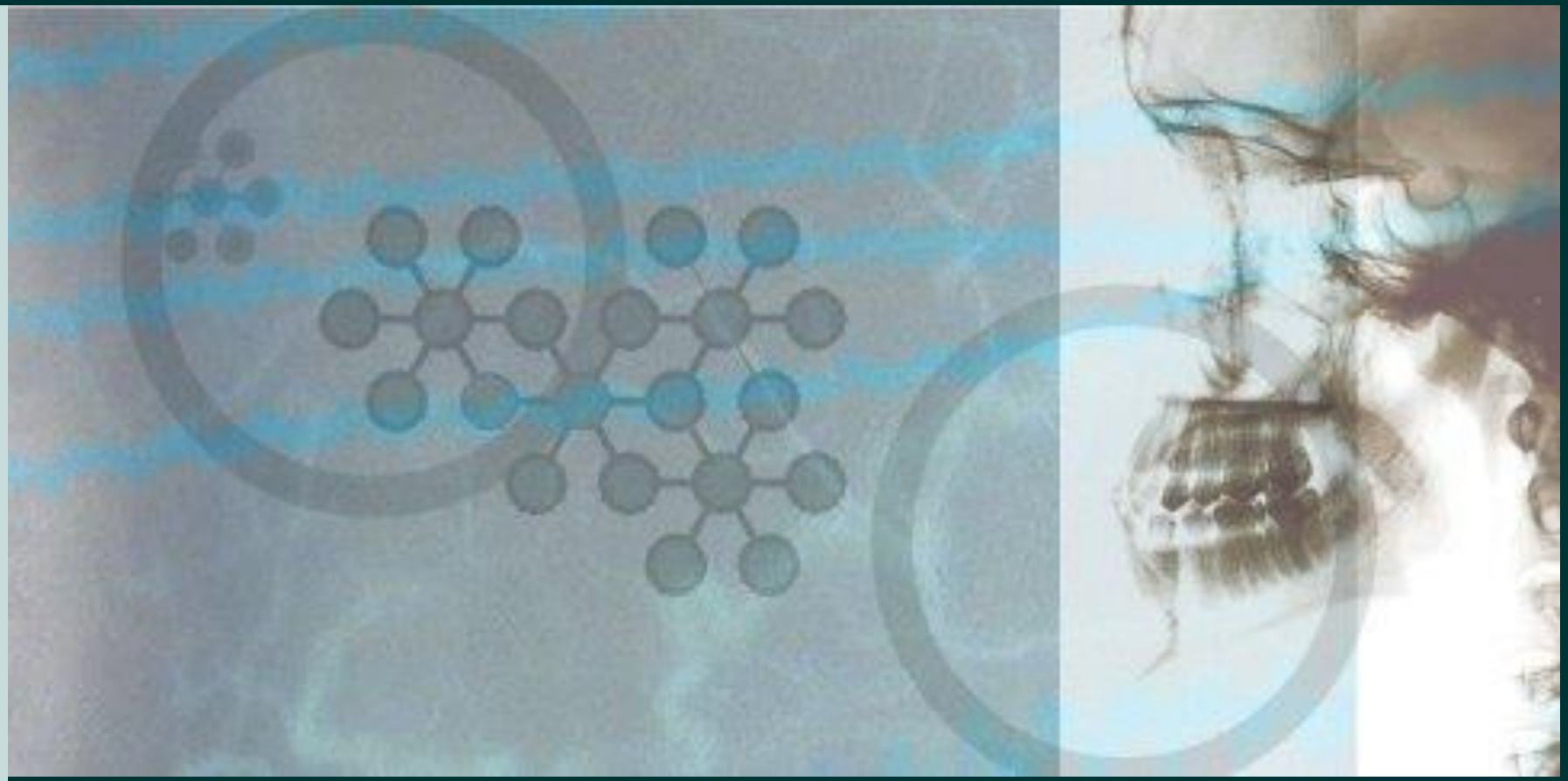
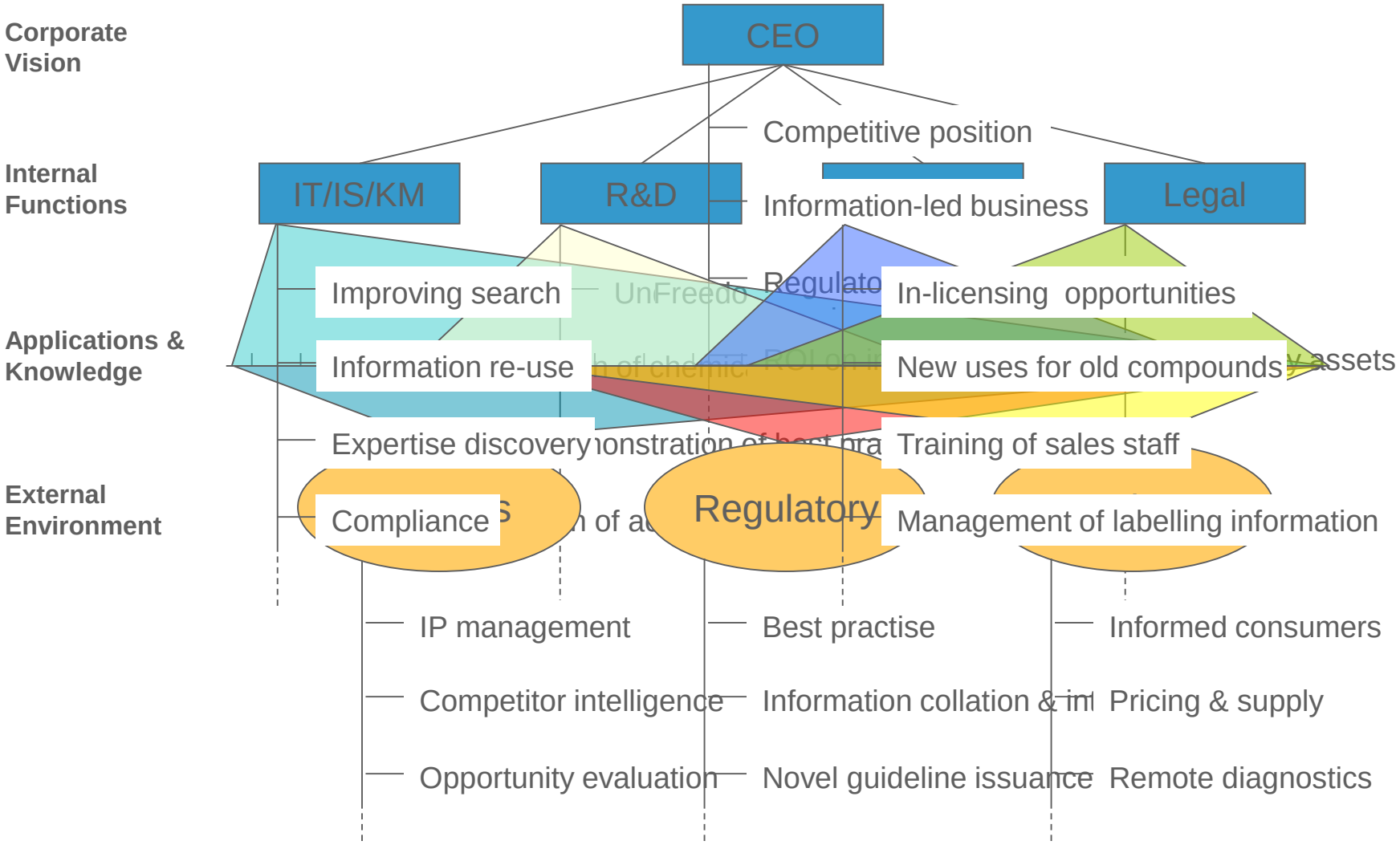
biowisdom

Knowledge Representations

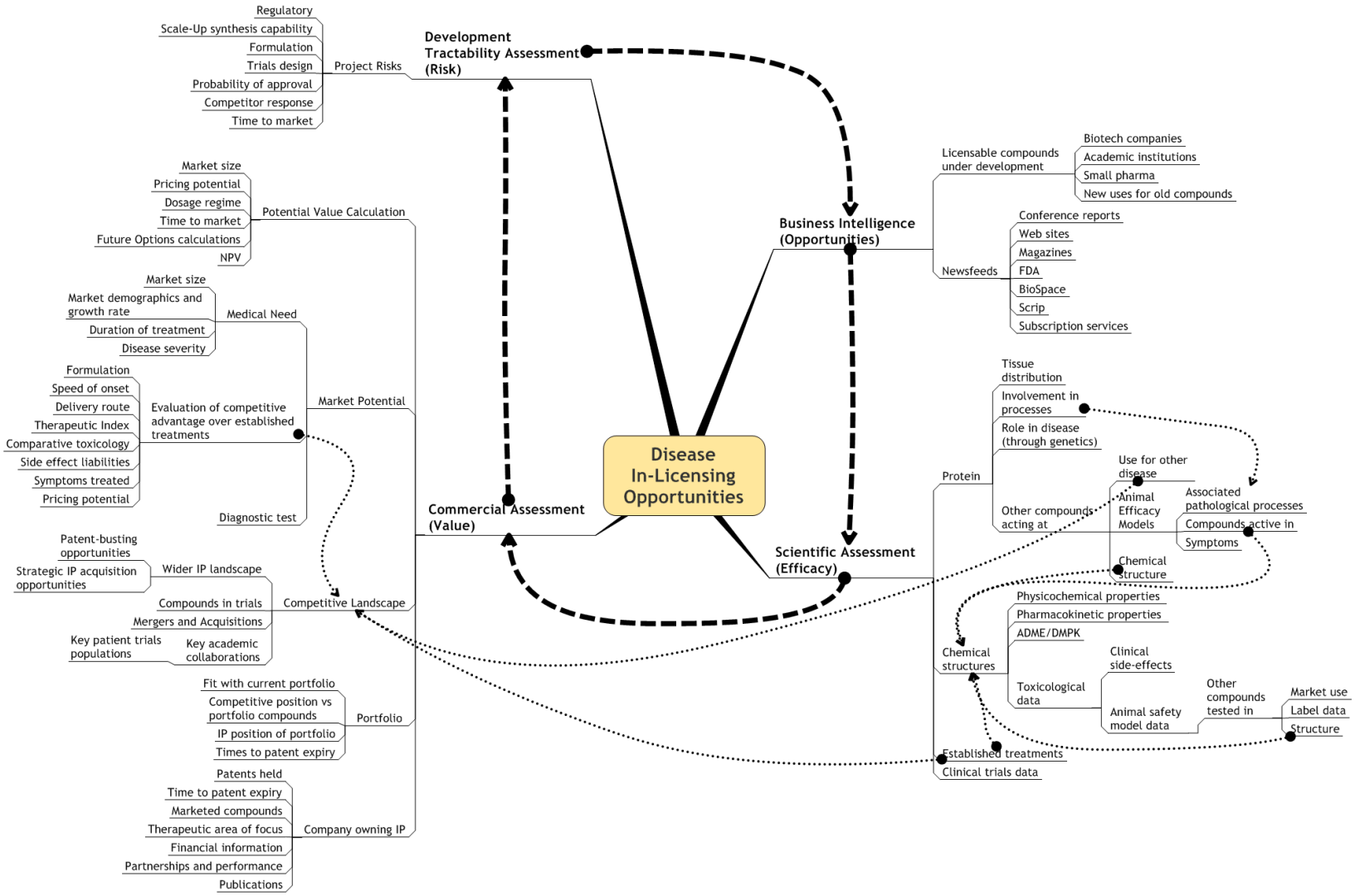
Prepared for PRISM Forum Oct 6th 2004



Integrated Intelligence for Business



In-Licensing Sources



Data Integration Methodologies

- Rules based
  - Matches values in tagged fields
- Data warehousing
  - Specialised database schema developed to optimise repetitive analysis in ‘same question, different data’ applications
- Federated middleware
  - Use of middleware to connect distributed data sources to various client applications via shared data model
- Ad hoc query optimization
  - Query normalisation and distribution across multiple source databases

The Importance of Semantics

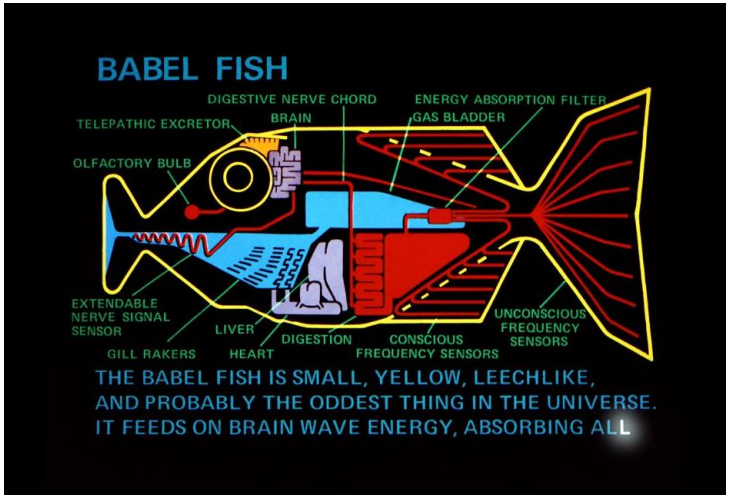
- Identity based semantics are very limiting
  - is-equivalent-of, is-same-as, is-a, is-part-of
- Descriptive relationships are much more valuable

COMPOUND	COMPOUND	AFFECTS
COMPOUND	COMPOUND	CONTAINS
COMPOUND	COMPOUND	HAS AFFINITY FOR
COMPOUND	COMPOUND	HAS DERIVATIVE
COMPOUND	COMPOUND	INCREASES
COMPOUND	COMPOUND	INDUCES
COMPOUND	COMPOUND	INHIBITS
COMPOUND	COMPOUND	INTERACTS WITH
COMPOUND	COMPOUND	IS ACTIVE INGREDIENT IN
COMPOUND	COMPOUND	IS ADMINISTERED WITH
COMPOUND	COMPOUND	IS ANALOGUE OF
COMPOUND	COMPOUND	IS INDUCED BY
COMPOUND	COMPOUND	IS METABOLITE OF
COMPOUND	COMPOUND	REDUCES
mRNA	COMPOUND	IS AFFECTED BY
mRNA	COMPOUND	IS DECREASED BY
mRNA	COMPOUND	IS DOWNREGULATED BY
mRNA	COMPOUND	IS INCREASED BY
mRNA	COMPOUND	IS INDUCED BY
mRNA	COMPOUND	IS INHIBITED BY
mRNA	COMPOUND	IS REGULATED BY
mRNA	COMPOUND	IS UPREGULATED BY
mRNA	PROTEIN	CODES FOR

The Importance of Semantics

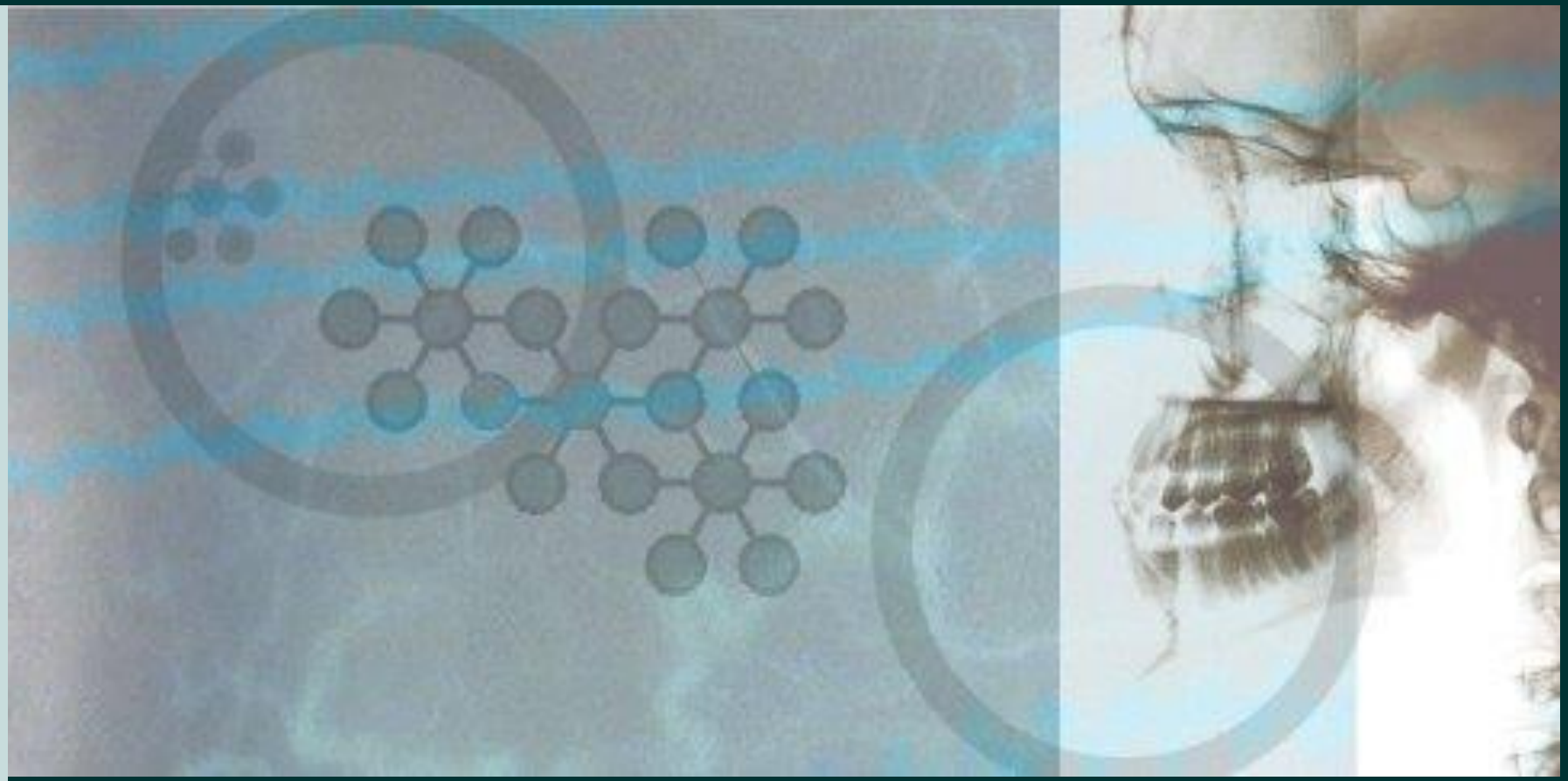
- Semantic Normalization
  - Disambiguation
    - Cold – rhinoviral disease or Chronic Obstructive Lung Disorder
  - Aggregation
    - Diazepam – 197 synonyms

