



The Pistoia SESL Project

Pistoia Alliance SESL Pilot for a Biomedical Brokering Service

Dietrich Rebholz Schuhmann, Christoph Grabmueller

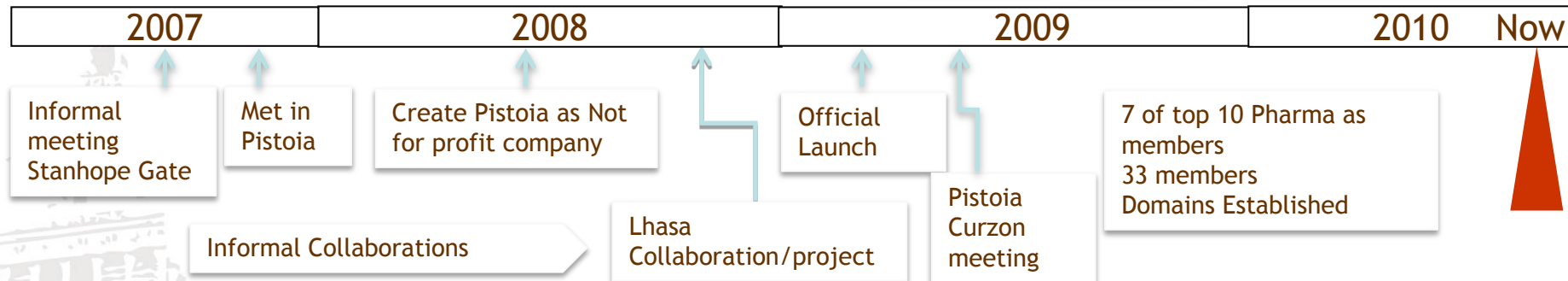
8th PharmaTech IT Congress

8th Nov 2010

<http://pistoiaalliance.org>



Pistoia Background - How it all started



Pistoia Description

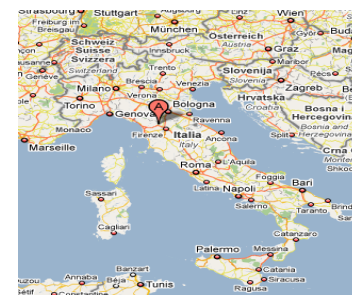
The primary purpose of the Pistoia Alliance is to streamline non-competitive elements of the life science workflow by the specification of common standards, business terms, relationships and processes

Pistoia Goals

- to allow this framework to encompass/support most pre-competitive work between the organisations
- to support life science workflow prior to submission
- to work with other Standards organisations

History

Initial Meeting with GSK, AZ, Pfizer and Novartis - outlined similar challenges and frustrations in the Informatics sector of Discovery



