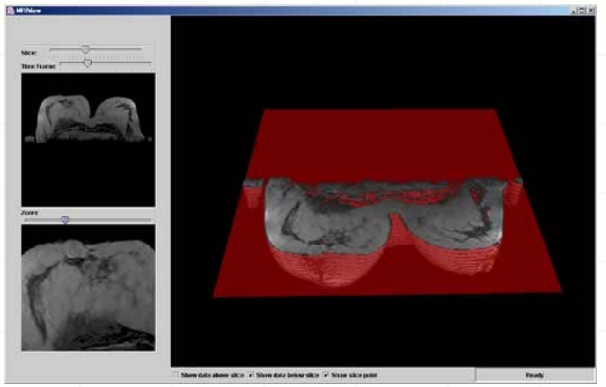
- exploitation of semantically heterogeneous and physically disparate information sources, e.g.
  - tactical datalinks
  - METAR weather reports
  - BBC monitoring service
  - other news feeds
  - NGO reports
  - institutional websites, e.g. NGDC, NOAA, SPC

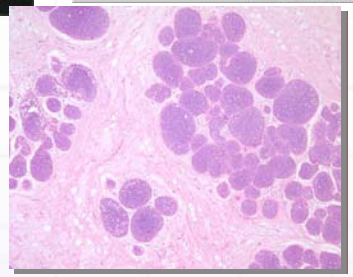


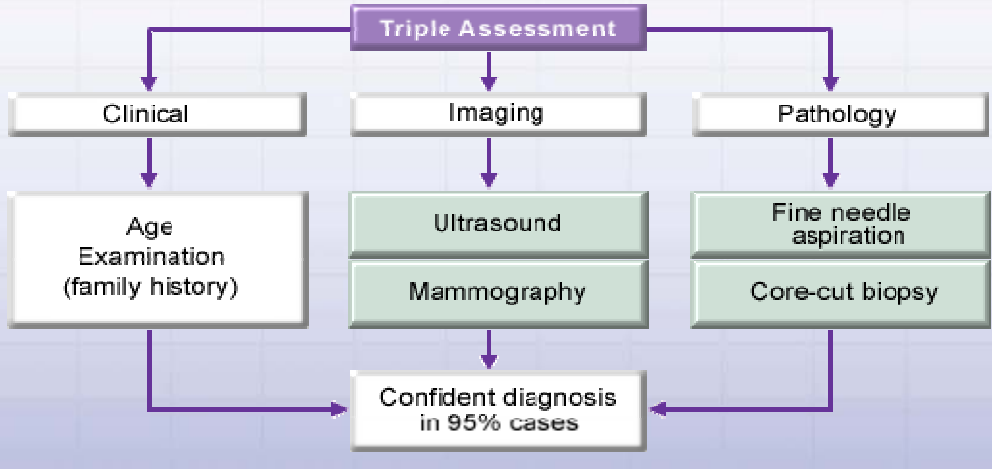
Complex Ontologies: MIAKT

- Multiple stakeholders
- Multiple viewpoints and ontologies (some implicit)
  - Breast imaging – X-ray, ultrasound, MRI
  - Clinical examination
  - Microscopy – cells and tissues (also, hormone receptors)
- Local dialects in use
- Variation between countries due to factors such as insurance claims!
- Protégé Requirement - Support for multimedia annotation
- Protégé Requirement - Supporting and Mapping Between Multiple Perspectives