Aliseum; Amiprol An-Ding Ansilive Ansiolin Ansiolisina Antenex Anxicalm Anxionil Apaurin Apo-diazepam Apozepam Armonil Arzepam Assival Atensine Atilen Azedipamin BRN 0754371 Baogin Bensedin Benzopin Best Betapam Bialzepam Britazepam CB 4261 CCRIS 6009 Calmaven Calmocitene Calmociteno Calmod Calmpose Caudel Centrazepam Cercine Ceregulart Chuansuan Condition D-Pam DZP Desconet Desloneg Diacepan Diaceplex Dialag Dialar Diapam Diapax Diapine Diaquel Diastat Diatran Diazem Diazemuls Diazepan Diazepan leo Diazepin Diazetard Dienpax Dipaz Dipezona Disopam Dizac Domalium Doval Drenian Ducene Duksen Dupin Duxen EINECS 207-122-5 Elcion CR Eridan Euphorin P Eurosan Evacalm Faustal Faustan Freudal Frustan Gewacalm Gihitan Gradual Gubex HSDB 3057 Horizon Iazepam Jinpanfan Kabivitrum Kiatrium Kratium Kratium 2 LA III LA-111 Lamra Lembrol Levium Liberetas Lizan Lovium Mandro Mandro-Zep Medipam Mentalium Metamidol Methyl diazepinone Methyldiazepinone Methyldiazepinone Metil Gobanal Morosan NSC 169897 NSC-77518 Nellium Nerozen Nervium Neurolytril Nivalen Nixtensyn Noan Notense Novazam Novodipam Ortopsique Paceum Paralium Paranten Parzam Pax Paxate Paxel Paxum Placidox 10 Placidox 2 Placidox 5 Plidan Pomin Propam Prozepam Psychopax Quetinil Quiatril Quievita Radizepam Relaminal Relanium Relax Reliver Renborin Ro 5-2807 Ruhsitus Saromet Sedipam Seduksen Seduxen Serenack Serenamin Serenzin Setonil Sibazon Sico Relax Simasedan Sipam Solis Sonacon Stesolid Stesolin Tensopam Tranimul Trankinon Tranqdyn Tranquirit Trazepam Umbrium Unisedil Usempax AP Valaxona Valeo Valiquid Valitran Valium Valrelease Valuzepam Vanconin Vatran Vazen Velium Vival Vivol WY-3467 Winii Zepaxid Zipan e-Pam



biowisdom

Background of Knowledge Representation



What do we Mean, Knowledge Representation?

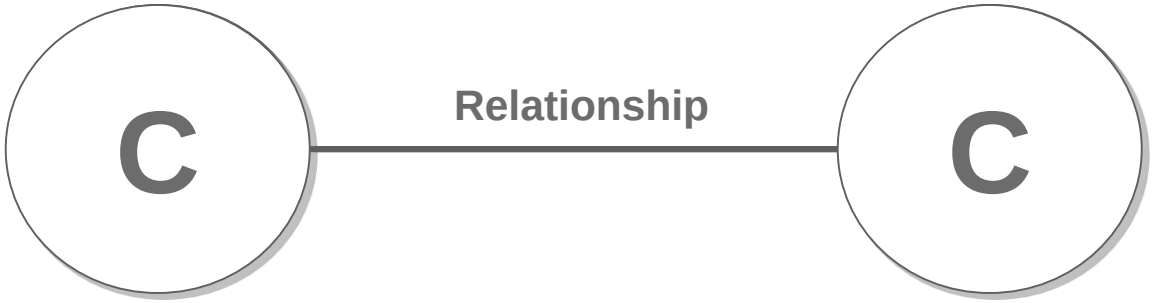
- Based in philosophy, applied in artificial intelligence
- 3 main components:
  1. Logic – provides the formal structures and rules of inference
  2. Ontology – defines the kinds of things that exist in the application domain and the relationships between them
  3. Computation – supports the business applications

Knowledge Representation: Logical, Philosophical and Computational Foundations, John F. Sowa
ISBN 0-534-94965-7

What is Logic?

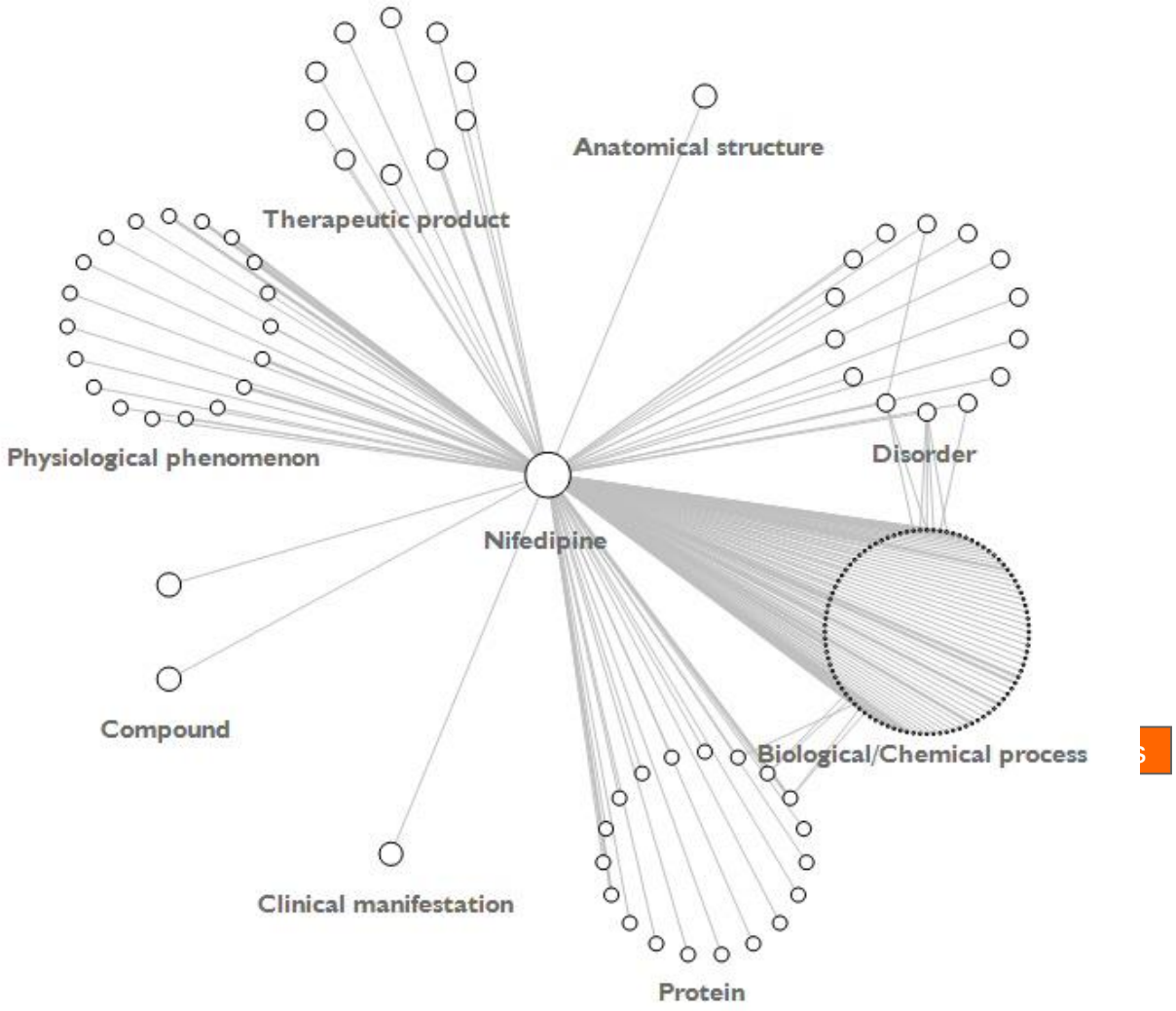
- Aristotle’s syllogisms
- Predicate calculus and conceptual graphs
- Graph theory

Building Blocks of Ontology

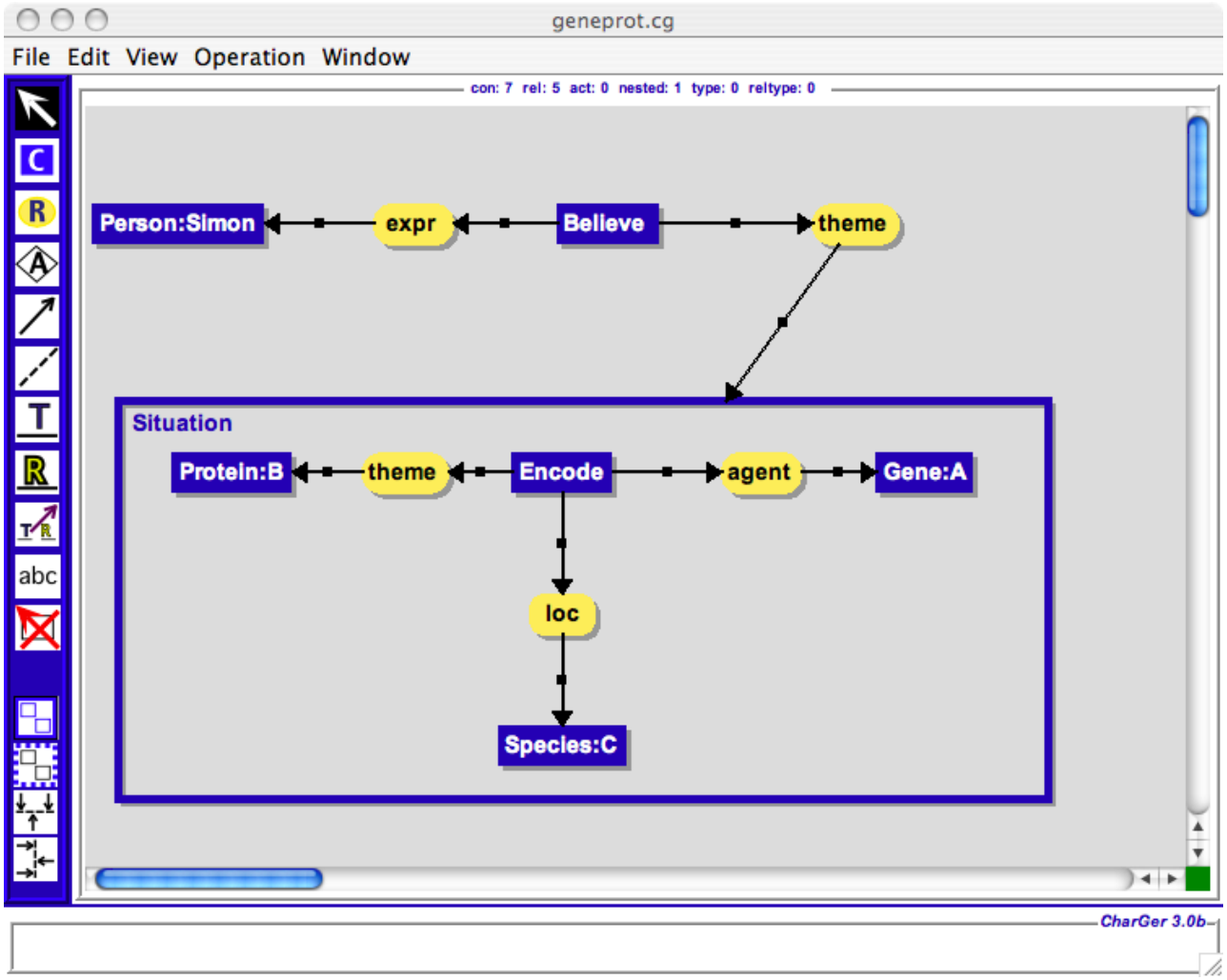


<subject> <predicate>
<object>

Graph Theory Representations



Binary vs n-ary



What is Ontology?

- Quirn's fundamental question of ontology:

  Q: What is there?

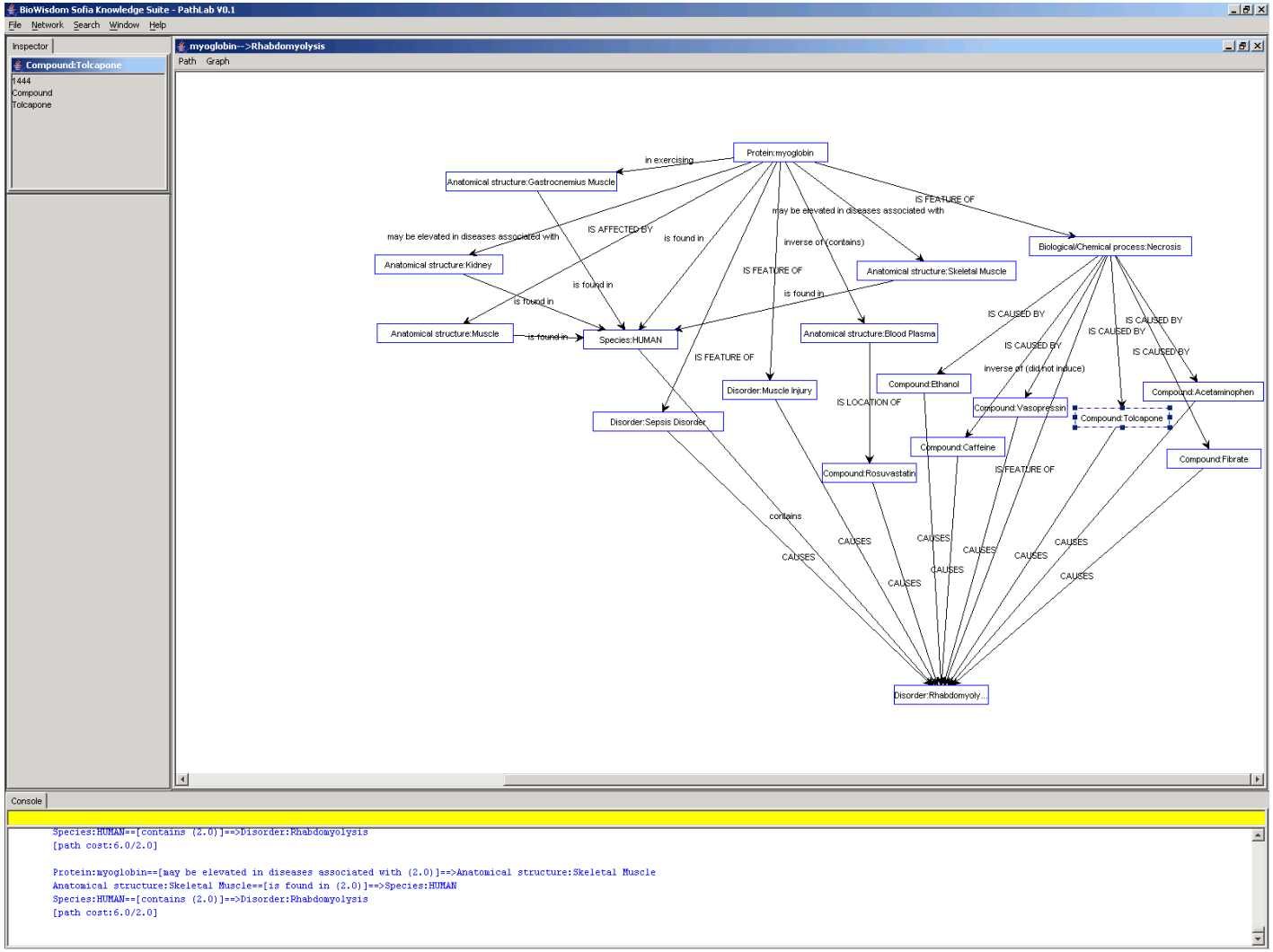
  A: Everything

- The study of 'things' that exist and the relationships that exist between them

What is Computation?