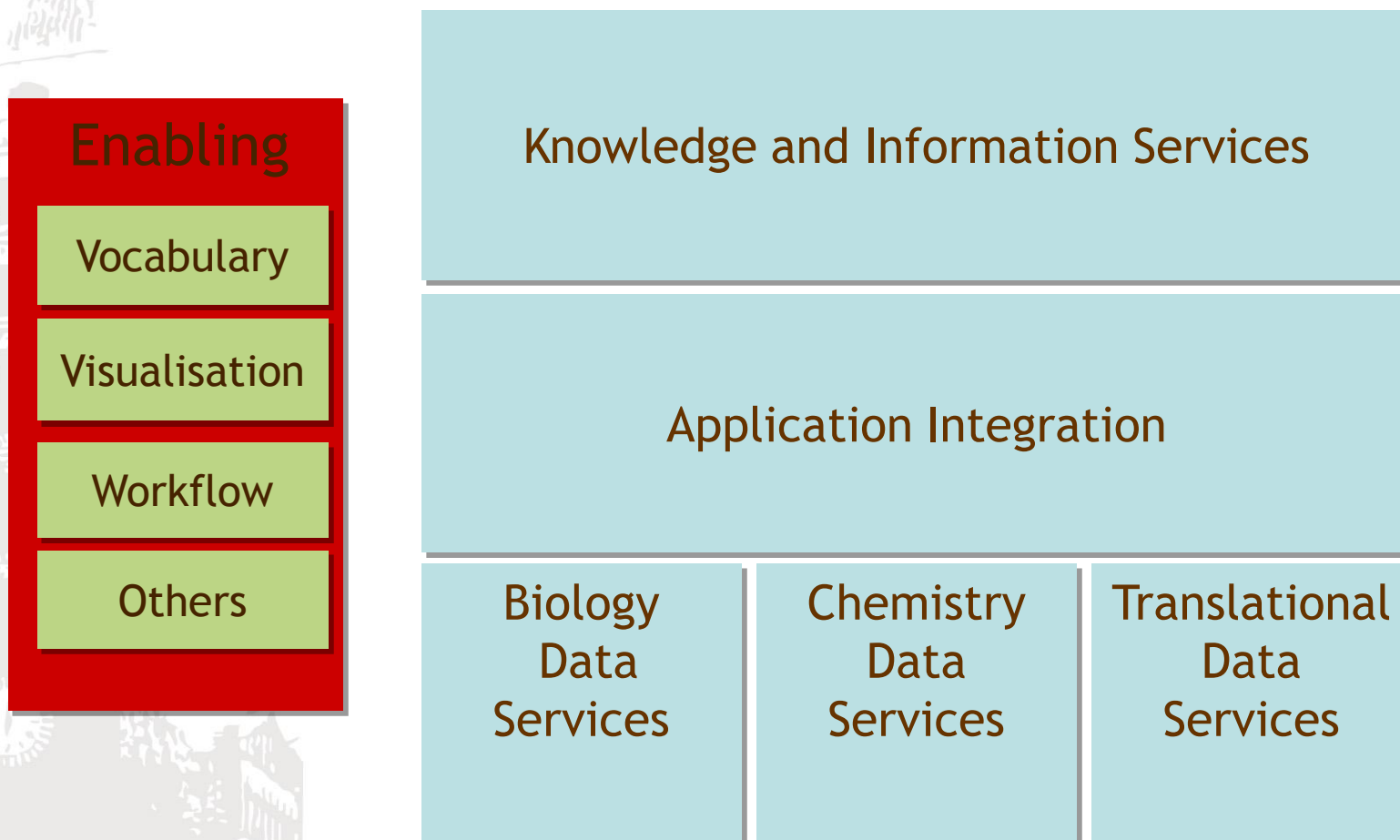
The advent of Web Services and Web2.0 allow for decoupling of proprietary data from technology

Publicly available structural and biological DBs allow for a non-IP related analysis and as a scientific test suite.