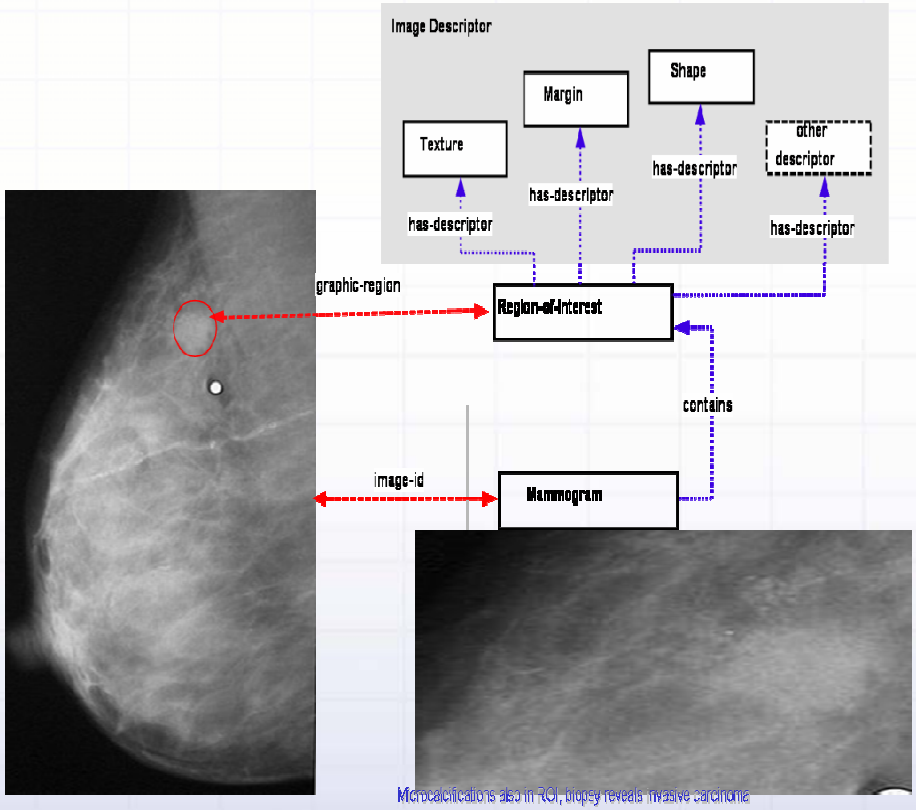




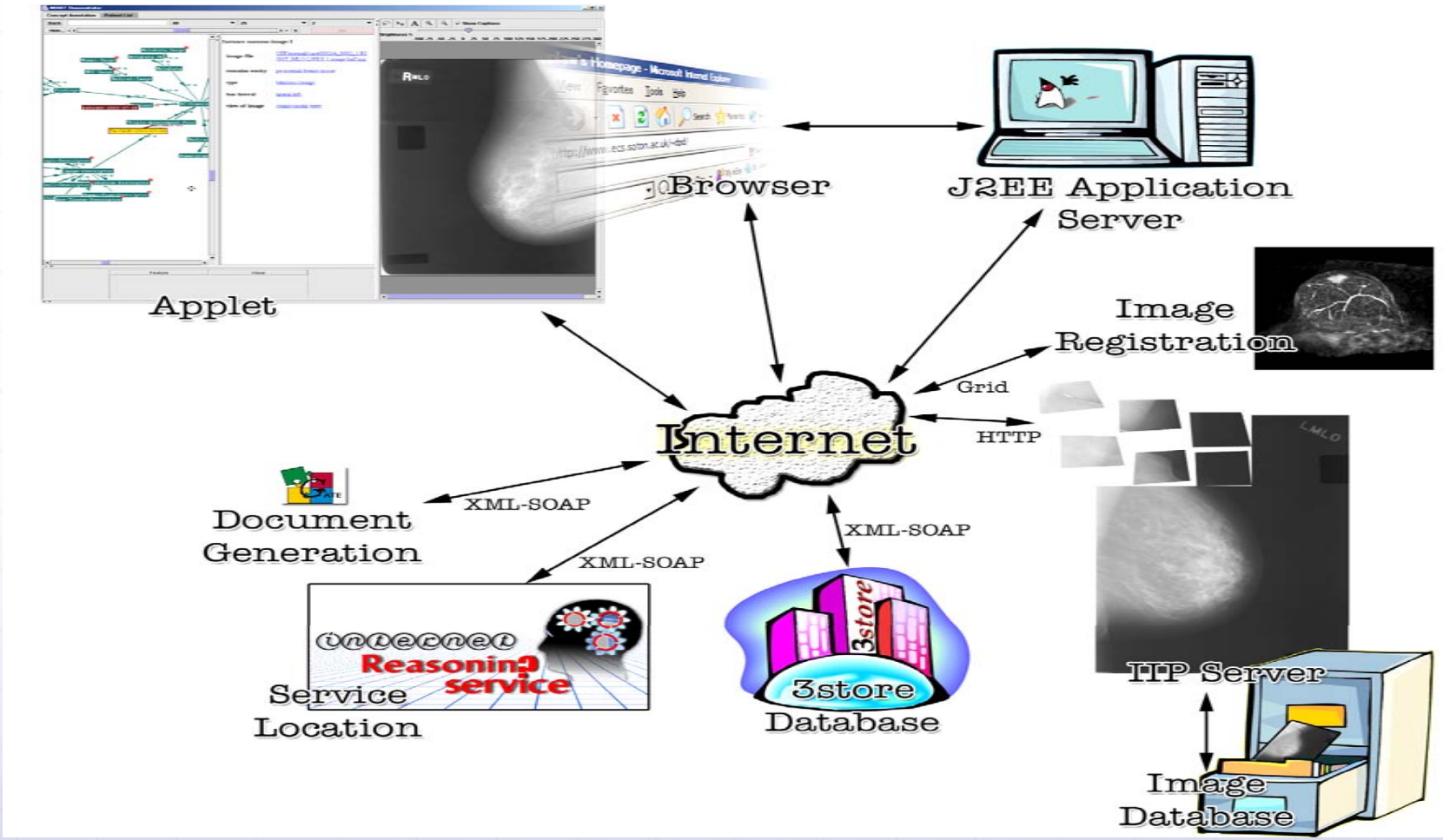


Ontologies in MIAKT

- Information indexed against ontologies can be retrieved via concept labels
- Image retrieval for annotated images
- Recognition of “significant” condition necessary
- Labels are outcome of classification
- Entered into ontology as declarative concepts