- Reasoning
- Path-finding
- Inference

Path-Finding

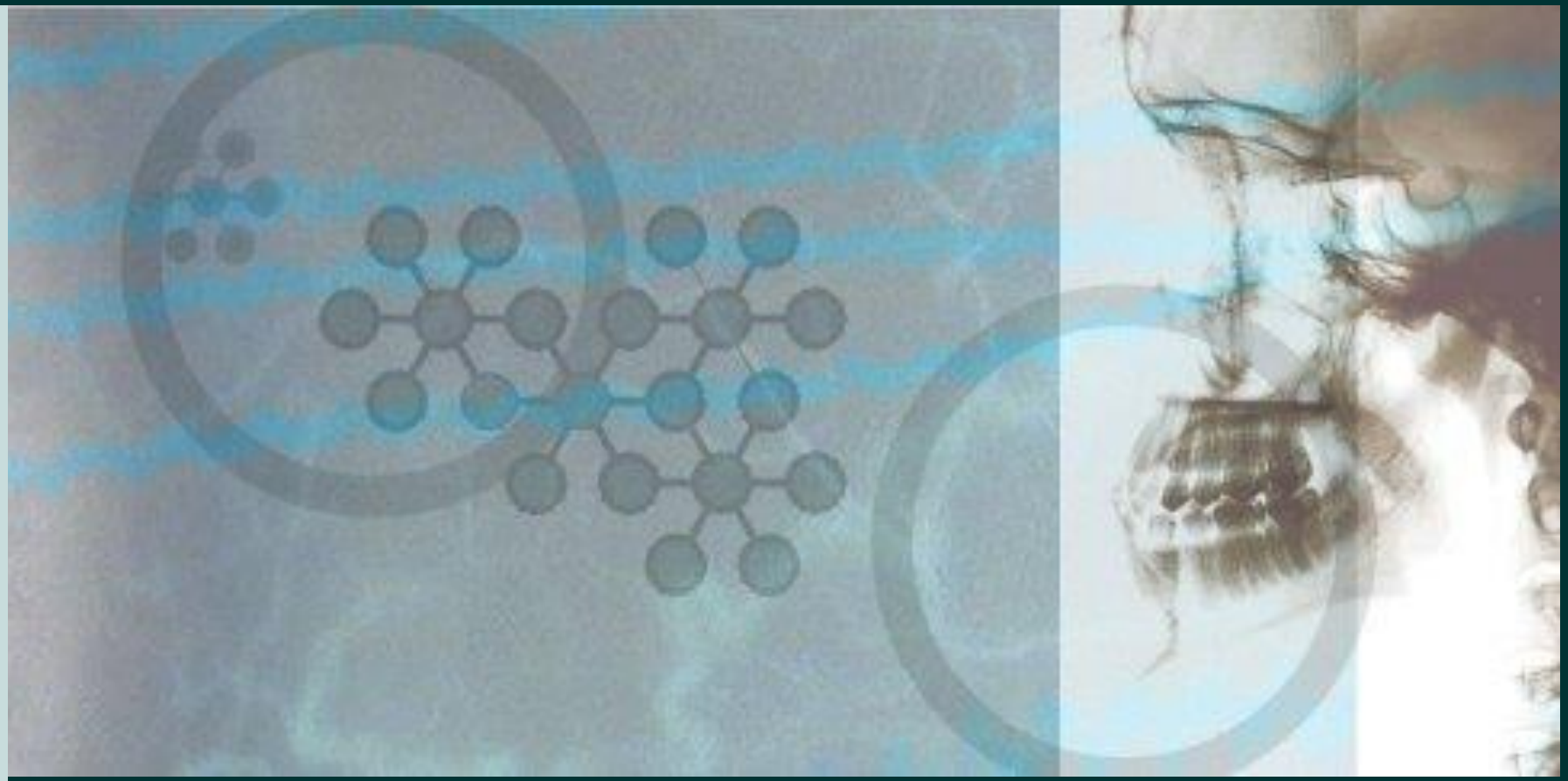


What is Computation In the Real World?

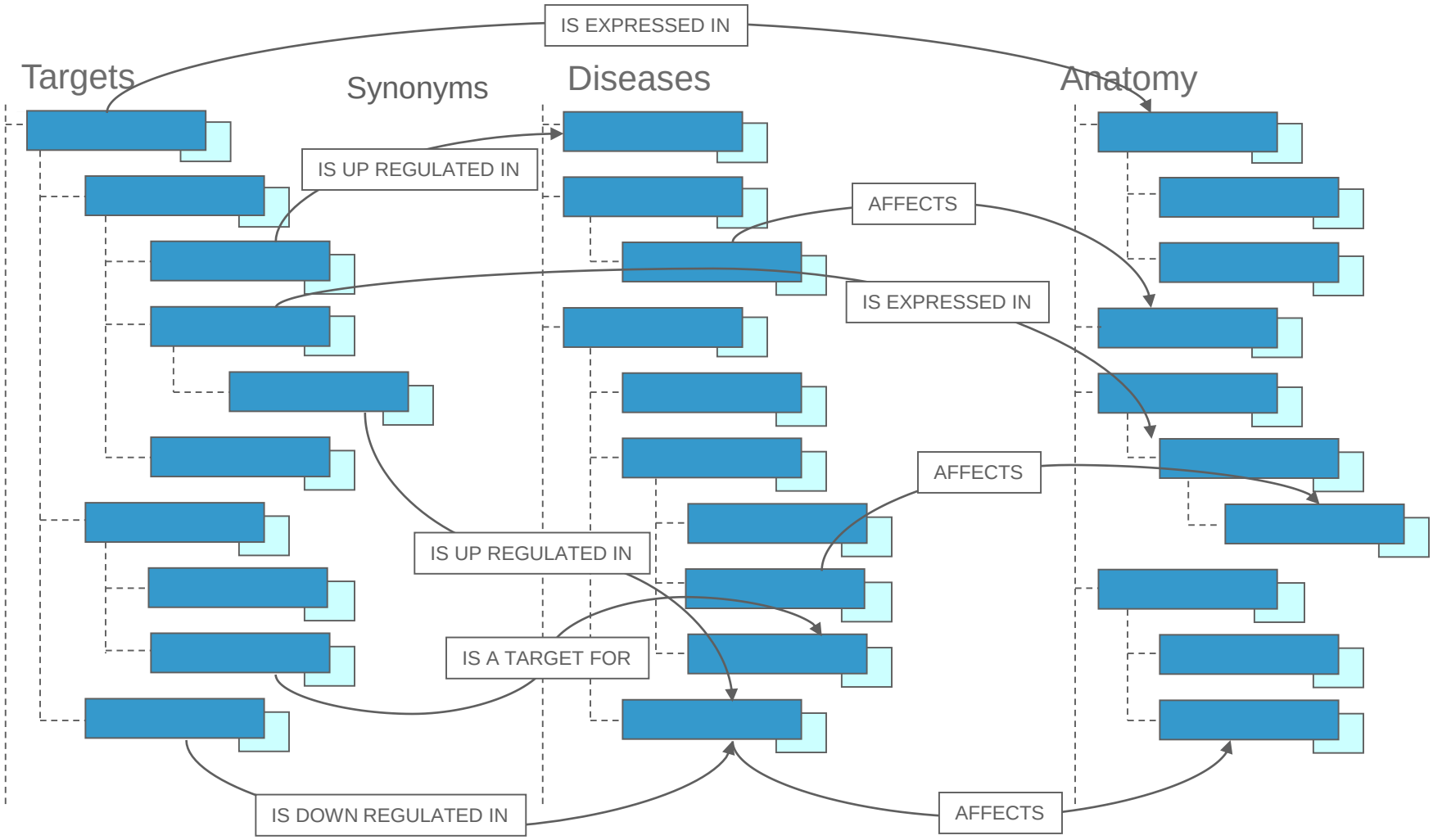
- Hypothesis generation for mechanisms of side effect liability
- Identification of potential biomarkers
- Structure based freedom to operate searches
- Extended high-dimensional SAR analysis using biological and chemical information
- Risk/reward evaluation for in-licensing opportunities
- Information auditing for regulatory compliance
- Smart spell-checker
- Smart phone book with expertise location
- 21st century search

biowisdom

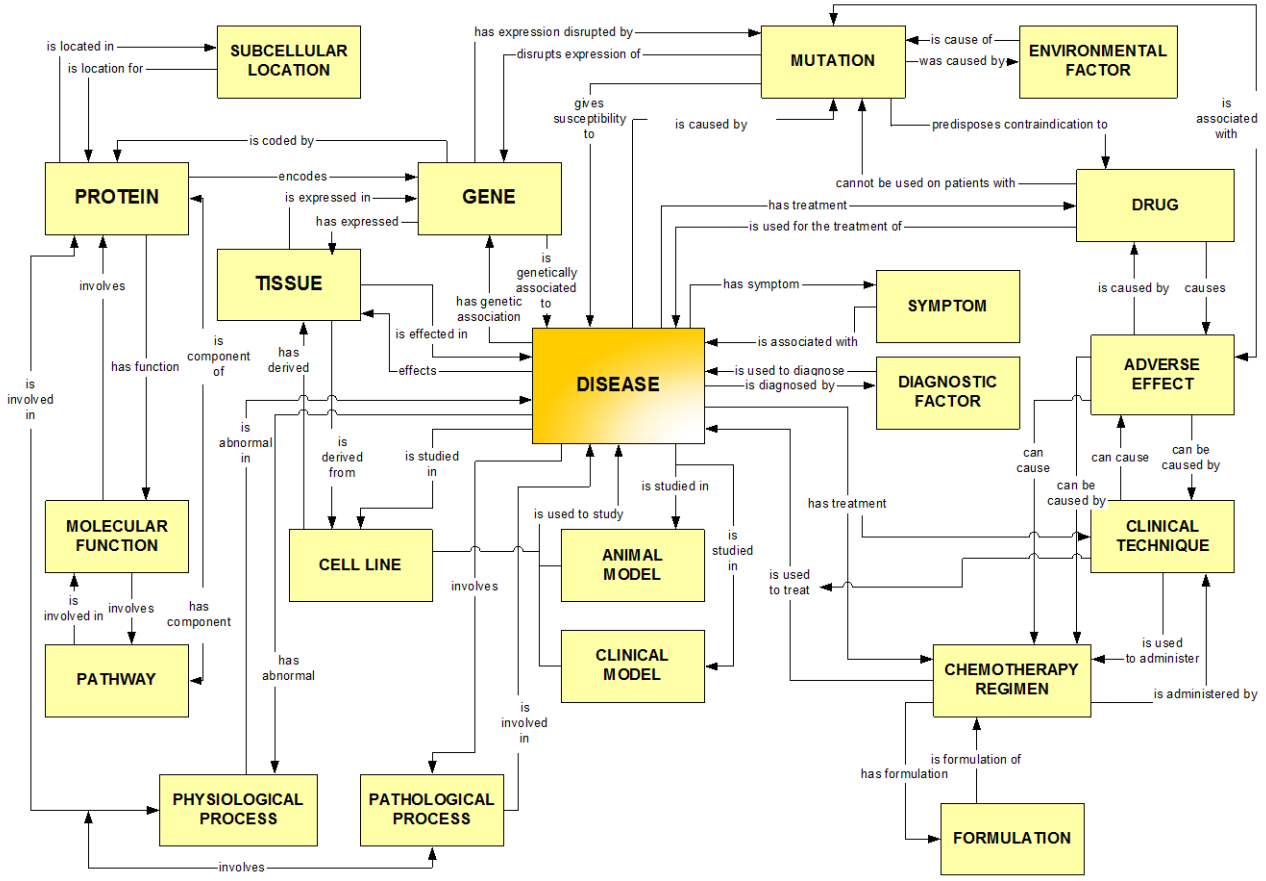
Types of Knowledge Representation



Knowledge Representations - Ontologies

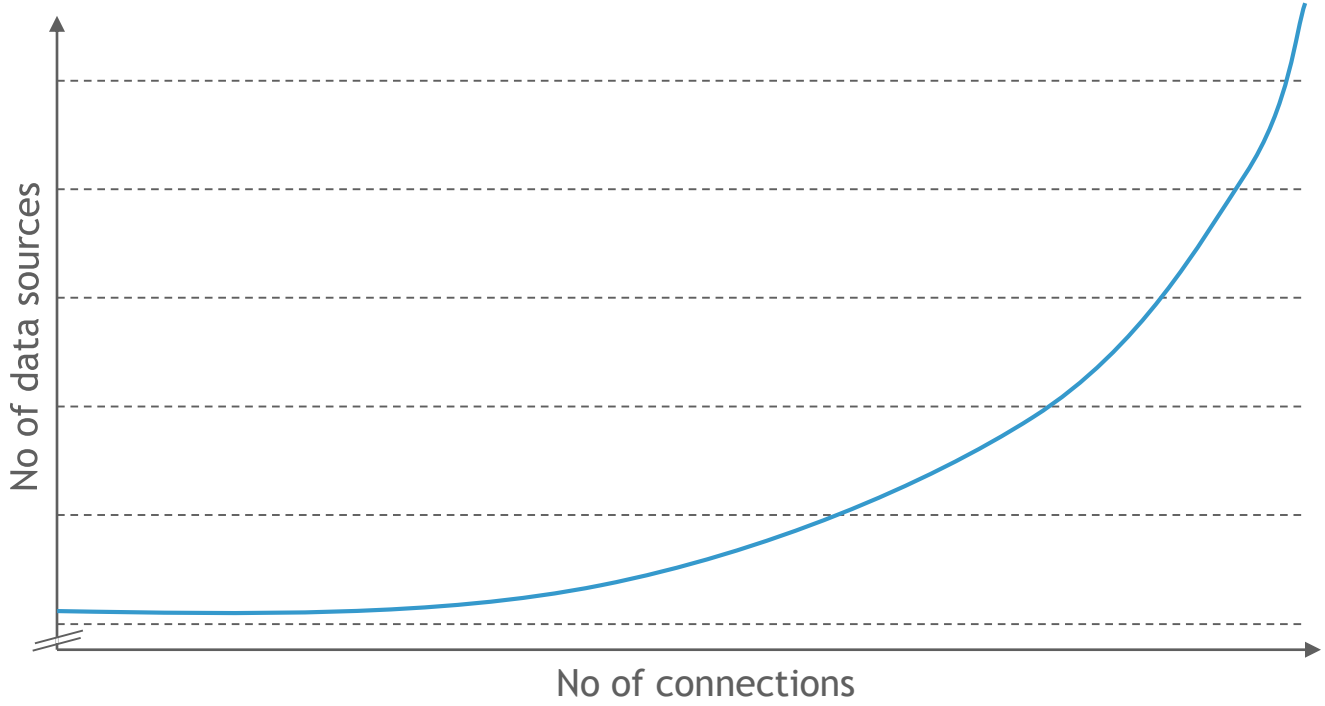


Multi-Relational Ontology



- Integrates information from multiple sources into single coherent view
- Connections are made at a semantic level, not by a common rule

Scalable Multi-Relational Ontology



- Constant level of effort results in an exponential increase in number and complexity of relationships between concepts
- Power of an ontology based system grows as the coverage, content and number of relations grows

Knowledge Representation

Taxonomies

- Manually curated
- Simple parent-child relationships
- Connect single type of concept
- Tend to invisibility
- Become harder to use as they grow
- Become harder to maintain as they grow
- Limited reusability

Ontologies

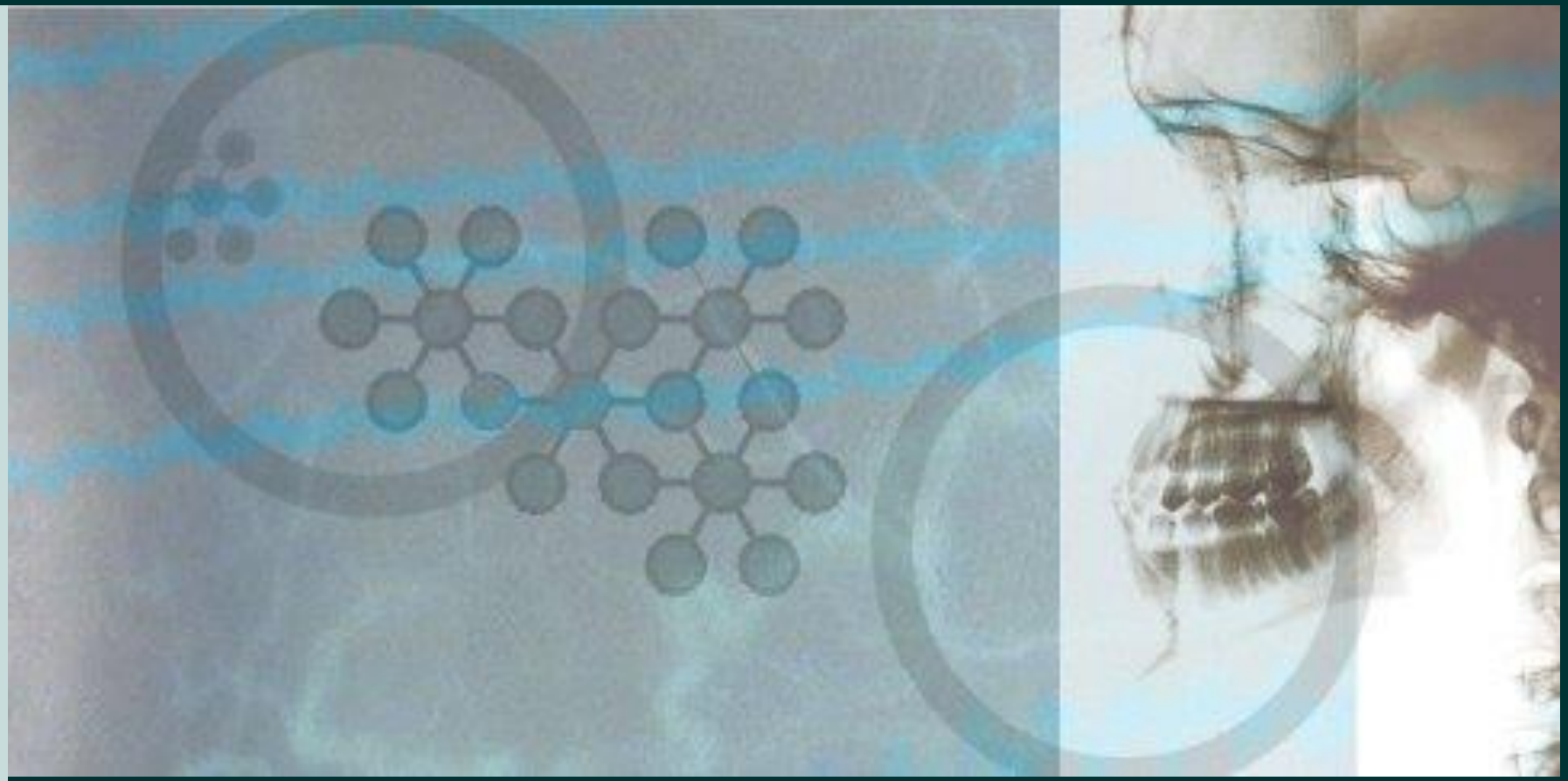
- Semi-automatic curation
- Multiple descriptive relationships
- Connect multiple types of concepts
- Tend to visibility
- Become more valuable and as they grow
- Become easier to maintain as they grow
- Widely reusable