Sponsorship from R&D IS heads within Life Science industry

Pistoia Domains

Pistoia Domains focus on business workflows /supply chains

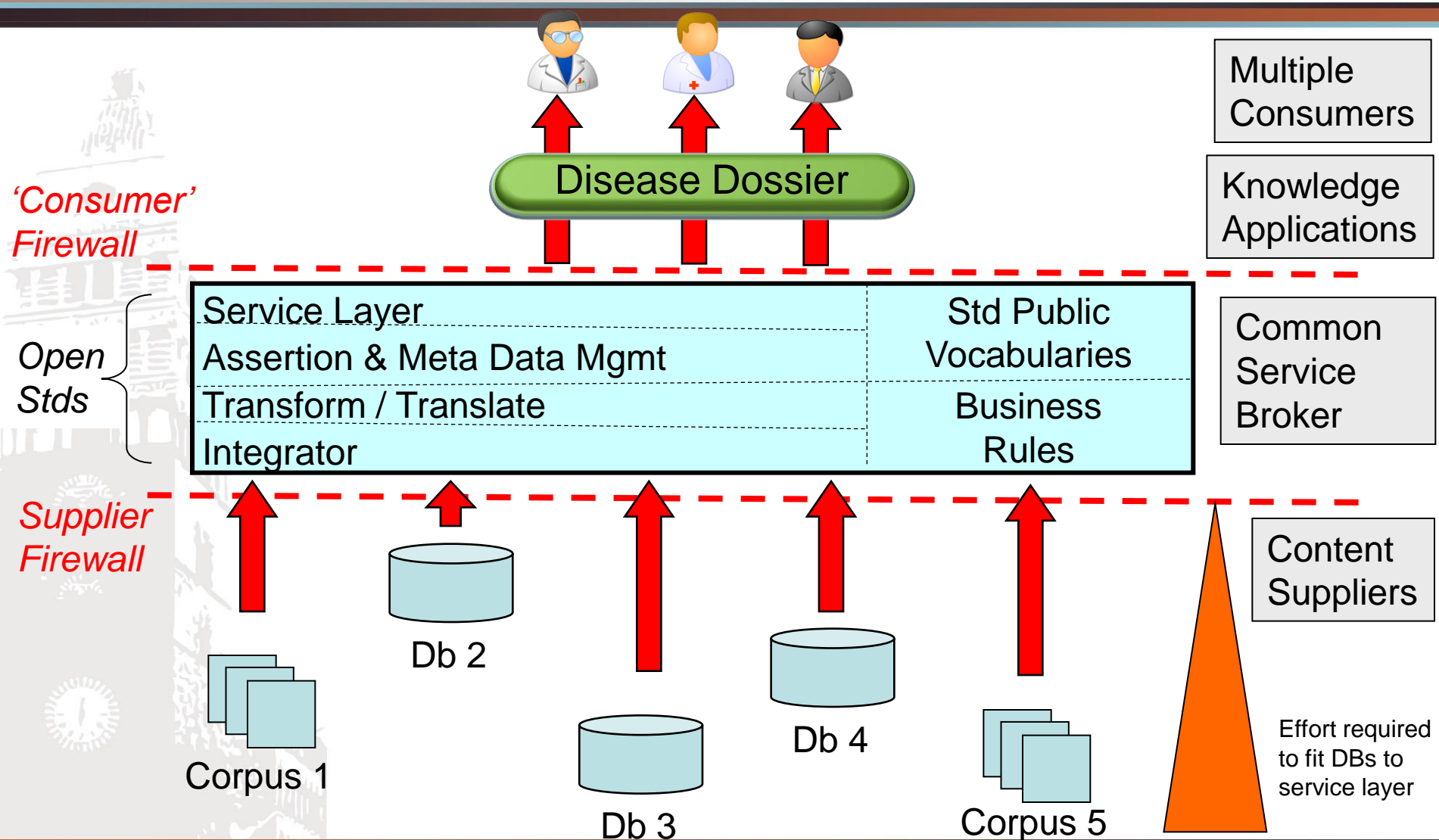


SESL: Biomedical Knowledge Brokering

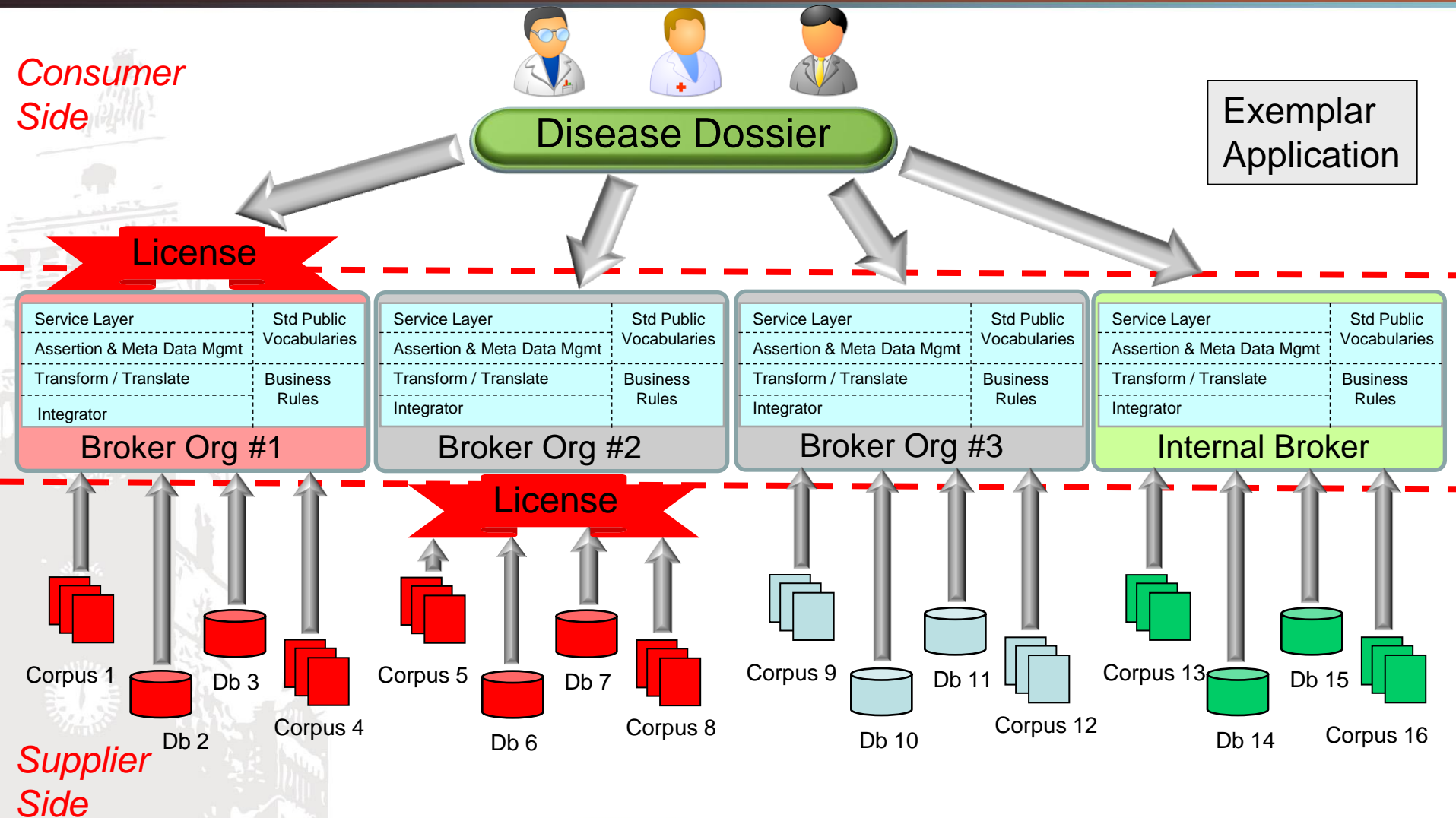


- Challenge:
 - No single system for retrieving gene to disease relationships contained in both published & biological database content
 - Need a 'push model' for biomedical knowledge access: the current model requires the consumer to search 1000's of content sources
- Opportunity: Pilot Project with key stakeholders
 - Pilot a 'push model' for biomedical knowledge brokering
 - Engage multiple consumers, content providers and a single, public group to develop the necessary infrastructure to explore the standards required for the model to work in production
- History:
 - May 2008: Common Disease Knowledge Environment (CDKE) IMI call drafted
 - Sep 2008: postponed call publication
 - Jan 2009: x-pharma meeting in London on how to progress CDKE
 - Apr 2009: CDKE presented at SESL workshop
 - Oct 2009: SESL Pilot meeting (funders)
 - Jan 2010: Pilot launch