The MIAKT Framework

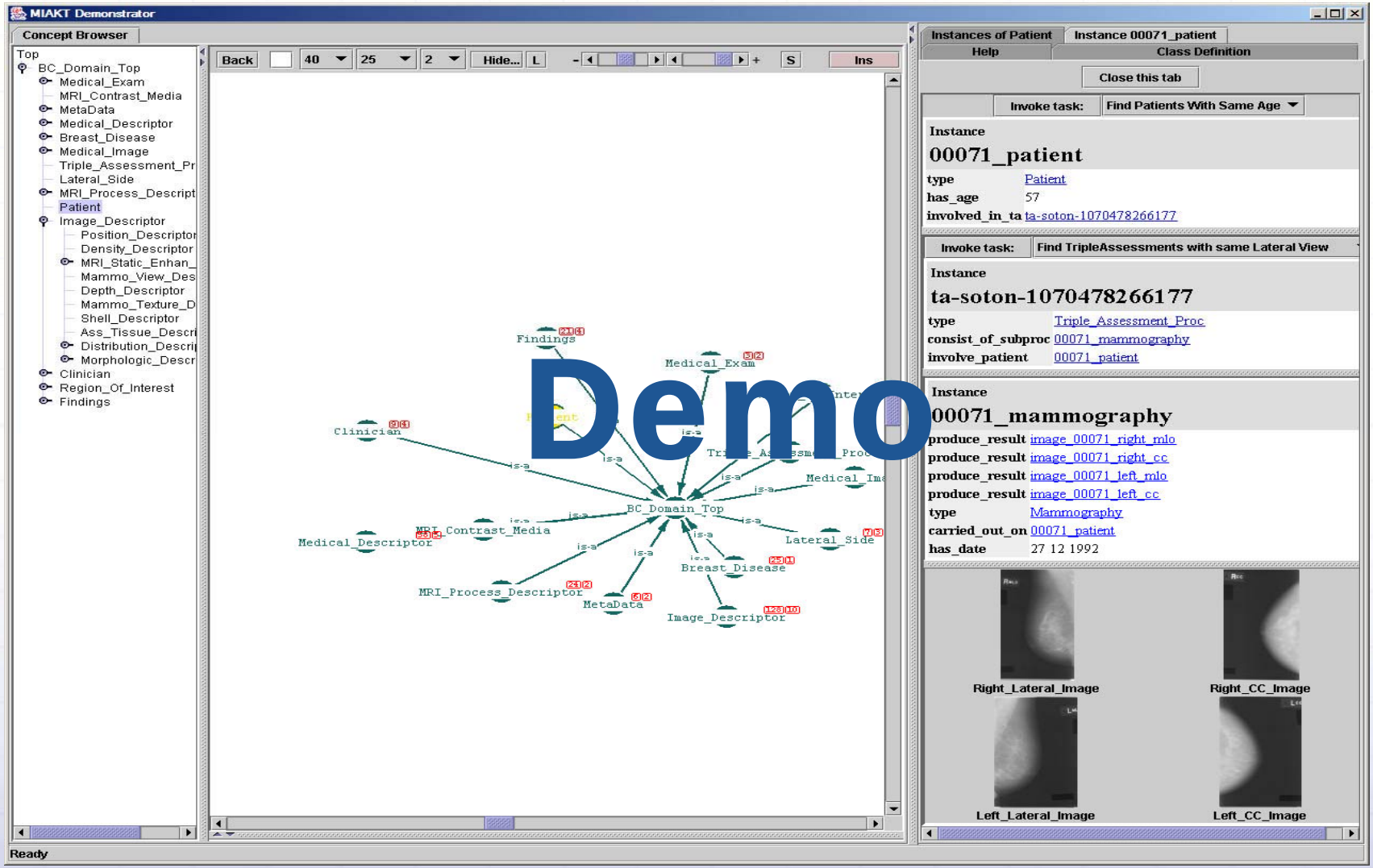


Patient Cases in RDF

```
<rdf:Description rdf:about='#g1p78_patient'>
  <rdf:type rdf:resource='#Patient'/>
  <NS2:has_date_of_birth>01.01.1923</NS2:has_date_of_birth>
  <NS2:involved_in_ta rdf:resource='#ta_soton_000130051992'/>
</rdf:Description>

<rdf:Description rdf:about='#ta_soton_000130051992'>
  <rdf:type rdf:resource='#Multi_Disciplinary_Meeting_TA'/>
  <NS2:involve_patient rdf:resource='#g1p78_patient'/>
  <NS2:consist_of_subproc rdf:resource='#oe_00103051992'/>
  <NS2:consist_of_subproc rdf:resource='#hp_00117051992'/>
  <NS2:consist_of_subproc rdf:resource='#ma_00127051992'/>
  <NS2:has_overall_impression rdf:resource='#assessment_b5_malignant'/>
  <NS2:has_overall_diagnosis>invasive carcinoma</NS2:has_overall_diagnosis>
</rdf:Description>

<rdf:Description rdf:about='#oe_00103051992'>
  <rdf:type rdf:resource='#Physical_Exam'/>
  <NS2:has_date>03.05.1992</NS2:has_date>
  <NS2:produce_result rdf:resource='#oereport_glp78_1'/>
  <NS2:carried_out_on rdf:resource='#g1p78_patient'/>
</rdf:Description>

<rdf:Description rdf:about='#oereport_glp78_1'>
  <NS2:type rdf:resource='#Lateral_OE_Report'/>
  <NS2:contains_roi rdf:resource='#oe_roi_00103051992'/>
  <NS2:has_lateral rdf:resource='#lateral_left'/>
</rdf:Description>
```

MIAKT Services

- Image Analysis Services
  - Oxford's XRay Mammogram Analyser
  - KCL MRI Mammogram Analyser/Classifier
- Classification Services
  - Abnormality Naïve Bayes Classifier (Soton)
  - MRI Lesion Classifier (KCL)
- Patient Data Retrieval Services (OU)
  - For example, "Find Patients With Same Age"
- Image Registration (KCL)
  - GRID service invoked via web-service
- Natural Language Report Generation (Sheffield)
  - Generate a patient report from RDF description
- UMLS Lookup (Sheffield)
  - Lookup term definitions in the UMLS
- Patient Records also accessed through web-service (Soton)
  - Web-service enabled AKT 3store