Top-Down vs Bottom-Up

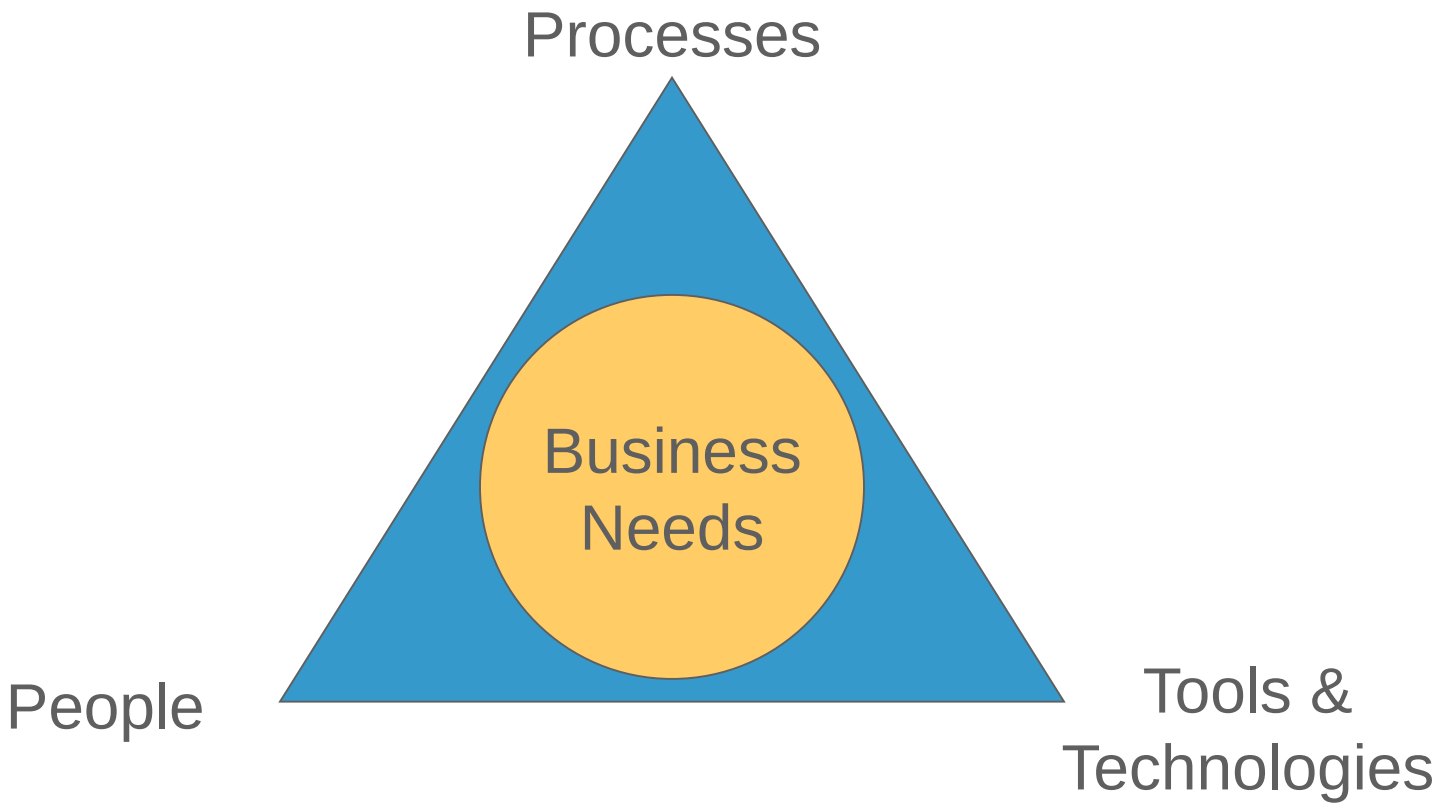
- Top-down approach
  - Segregation into Abstract ¦ Concrete classes
  - Limited relational complexity
  - Manual design and population
  - Guaranteed computability, but limited data
- Bottom-up approach
  - Analyse available data
  - Semi-automated identification of concepts and relationships in data
  - All concepts and relationships structured
  - Potentially limited computability

biowisdom

Applications of Ontology



Background to Knowledge Management

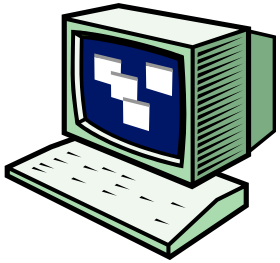


Knowledge Management Pitfalls

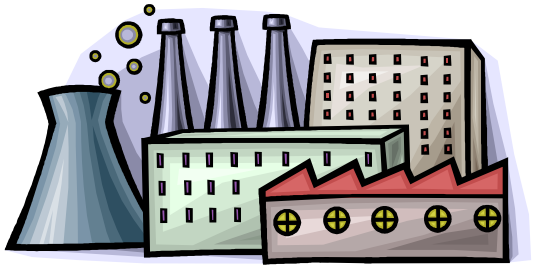
- 60% of KM budgets is spent on high-risk, closed architecture data integration projects
- Lack of business buy-in
  - Often caused by focussing too much on the tools and technologies
  - ‘So what does it mean to me?’
- Too complex a vision means that nothing is delivered until after the business needs have changed
- Poor execution and risk management

4 Pillars of Knowledge Management

People



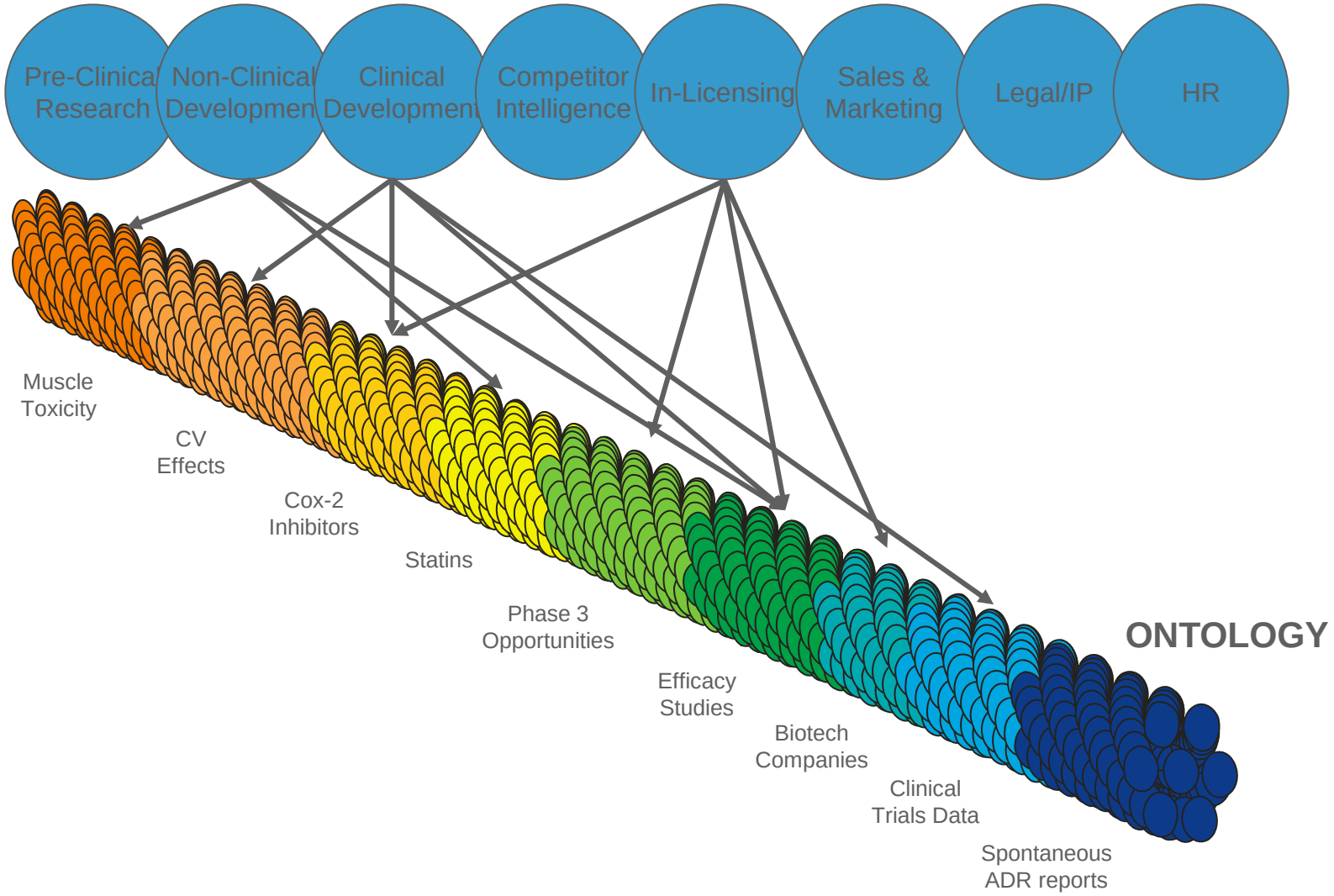
Tools



Organizational Network

Global Network

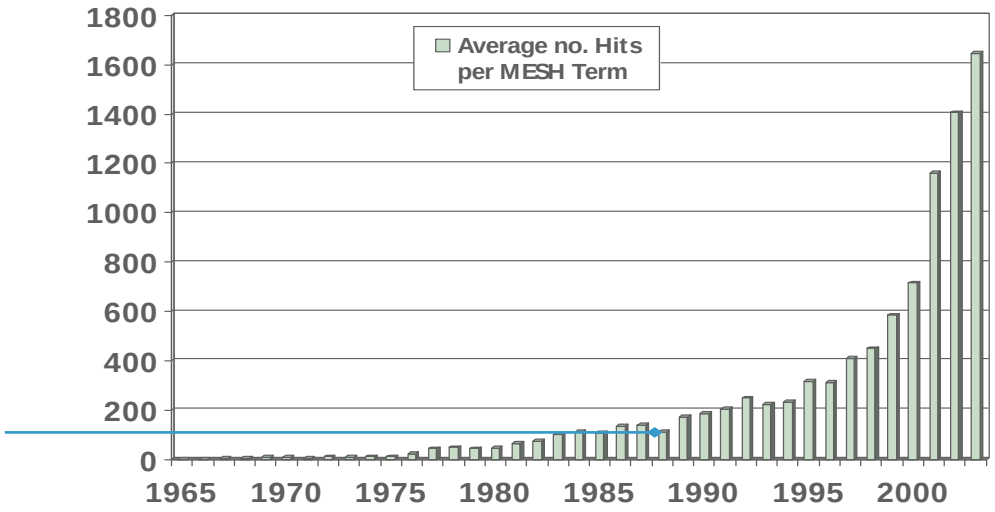
Reusing the Thread of Knowledge



The Tools Don't Work Anymore

- Average scientist or business analyst spends 20-25% of their time looking for information in text sources
- Search recall is only 25-35% as they miss synonyms
- Co-occurrence of terms only works across whole documents
- They get thousands of hits, so they skim the top 100 titles
- They read the top 10 abstracts, and select the top 5 papers

- Chance of reading the 'right' paper is <2%
- Cost to business is $900 per scientist per week *



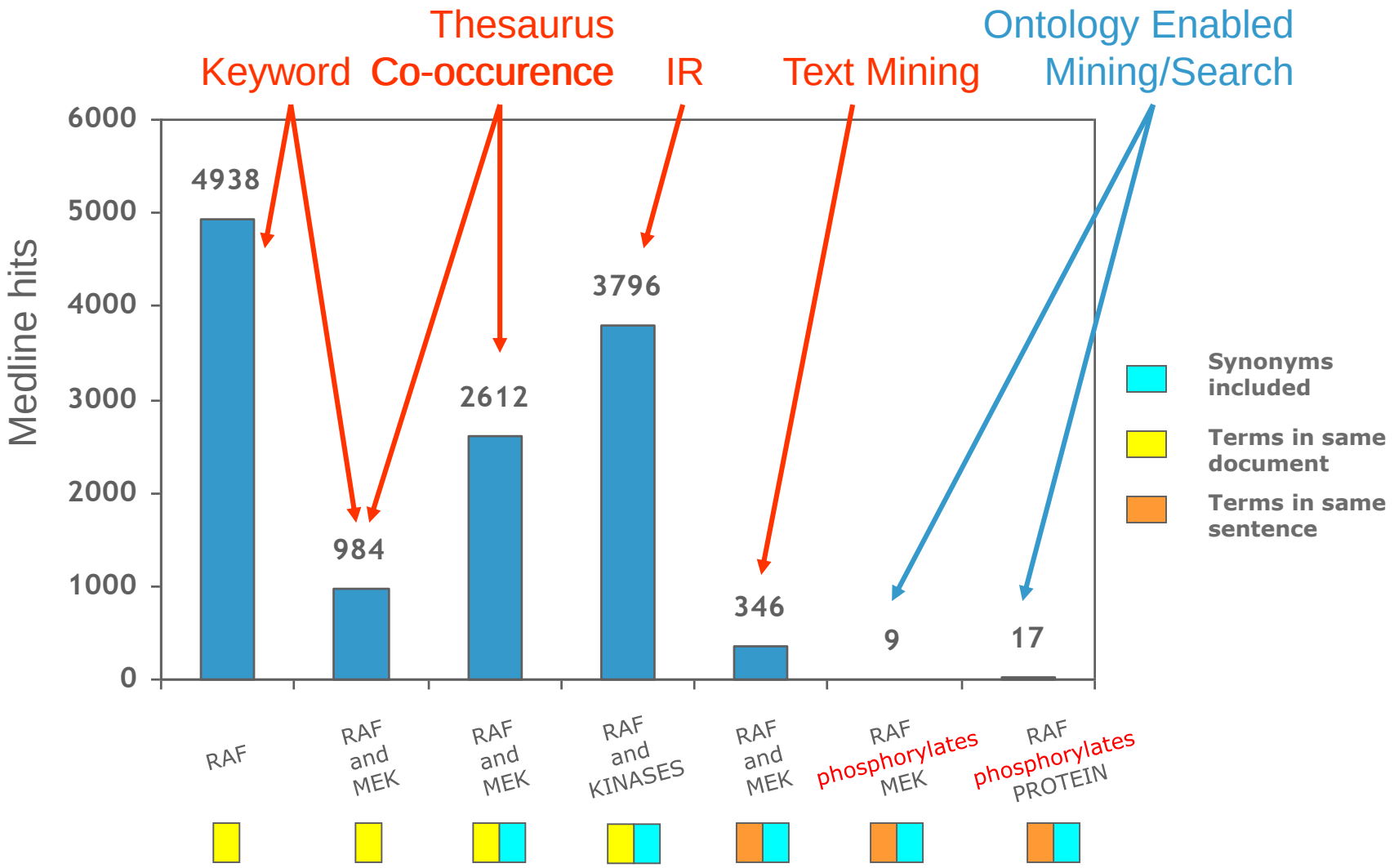
*Based on $200K/yr FTE rate

Progression of Searching

Example Query: ‘RAF phosphorylates MEK’

- PubMed keyword:
  - Articles that contain the word ‘RAF’
- Taxonomy/thesaurus based search:
  - Articles that contain ‘RAF’ or any synonym
- Co-occurrence:
  - Articles that contain both ‘RAF’ and ‘MEK’ (or any synonym)
- Information Retrieval (Verity, Convera, Inxight etc.):
  - Articles that are about ‘RAF’ and other kinases
- Text Mining (ClearForest, Inxight, I2E etc.):
  - Articles that contain the concepts ‘RAF’ and ‘MEK’ (or synonym) linguistically bounded in phrase, sentence or section with verb
- Thematic (Ontology):
  - Articles that contain references to ‘RAF phosphorylating MEK’ or any concept/relationship synonym
  - All other things that ‘RAF’ (or its synonyms) interacts with grouped by type or relationship

Ontology Improves Search Accuracy

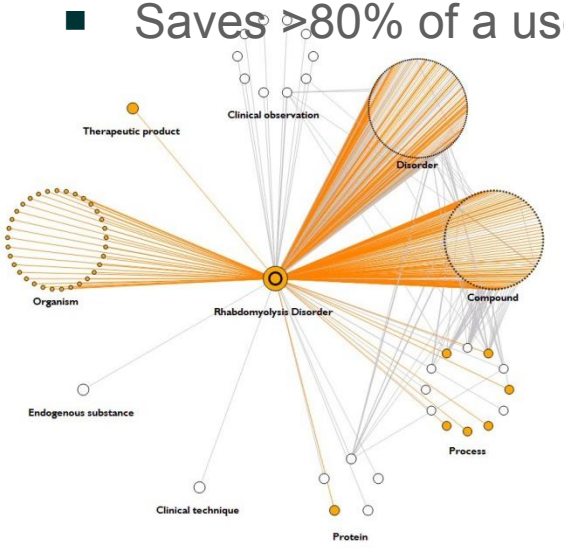


RAF - phosphorylates - TARGET

v-raf-1 murine leukemia viral oncogene 1	phosphorylates*/vg	target	Sentence	Link	Score
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	efficiently phosphorylate	mitogen-activated protein kinase kinase *(MAP kinase kinase)*	B-Raf and Raf-1 both efficiently phosphorylate MAP kinase kinase (MEK-1).	doc10741	99
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	phosphorylated	mitogen-activated protein kinase kinase 1 *(MKK1)*	Raf-1 phosphorylated MKK1 on one major tryptic phosphopeptide, the phosphorylation of which increased with time.	doc10965	100
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	was extensively phosphorylated	camp-dependent protein kinase *(PKA)*	in vitro phosphorylation experiments showed that Raf-1 was extensively phosphorylated by PKA, while ERK2 and MEK were not.	doc17924	99
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	phosphorylates	mitogen-activated protein kinase kinase *(mitogen-activated protein kinase kinase)*	upon activation, Raf-1 phosphorylates mitogen-activated protein kinase kinase (MEK), which in turn activates mitogen-activated protein kinase/extracellular signal-regulated kinases (MAPK/ERKs), leading to the propagation of signals.	doc22168	100
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	phosphorylate	mitogen-activated protein kinase kinase 1 *(MKK1)*	this could be partly explained by the inability of Raf-1 to phosphorylate MKK1 C-terminal deletion mutants even though the phosphorylation sites were intact in these mutants.	doc24780	99
v-raf-1 murine leukemia viral oncogene 1 *(both Raf-1)*	are phosphorylated	p21 activated kinase 1 *(PAK1)*	here we show that both Raf-1 and MEK1 are phosphorylated by PAK1 and that mutations at PAK1 phosphorylation sites in either protein prevent cross-cascade activation.	doc26770	97
		protein kinase c-like 1 *(PAK1)*	here we show that both Raf-1 and MEK1 are phosphorylated by PAK1 and that mutations at PAK1 phosphorylation sites in either protein prevent cross-cascade activation.	doc26770	97
v-raf-1 murine leukemia viral oncogene 1 *(Raf-1)*	phosphorylate	mitogen-activated protein kinase kinase 1 *(MEK1)*	phosphorylation of MEK1 on serine 298 does not appear to regulate the interaction between Raf-1 and MEK1, but rather the ability of Raf-1 to phosphorylate MEK1 with which it is complexed in vivo.	doc26770	99

Finding Information Effectively Using Ontology

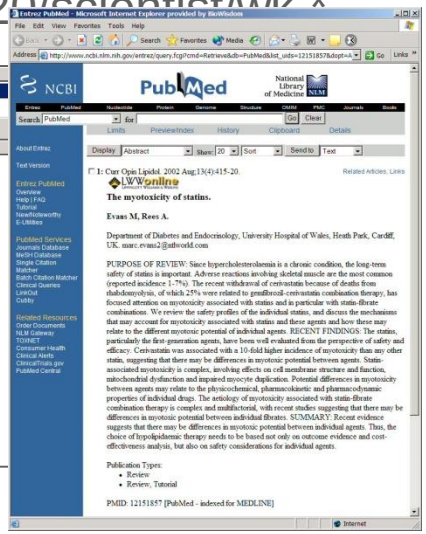
- Text resources have been mined for all concepts and relationships
- Recall is >90% as synonyms are automatically appended to the search
- User can choose the themes and topics that they wish to see
- Precision is >90% for the specific relationship between the terms
- Users get presented with an overview of the contexts in which their concept occurs, and the best papers connecting multiple concepts
- Saves >80% of a user's search time - $720/scientist/wk *



Name	Type	Relationship	Name
Rhabdomyolysis	Disorder	IS CAUSED BY	Cerivastatin
Acute Rhabdomyolysis	Disorder	IS CAUSED BY	Cerivastatin

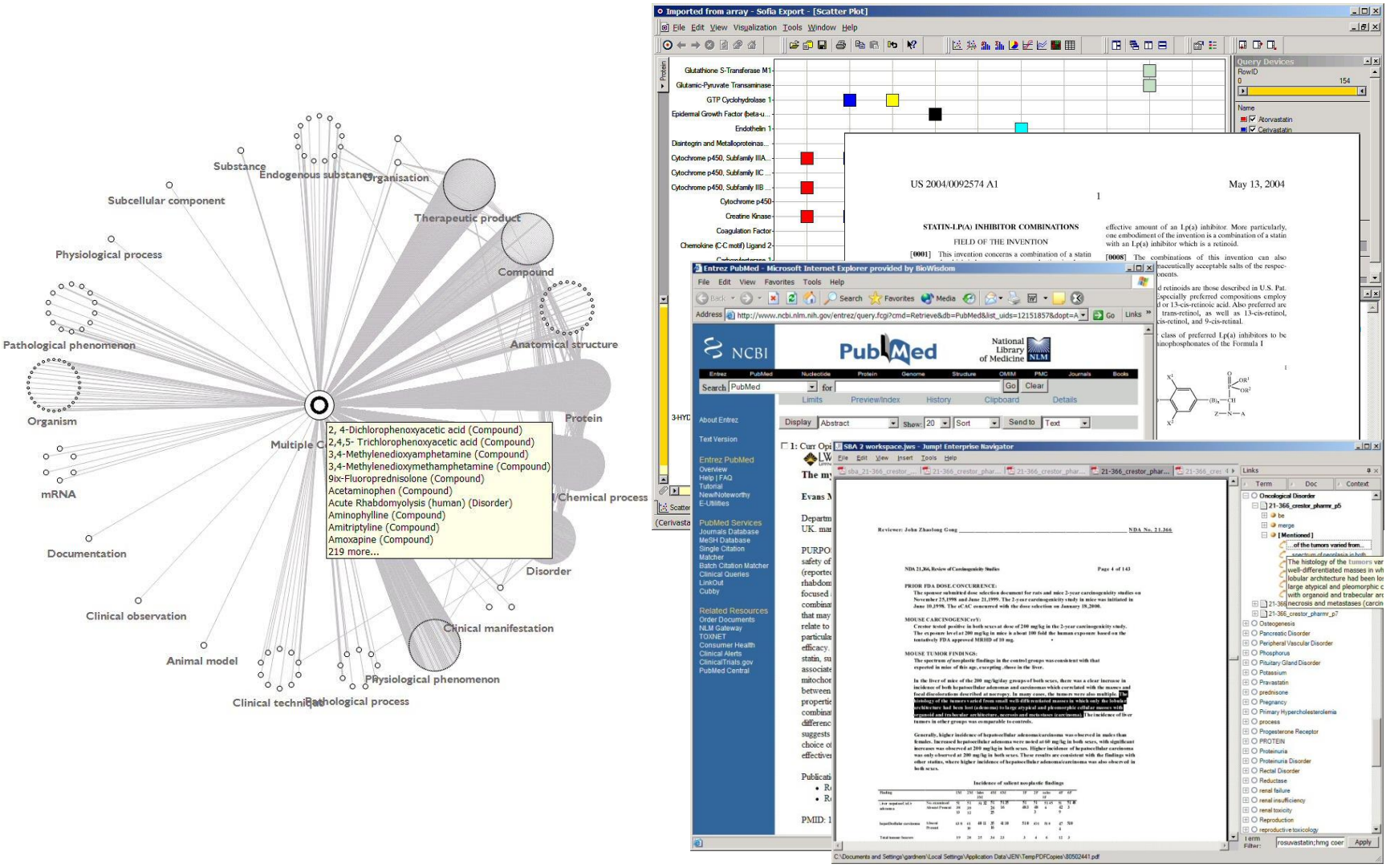
Assertion Evidence

Data Source	Version	Record
Disease Database	1	
MEDLINE	1	12151857
MEDLINE	1	12372976
MEDLINE	1	12496749
MEDLINE	1	11836874
MEDLINE	1	12093269
MEDLINE	1	11602127
MEDLINE	1	12596018
MEDLINE	1	12426921
MEDLINE	1	11129127
MEDLINE	1	11417566
MEDLINE	1	11371708
MEDLINE	1	11773716
MEDLINE	1	12066968
MEDLINE	1	12058707
MEDLINE	1	12030671
MEDLINE	1	11856903
MEDLINE	1	11718617
MEDLINE	1	11562335
MEDLINE	1	11502997
MEDLINE	1	11718614
MEDLINE	1	11905593
MEDLINE	1	11939690
SBA	1	081102en.pdf



*Based on $200K/yr FTE rate

Searches Lead to More Relevant Knowledge



biowisdom

Systematic Knowledge Analysis

What is the mechanism of toxicity associated with a class of drugs?

Identify all known side effects

- Identified forms of side effect (42)
- Search PubMed extract clinical subset
- 500,000 papers Read 100/day 20 man years

Extract all compounds

- Read abstracts
- Extract all references to compounds (140)

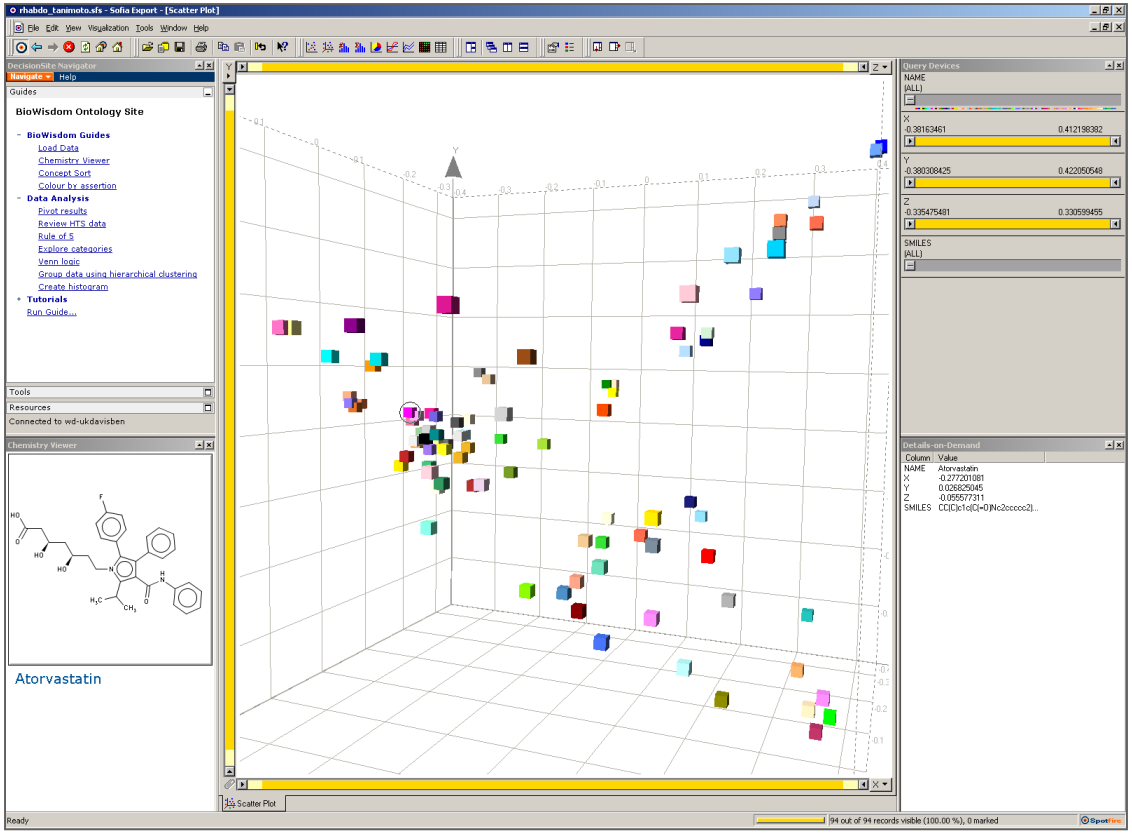
Identify all proteins

- Search Pubmed for protein interactions for each of 140 cmpds
- 3,000,000 papers Read 100/day 120 man years
- Proteins (500)

Manual, systematic aggregation of all the knowledge to enable comparative analysis is not tractable

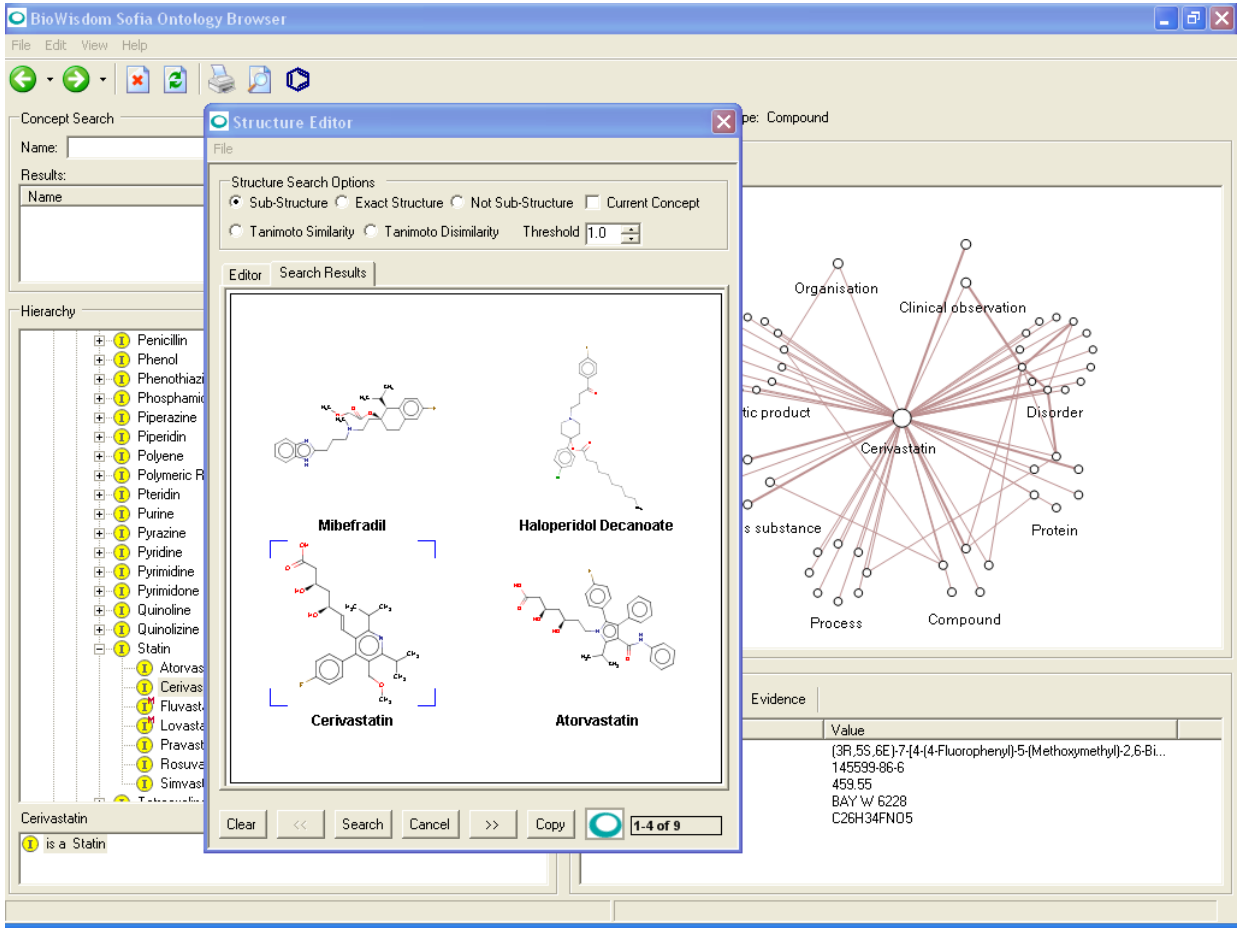
Analysis and Data Mining

- Aggregates relevant information from many sources
- Exported for analysis in data mining tools of choice, e.g. Spotfire