MIAKT Demonstrator
Concept Browser
Top
BC_Domain_Top
Medical_Exam
MRI_Contrast_Media
MetaData
Medical_Descriptor
Breast_Disease
Medical_Image
Triple_Assessment_Pr
Lateral_Side
MRI_Process_Descript
Patient
Image_Descriptor
Position_Descriptor
Density_Descriptor
MRI_Static_Enhan_
Mammo_View_Des
Depth_Descriptor
Mammo_Texture_D
Shell_Descriptor
Ass_Tissue_Descri
Distribution_Descri
Morphologic_Descr
Clinician
Region_Of_Interest
Findings
Back
Hide...
Ins
Findings
Medical_Exam
Clinician
Medical_Descriptor
MRI_Contrast_Media
BC_Domain_Top
Lateral_Side
Breast_Disease
MRI_Process_Descriptor
MetaData
Image_Descriptor
is-a
Instances of Patient
Instance 00071_patient
Help
Class Definition
Close this tab
Invoke task:
Find Patients With Same Age
Instance
00071_patient
type Patient
has_age 57
involved_in_ta ta-soton-1070478266177
Invoke task:
Find TripleAssessments with same Lateral View
Instance
ta-soton-1070478266177
type Triple_Assessment_Proc
consist_of_subproc 00071_mammography
involve_patient 00071_patient
Instance
00071_mammography
produce_result image_00071_right_mlo
produce_result image_00071_right_cc
produce_result image_00071_left_mlo
produce_result image_00071_left_cc
type Mammography
carried_out_on 00071_patient
has_date 27 12 1992
Right_Lateral_Image
Right_CC_Image
Left_Lateral_Image
Left_CC_Image
Ready

Demo

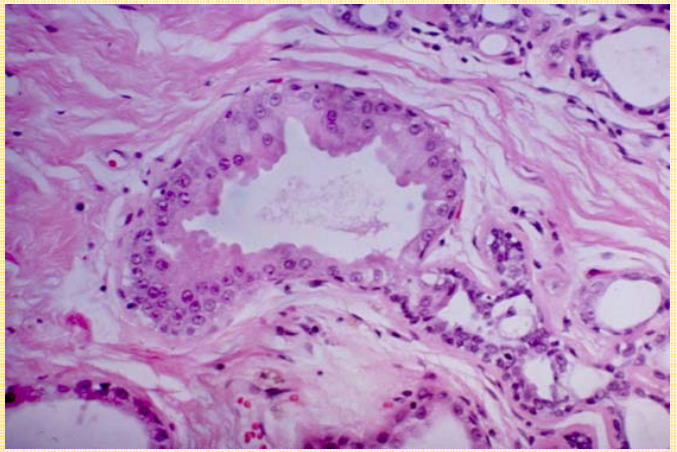
What are the ontological classes in MIAKT?

- After Dasmahapatra and O'Hara 2005
- They are end-products of epistemological and/or decision-making procedures
- One needs to "recognise" instances of a particular class as such
- Information indexed against an ontology can be treated declaratively (Tarski, OWL), but …
- … they come into being procedurally against social and institutional norms

Institutional Norms

- NHS guidelines suggests for identification of apocrine cells (common false positive):