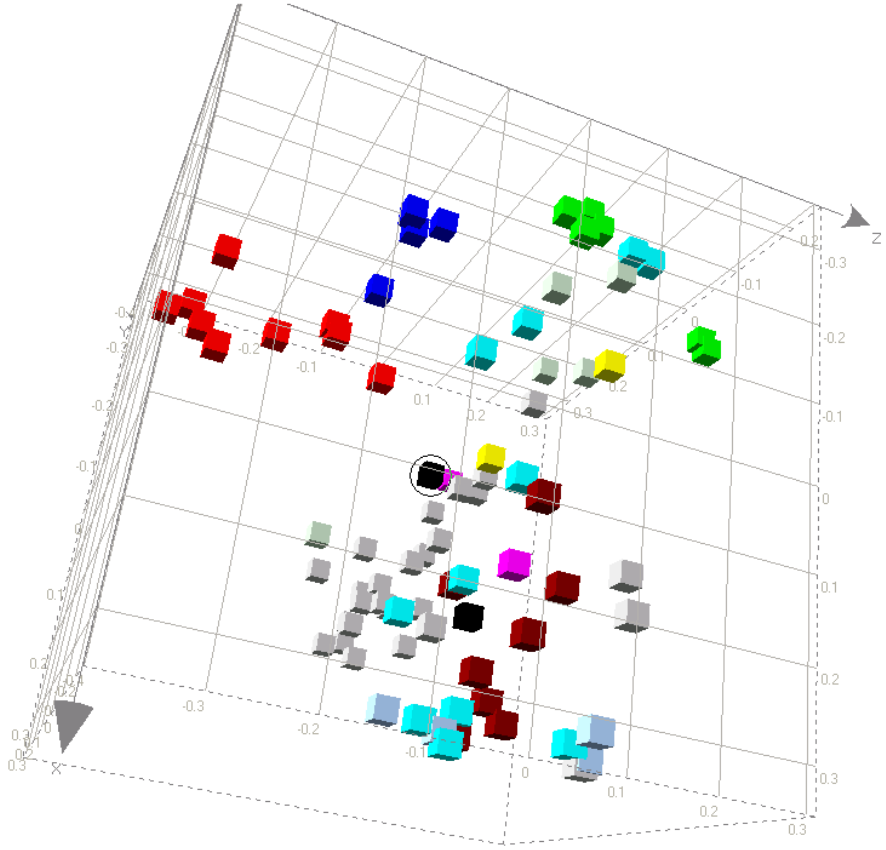


Linking Structure to Function

for Medicinal Chemists/Toxicologists

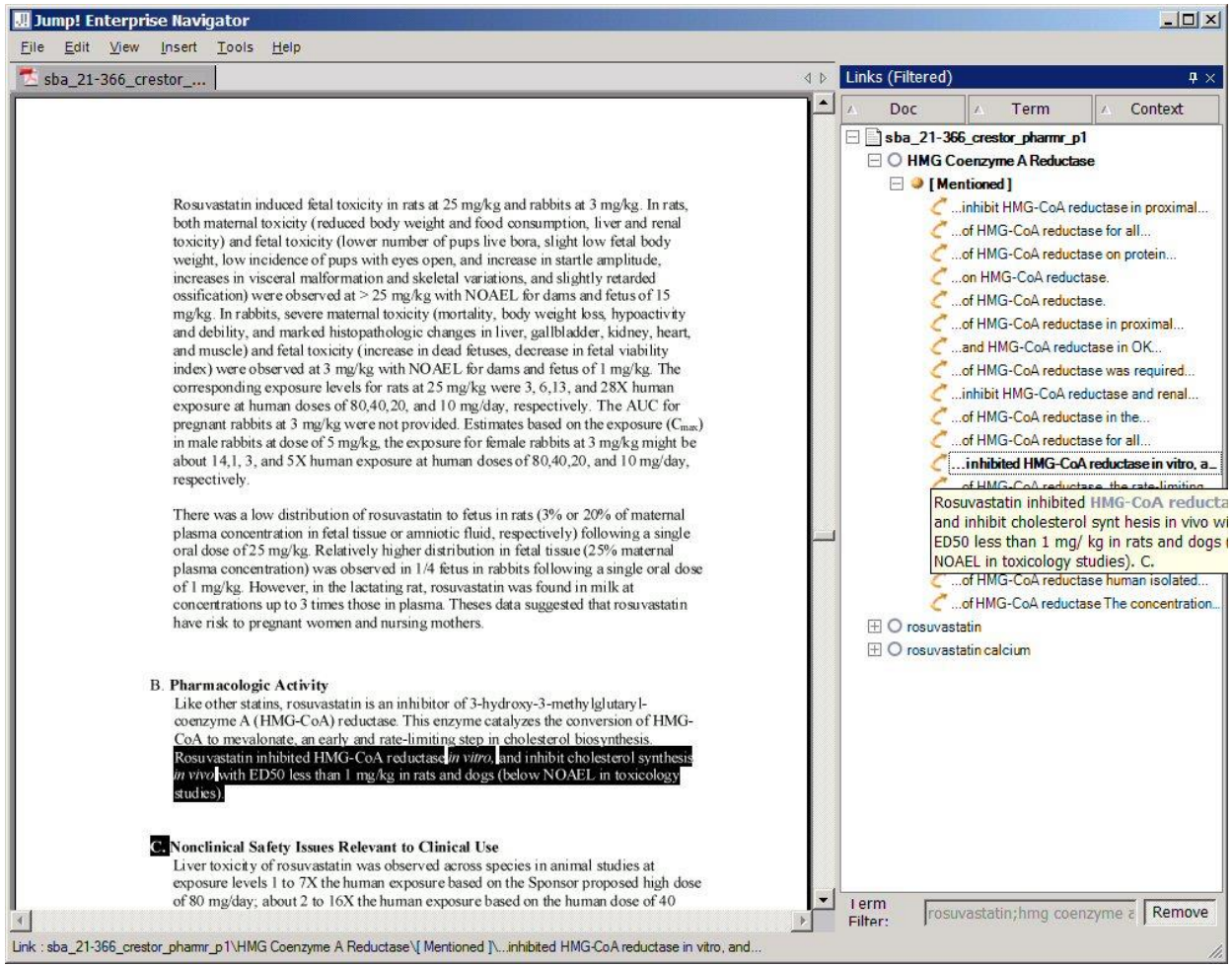


Extended SAR using Biological + Chemical Data

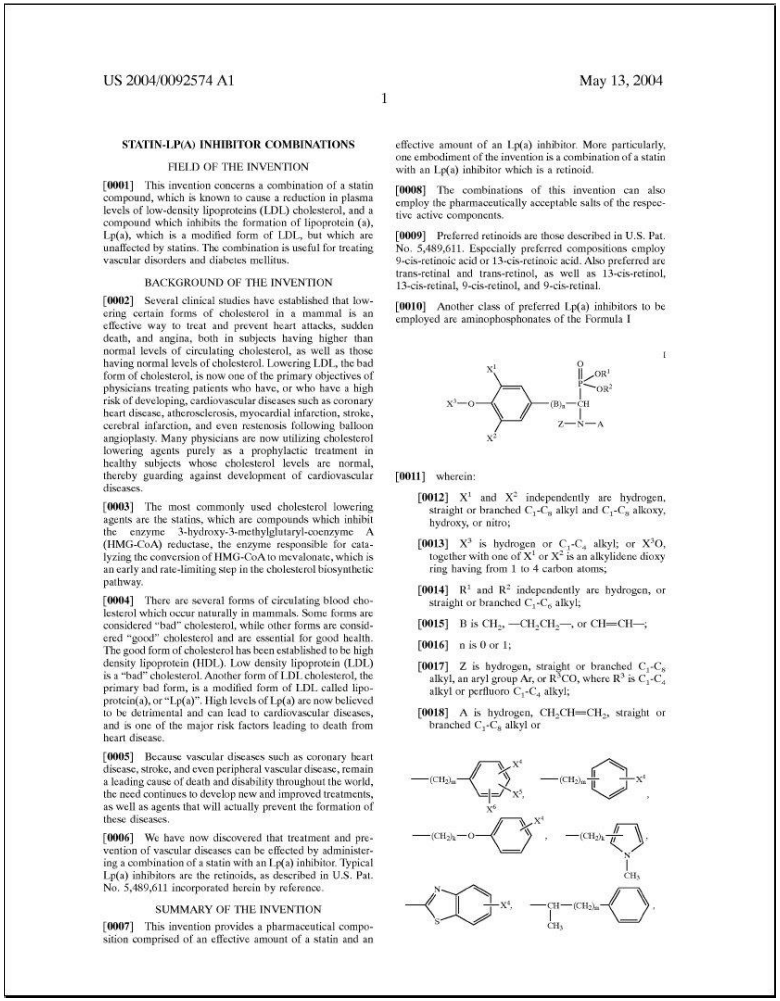


Speeding up Analysis of FDA Documents

for Regulatory Scientists



Freedom to Operate



US 2004/0092574 A1 May 13, 2004
1

STATIN-LP(A) INHIBITOR COMBINATIONS

FIELD OF THE INVENTION

[0001] This invention concerns a combination of a statin compound, which is known to cause a reduction in plasma levels of low-density lipoproteins (LDL) cholesterol, and a compound which inhibits the formation of lipoprotein (a), Lp(a), which is a modified form of LDL, but which are unaffected by statins. The combination is useful for treating vascular disorders and diabetes mellitus.

BACKGROUND OF THE INVENTION

[0002] Several clinical studies have established that lowering certain forms of cholesterol in a mammal is an effective way to treat and prevent heart attacks, sudden death, and angina, both in subjects having higher than normal levels of circulating cholesterol, as well as those having normal levels of cholesterol. Lowering LDL, the bad form of cholesterol, is now one of the primary objectives of physicians treating patients who have, or who have a high risk of developing, cardiovascular diseases such as coronary heart disease, atherosclerosis, myocardial infarction, stroke, cerebral infarction, and even restenosis following balloon angioplasty. Many physicians are now utilizing cholesterol lowering agents purely as a prophylactic treatment in healthy subjects whose cholesterol levels are normal, thereby guarding against development of cardiovascular diseases.

[0003] The most commonly used cholesterol lowering agents are the statins, which are compounds which inhibit the enzyme 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the enzyme responsible for catalyzing the conversion of HMG-CoA to mevalonate, which is an early and rate-limiting step in the cholesterol biosynthetic pathway.

[0004] There are several forms of circulating blood cholesterol which occur naturally in mammals. Some forms are considered "bad" cholesterol, while other forms are considered "good" cholesterol and are essential for good health. The good form of cholesterol has been established to be high density lipoprotein (HDL). Low density lipoprotein (LDL) is a "bad" cholesterol. Another form of LDL cholesterol, the primary bad form, is a modified form of LDL called lipoprotein(a), or "Lp(a)". High levels of Lp(a) are now believed to be detrimental and can lead to cardiovascular diseases, and is one of the major risk factors leading to death from heart disease.

[0005] Because vascular diseases such as coronary heart disease, stroke, and even peripheral vascular disease, remain a leading cause of death and disability throughout the world, the need continues to develop new and improved treatments, as well as agents that will actually prevent the formation of these diseases.

[0006] We have now discovered that treatment and prevention of vascular diseases can be effected by administering a combination of a statin with an Lp(a) inhibitor. Typical Lp(a) inhibitors are the retinoids, as described in U.S. Pat. No. 5,489,611 incorporated herein by reference.

SUMMARY OF THE INVENTION

[0007] This invention provides a pharmaceutical composition comprised of an effective amount of a statin and an effective amount of an Lp(a) inhibitor. More particularly, one embodiment of the invention is a combination of a statin with an Lp(a) inhibitor which is a retinoid.

[0008] The combinations of this invention can also employ the pharmaceutically acceptable salts of the respective active components.

[0009] Preferred retinoids are those described in U.S. Pat. No. 5,489,611. Especially preferred compositions employ 9-cis-retinoic acid or 13-cis-retinoic acid. Also preferred are trans-retinal and trans-retinol, as well as 13-cis-retinol, 13-cis-retinal, 9-cis-retinol, and 9-cis-retinal.

[0010] Another class of preferred Lp(a) inhibitors to be employed are aminophosphonates of the Formula I

I

[0011] wherein:

[0012] X^1 and X^2 independently are hydrogen, straight or branched C_1-C_8 alkyl and C_1-C_8 alkoxy, hydroxy, or nitro;

[0013] X^3 is hydrogen or C_1-C_4 alkyl; or X^3O, together with one of X^1 or X^2 is an alkylidene dioxy ring having from 1 to 4 carbon atoms;

[0014] R^1 and R^2 independently are hydrogen, or straight or branched C_1-C_6 alkyl;

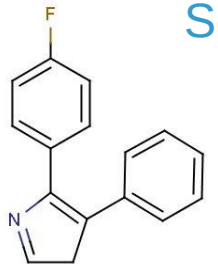
[0015] B is CH_2, $—CH_2CH_2—$, or $CH═CH—$;

[0016] n is 0 or 1;

[0017] Z is hydrogen, straight or branched C_1-C_6 alkyl, an aryl group Ar, or R^3CO, where R^3 is C_1-C_4 alkyl or perfluoro C_1-C_4 alkyl;

[0018] A is hydrogen, $CH_2CH═CH_2$, straight or branched C_3-C_8 alkyl or

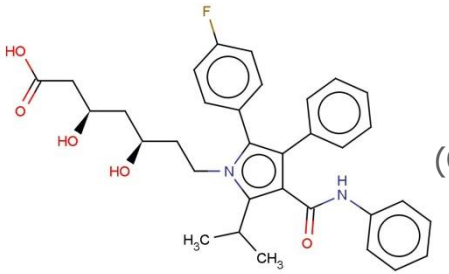
Search Input



Search Output

[R-(R*, R*)]-2-(4-fluorophenyl)-ß, d-dihydroxy-5-(1-methylethyl)-3-phenyl-4-[(phenylamino)carbonyl]-1H-pyrrole-1-heptanoic acid, calcium salt (2:1) trihydrate

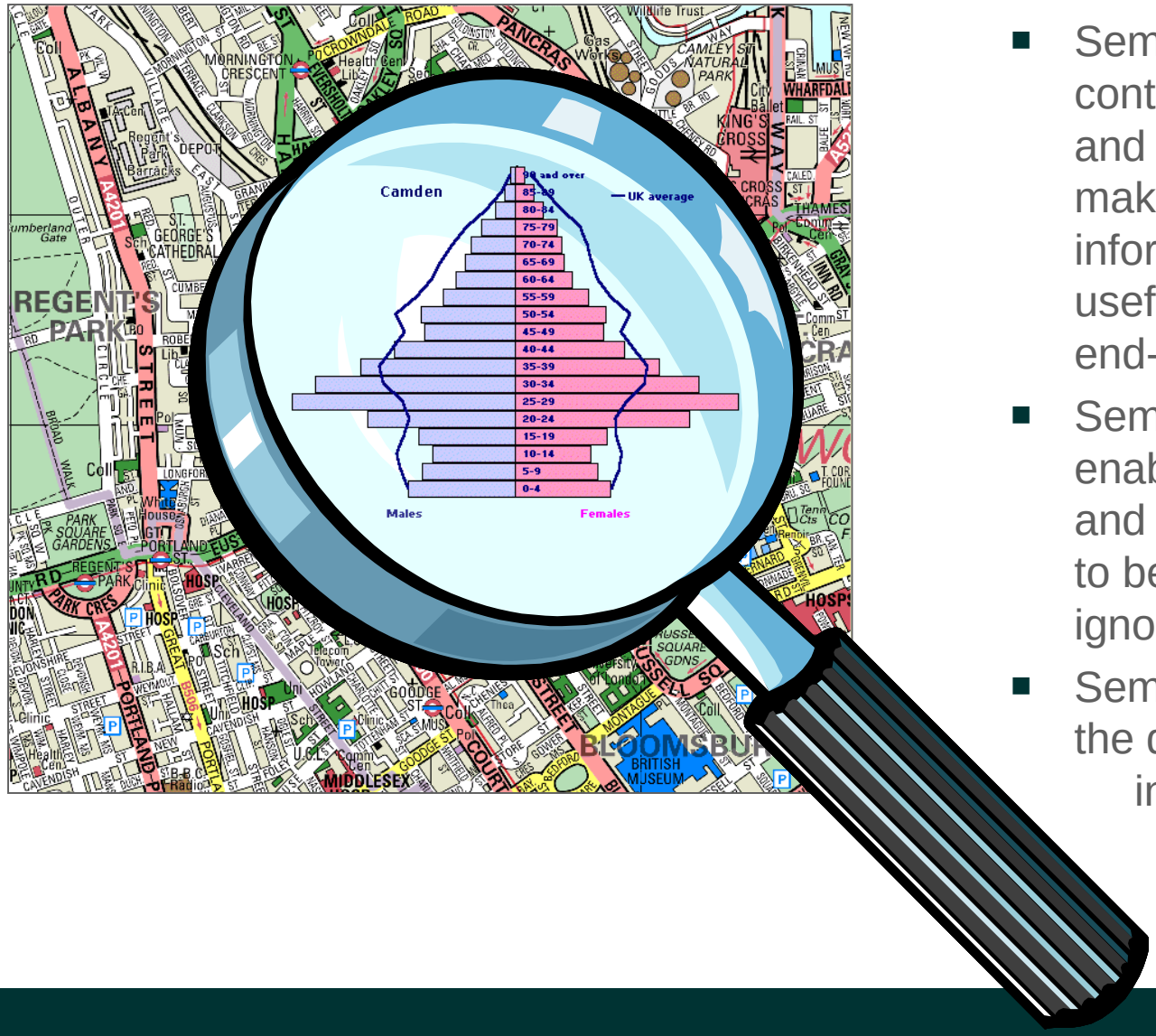
Atorvastatin
Lipitor
PD155-158



$(C_{33}H_{34}FN_2O_5)_2Ca \bullet 3H_2O$

CC(C)c1c(C(=O)Nc2ccccc2)c(-c3ccccc3)c(-c4ccc(F)cc4)n1CC[C@@H](O)C[C@@H](O)CC(O)=O

Semantic Lenses



- Semantic Lenses contain sets of filters and rules used to make the display of information more useful to a particular end-user
- Semantic Lenses enable specific data and evidence sources to be highlighted or ignored
- Semantic Lenses allow the display of information to be tailored to the type of data

Return on Investment Calculations

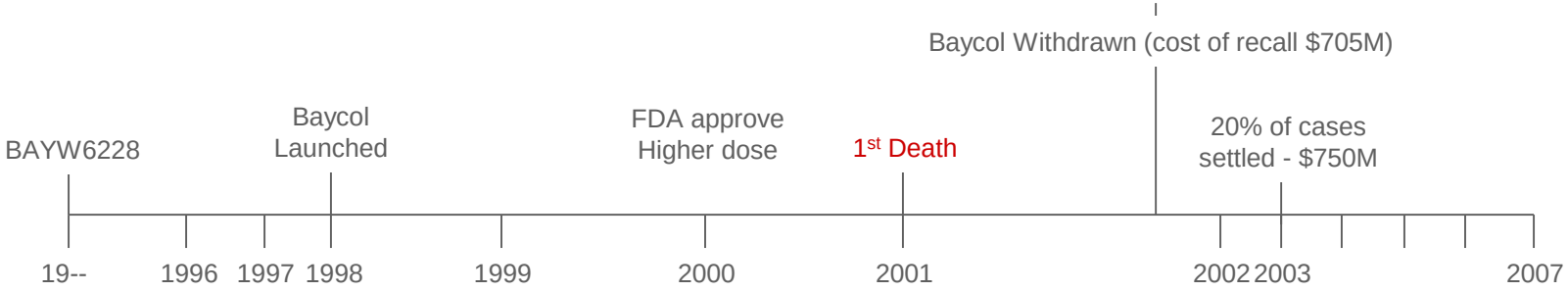
- Opportunity based
  - Do things you want to but can't/don't do now
    - Comprehensive systematic analysis
    - Identification of new business opportunities
    - Objective knowledge-led decision making
- Risk based
  - Protection from costly or negative outcome
    - Avoid missing side-effect liabilities
    - Assess opportunities quickly enough to secure position
    - Evaluate project risks and market potential accurately
- Productivity based
  - Improvement of existing processes
  - Savings of time, headcount or money

What Ontology can do for R&D

- Helps eliminate liabilities early

- >100 killed, 1000's injured
- Many information sources
  - Human genome / proteome
  - Clinical & pre-clinical experience
  - Similar cases
- Investigation hampered by lack of 'system'
  - Different people, different jurisdiction, different locations

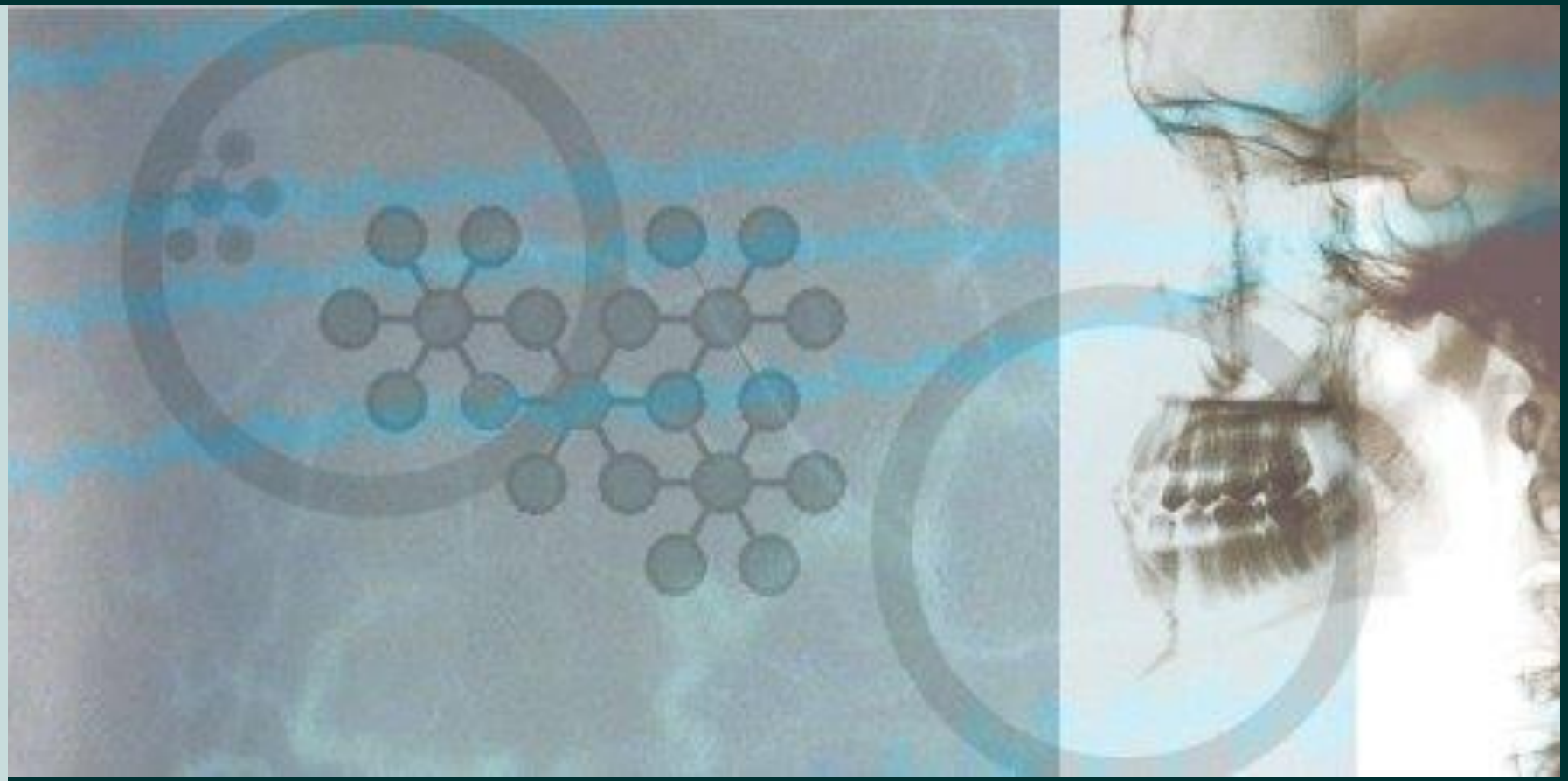
Cost
$1-4B cash



biowisdom

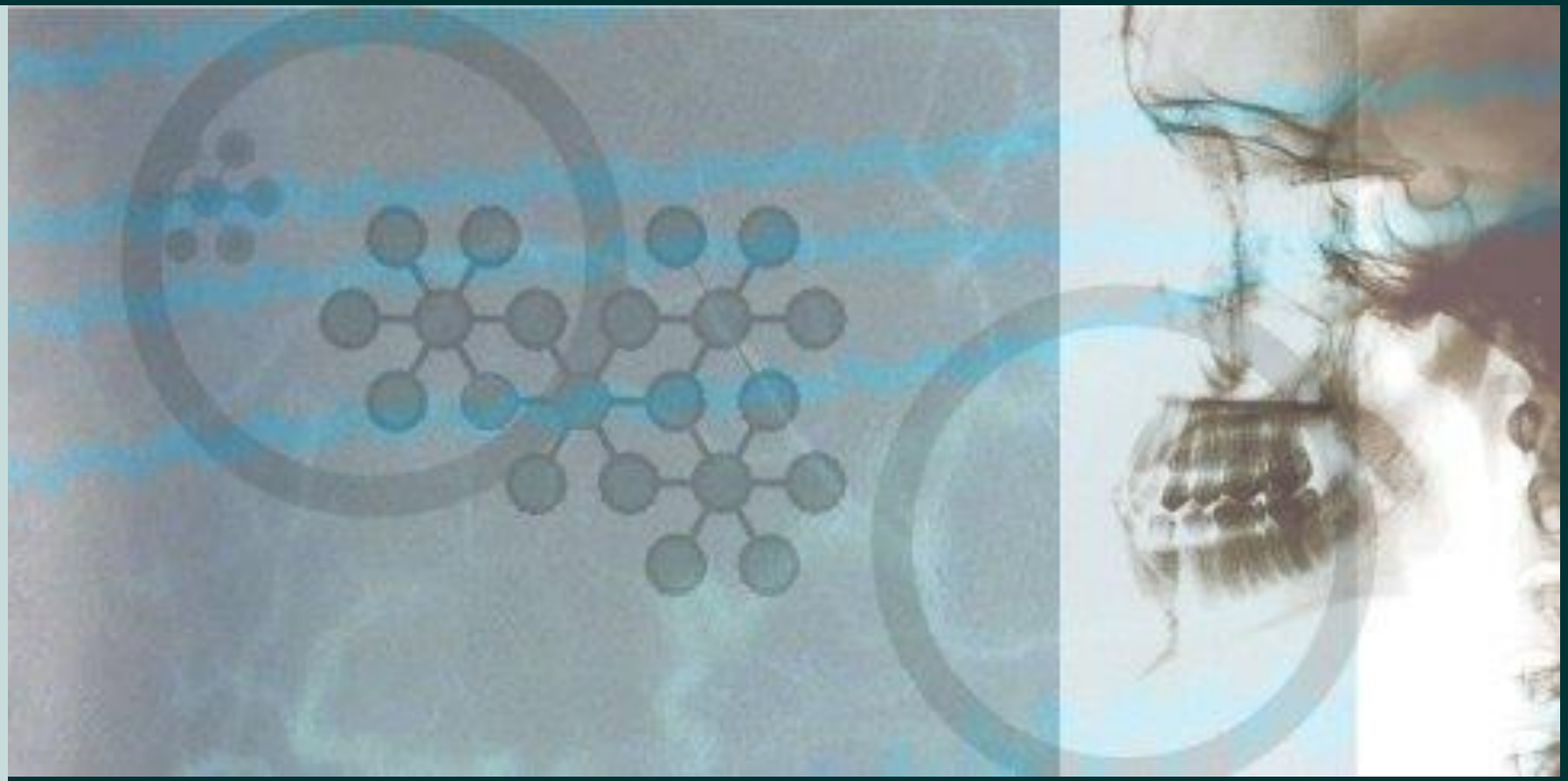
Case Studies

Sheryl Torr-Brown, Pfizer

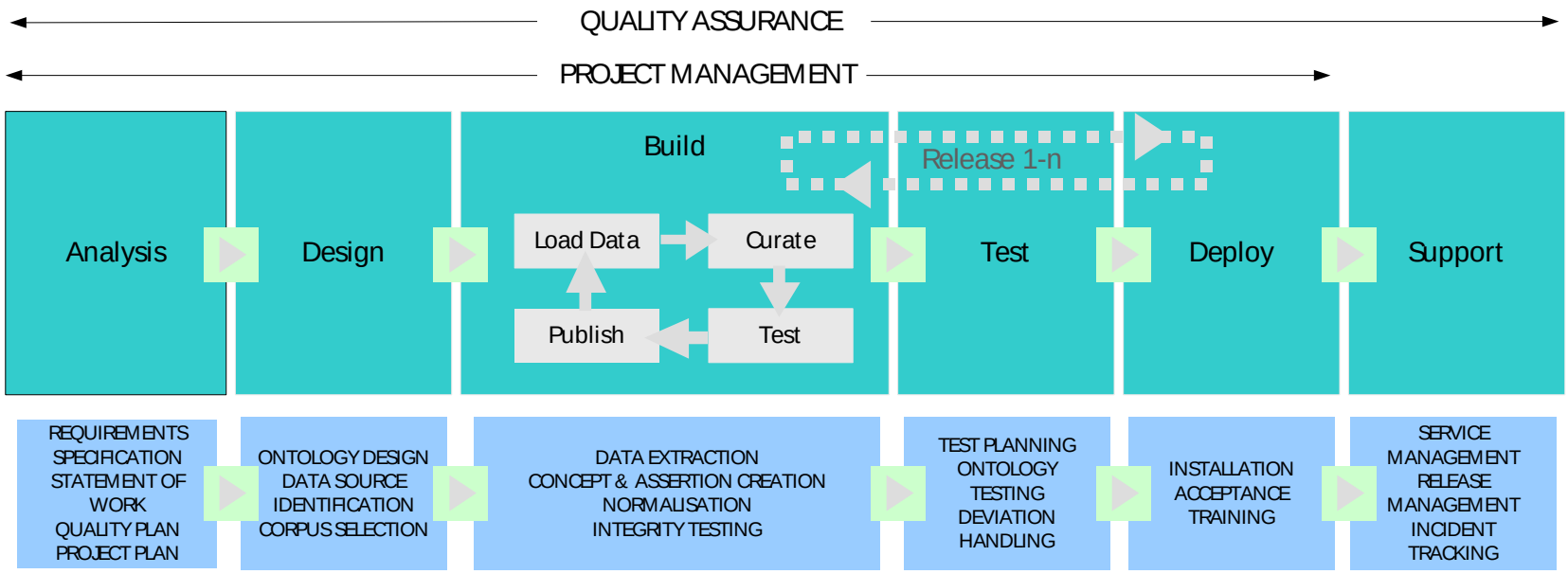


biowisdom

Construction of Ontologies

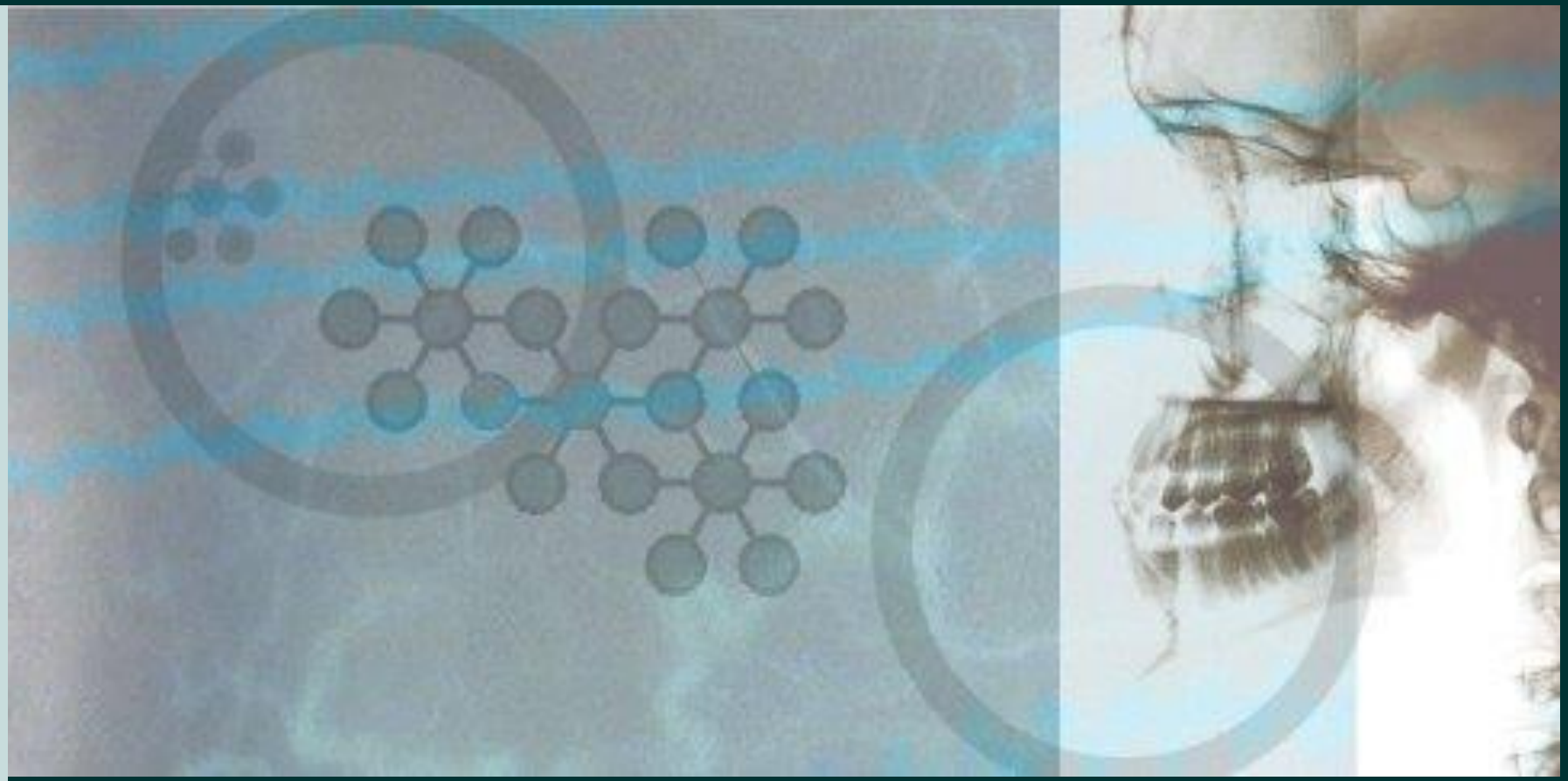


Ontology Curation Process



biowisdom

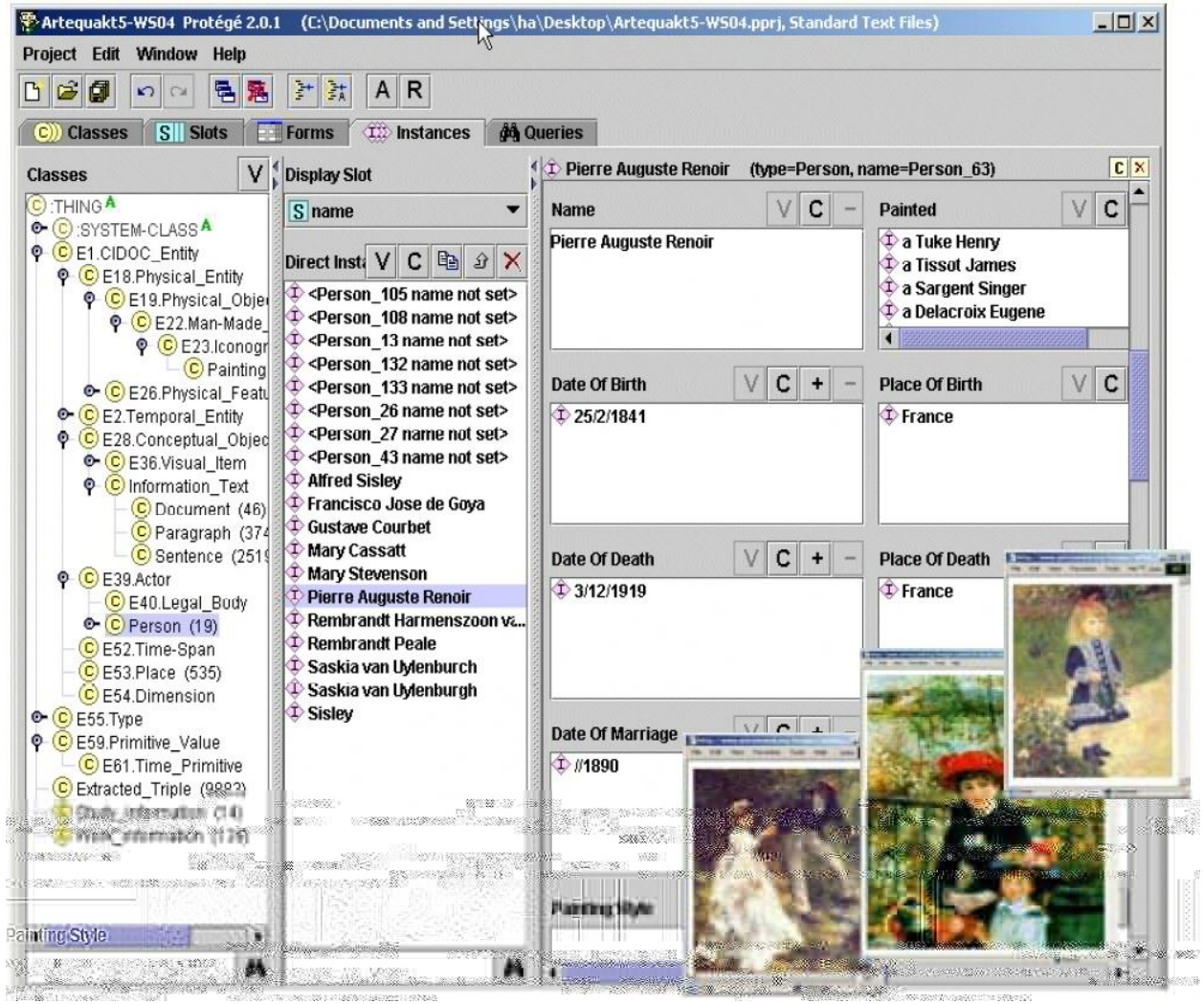
Standards & Ontology Languages



Ontology Standards

- XML
  - Structured information interchange format
- RDF
  - Designed for classification/search applications
  - Oriented around <subject> <predicate> <object> triple
  - Uses URIs (e.g. LSIDs) for resource location
  - Each triple can be joined with other triples, but retains its unique meaning regardless of the complexity of the model
- OWL (Lite, DL, Full)
  - Lite – limited language subset supporting taxonomies
  - DL – simple extensions supporting Description Logics
  - Full – full blown semantic ontology, not guaranteed to be computationally complete