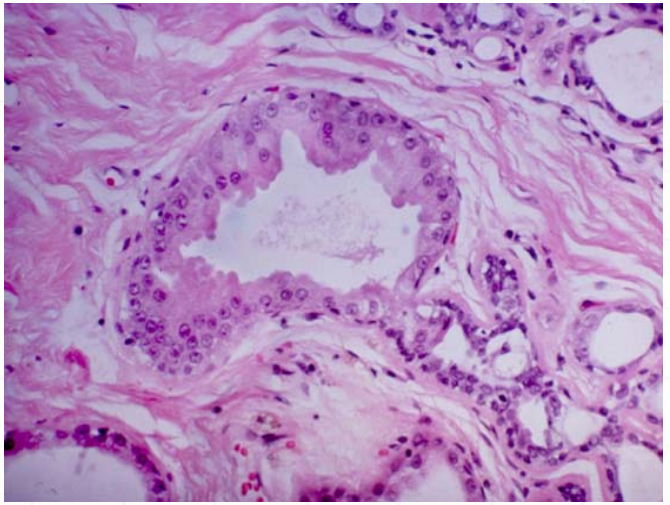
  "Recognition of the dusty blue cytoplasm, with or without cytoplasmic granules with Giemsa stains or pink cytoplasm on Papanicolaou or haematoxylin and eosin stains coupled with a prominent central nucleolus is the key to identifying cells as apocrine."



- Common false-positives in FNAC is misdiagnosis of apocrine cells as malignant condition (pleomorphic appearance signals malignancy; morphological characteristics trad. distinguishing classification criteria for pathologists)
- For KR support, need to record not just the label relevant for diagnosis ("apocrine cells") but also the means by which such a labelling was achieved

Formalised Procedures

$$\forall x \left(\text{fixed-sample}(x) \wedge \exists y_{\bullet}\left[\text{stain}(y_{\bullet}) \wedge \text{stained-with}(x, y_{\bullet}) \longrightarrow \text{Some-Pathological-Concept}(x) \longleftrightarrow \exists z_{\bullet}\left(\text{colour}(z_{\bullet}) \wedge \exists w\left(\text{cytoplasm}(w) \wedge \text{contains}(x, w) \wedge \text{has-colour}(w, z_{\bullet})\right)\right)\right]\right)$$



- For laboratory practice L(x, t) that specimen x is subjected to in context t (time, state variables for exptal/clinical conditions) a predicative attribute P(x) is identified with behavioural response B(x, t) leading to an implicit definition of P(x)

Procedures for Reproducibility

- Specific criteria for identifying histopathological slides as instances of particular lesions – rule following props – make concept labelling reproducible

Standardised Protocol 6 Pathologists		Non-standard Protocol 5 Pathologists	
# Agree	% of cases	# Agree	% of cases
6/6	58	5/5	0
5/6	71	4/5	20
4/6	92	3/5	50

For Ductal Carcinoma in situ, Atypical ductal hyperplasia, procedural criteria reduces inter-expert variability

Criteria of Page et al (Cancer 1982; **49**:751-758; Cancer 1985; **55**:2698-2708), reported by Fechner in MJ Silverstein (1997). Ductal Carcinoma In Situ of the Breast

Norms and Rule-following

- Concept use in medical practice requires the recognition of instances as instances of appropriate classes
- Classes are assigned as proxies of groups of instances to respond in coherent ways to patterns of questioning
- Class ascription needs to be reproducible
- Reproducibility is enhanced by rule-following

So Ironically…

- What was regarded as an implausible philosophical account of ontology (realist) now finds a new embodiment
  - Machines are able to support Tarski semantics
- There is a coming together of a procedural/constructivist account within an apparently traditional formal semantics
- There is a place for a denotational semantics that support ontologies
- But do not expect the meanings to remain stable – they are constructed – they have always been
- Need to understand how meaning will become more richly constructed by our machines and systems in the future

And Finally Requirements on any Ontology Engineering Framework

- Maintenance
  - How to support dynamic evolution
- Viewpoints
  - Mapping within and between perspectives
- Context
  - Design Rationale
- Reuse
  - Disaggregating, modularity, patterns
- Multimedia
  - Annotation and feature extraction
- Rules and procedures
  - Objects/Descriptions & Rules/Procedures

Real ontologists ....

- Real ontologists consider themselves well dressed if their socks match.
- Real ontologists have a non-technical vocabulary of 800 words.
- Real ontologists give you the feeling you're having a conversation with an dial tone.
- Real ontologists wear badges so they don't forget who they are.
- Real ontologists don't find the above at all funny.