Example Ontologies

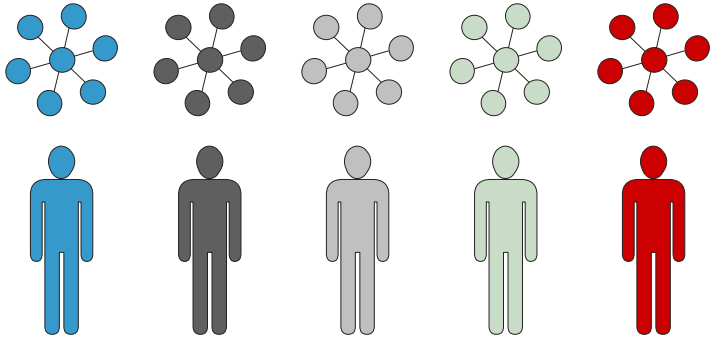


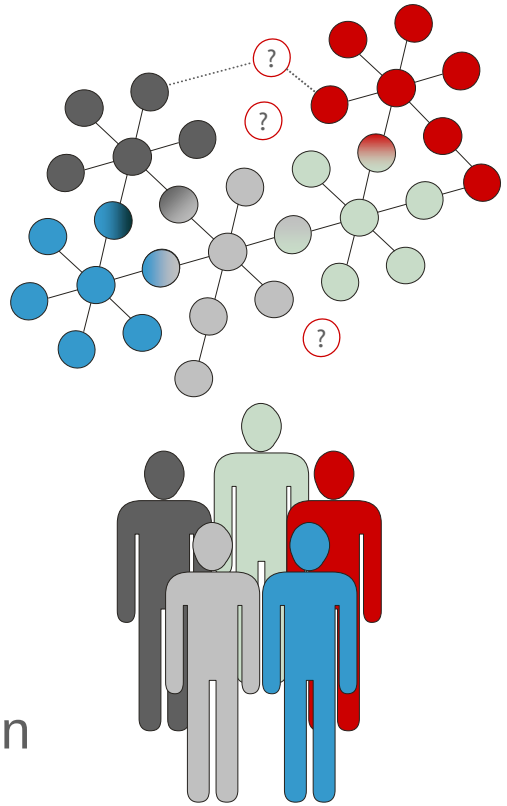
Example Ontologies

Public Domain Ontology Initiatives

- W3C
  http://w3c.org/
- Ontaria - 858 sources, 2.5M assertions
  http://www.w3.org/2004/ontaria/
- Ontoweb
  http://ontoweb.aifb.uni-karlsruhe.de/
- OpenRDF
  http://www.openrdf.org/
- Protégé
  http://protege.stanford.edu/
- Gene Ontology
  http://www.geneontology.org/
- Biological Processes Ontology
  http://smi-web.stanford.edu/projects/helix/pubs/process-model/
- HL7-RIM
  http://www.ics.mq.edu.au/~borgun/Software.html

Value of Ontology

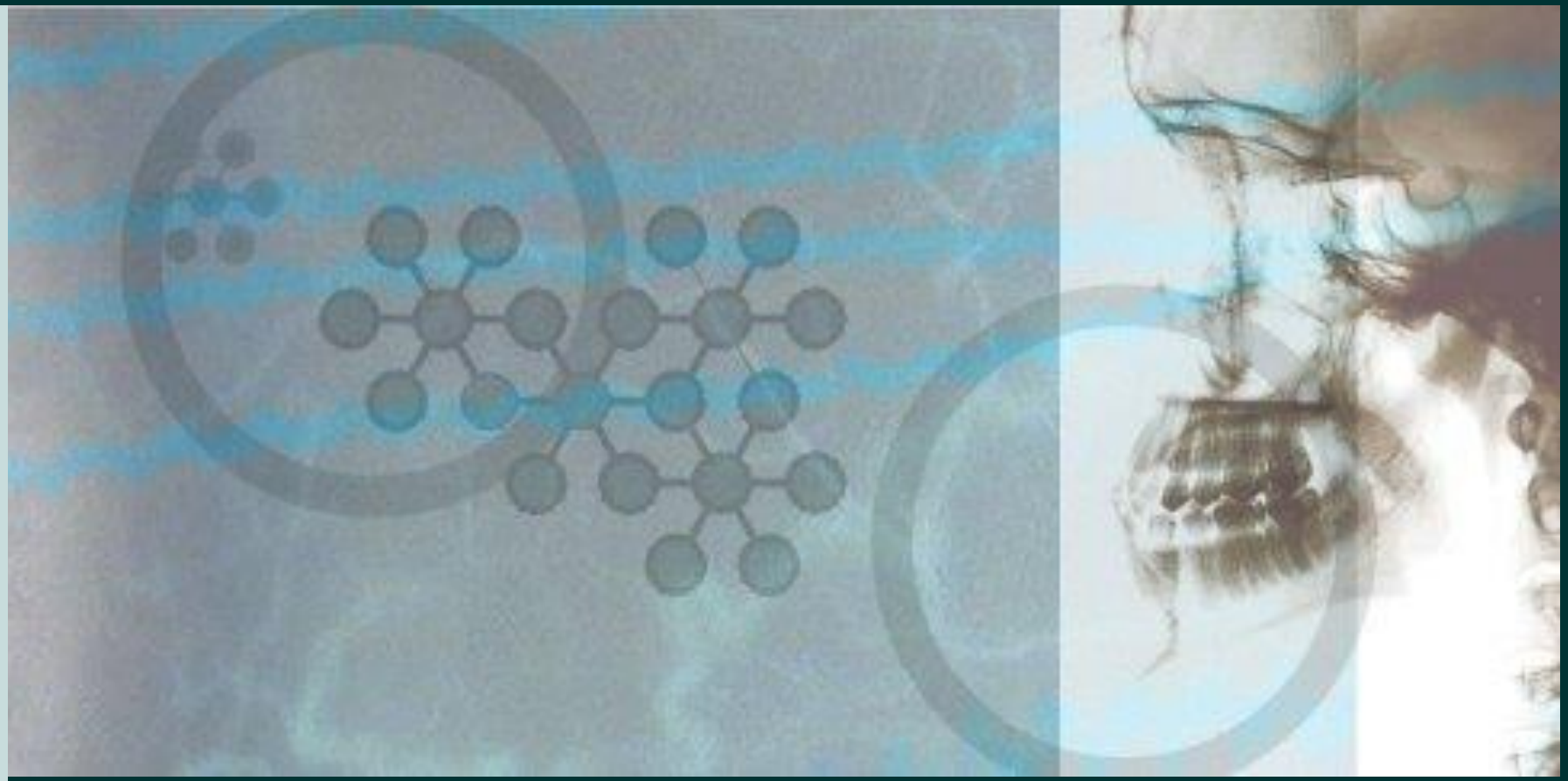




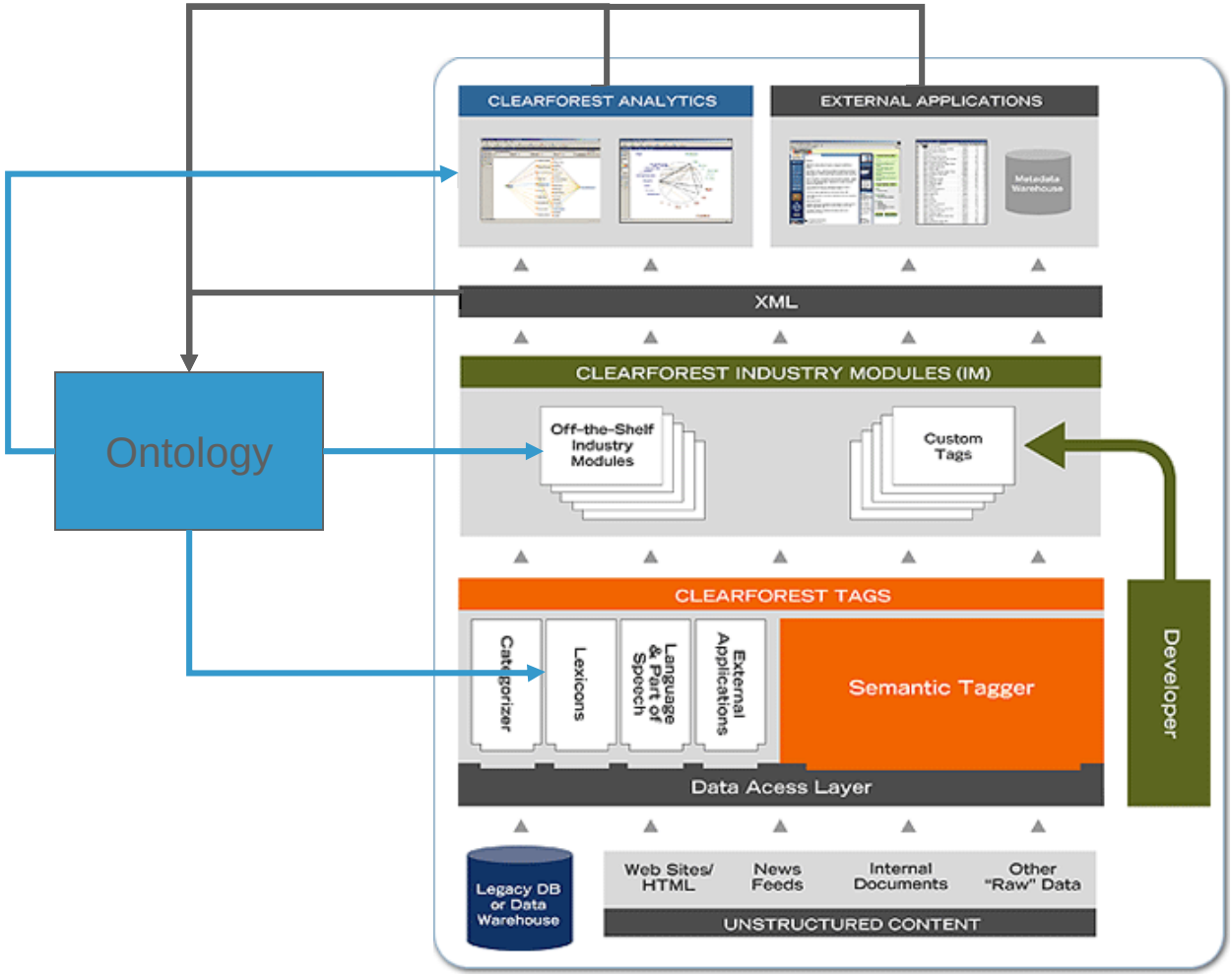
- Makes teams' knowledge visible
- Facilitates collaboration and communication
- Identifies knowledge gaps
- Supports multiple business applications
- Makes knowledge available for re-use on new projects

biowisdom

steve.gardner@biowisdom.com



Potential Integration with ClearForest

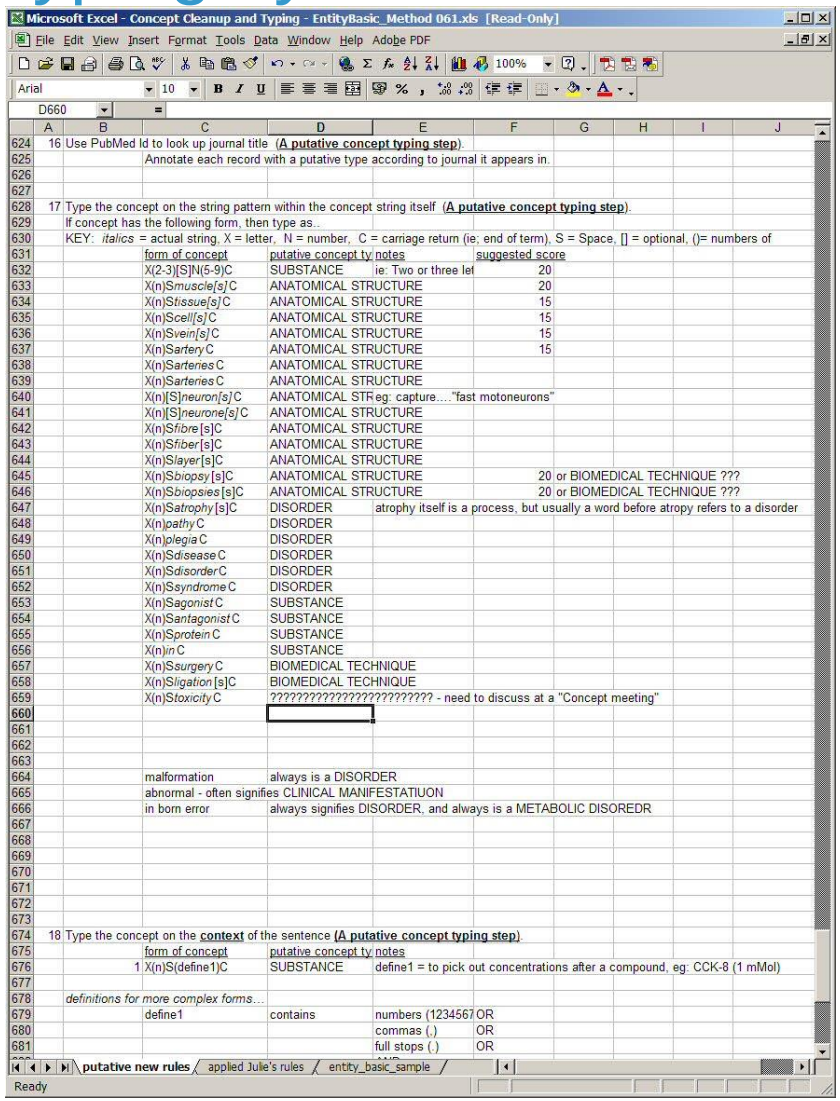


An Ontology is an Atlas

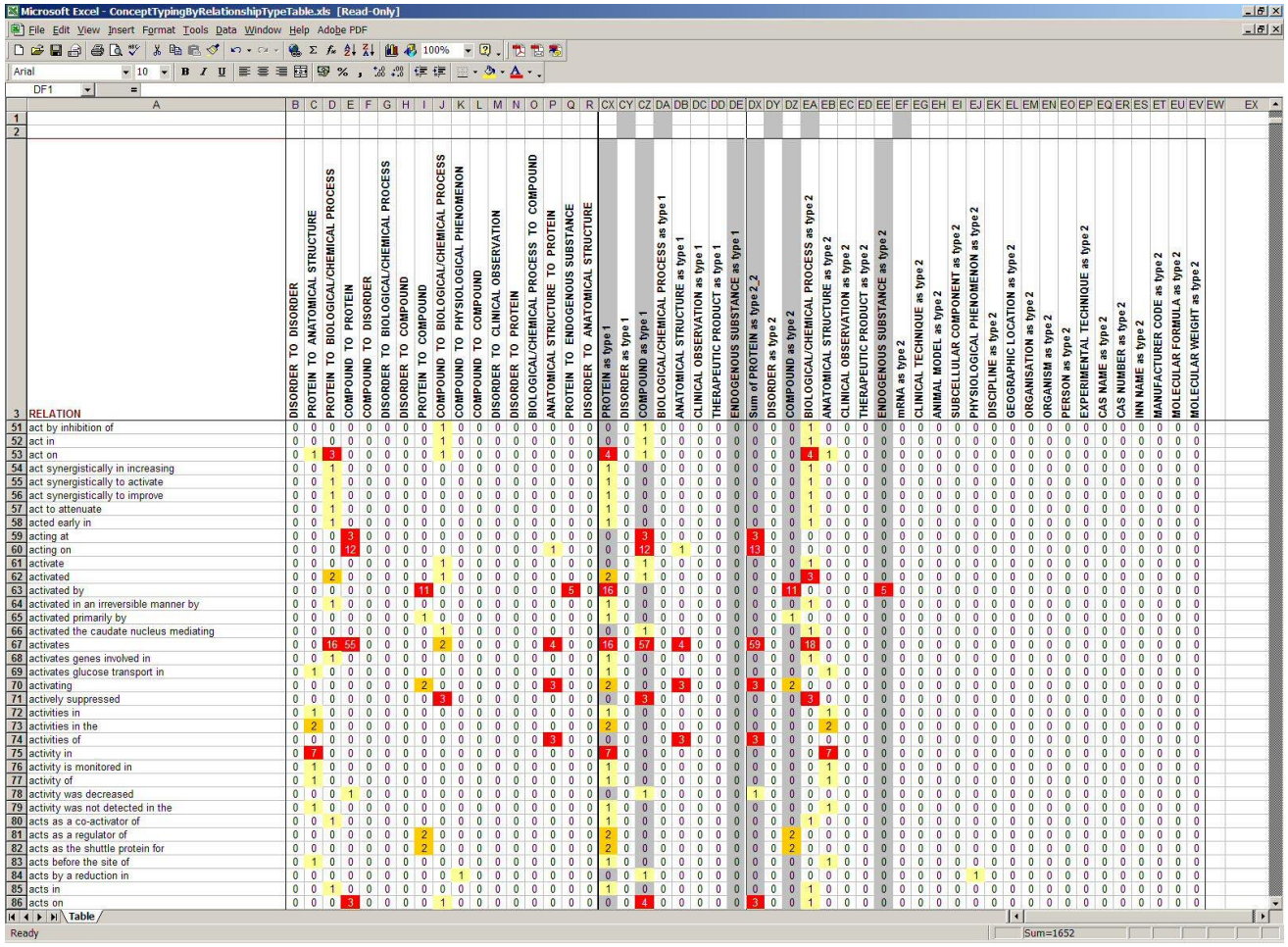
- Contains the names of all important things (places)
- Contains attributes of all things (size, postcode, counties, population, etc)
- Contains the links between one thing and all others it is connected to (routes)

- Everybody has ontologies in their head – they are our way of looking at and interpreting the world
- Relationships depend on context (tube, bus or car)

Concept Typing by Rules



Disambiguation by Relationship



Curation and Document Analysis Tools

for Information Scientists