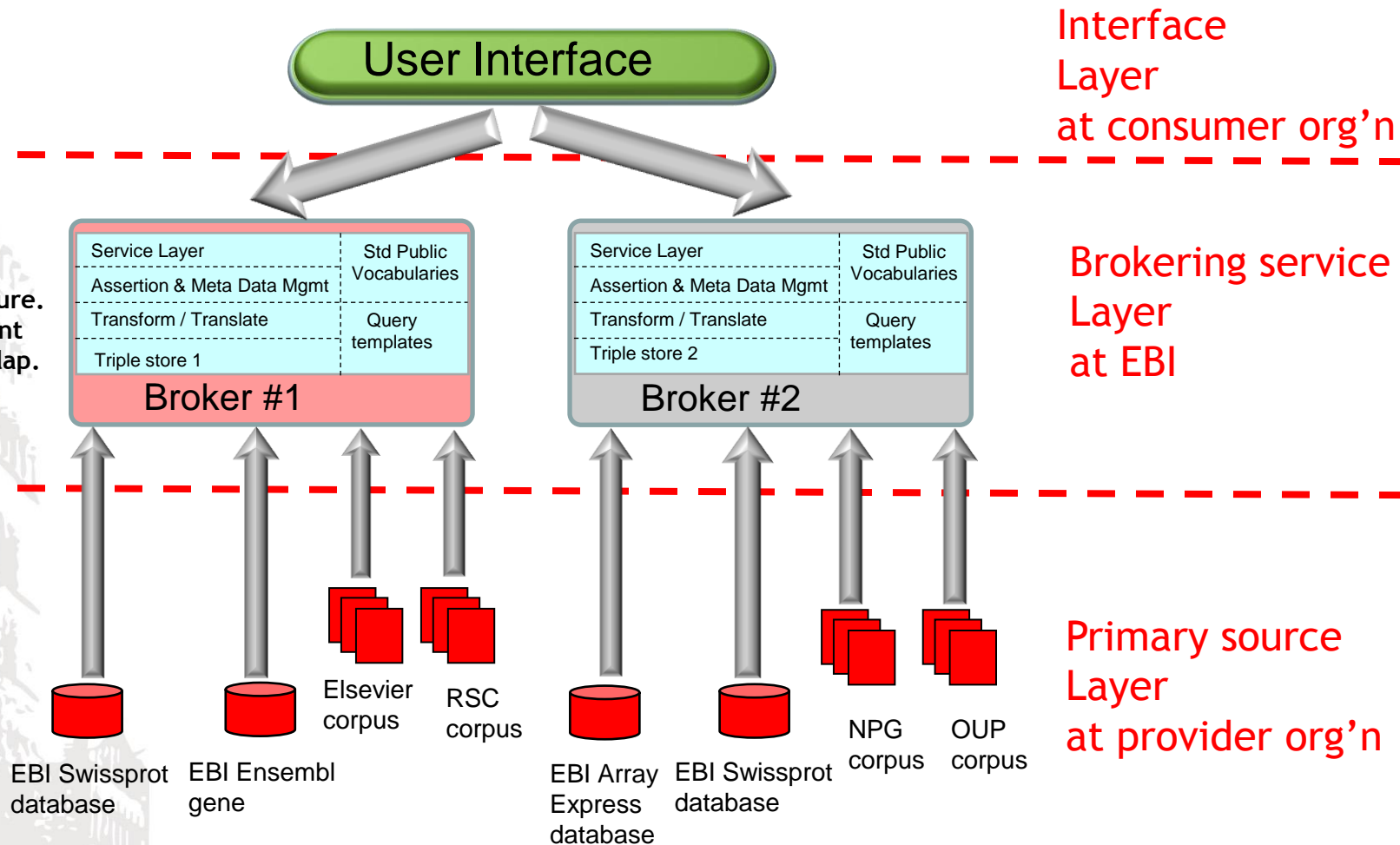
The Knowledge Service Framework



A Production Service vision...



Minimal configuration to test a brokering service



SESL user interface Mock Up: Results

SESL Results

	Gene	R'ship	Disease	Species	Evidence
1	<i>abc1</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Mus</i>	<u><i>Paper UID:1234</i></u>
2	<i>abc1</i>	<i>Up-Reg</i>	<i>Diabetes</i>	<i>Homo</i>	<u><i>ArrayExpress: XXX</i></u>
3	<i>abc2</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Homo</i>	<u><i>Paper UID:1344</i></u>
4	<i>abc13</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Mus</i>	<u><i>Paper UID:1314</i></u>
5	<i>abc7</i>	<i>Mutation</i>	<i>Diabetes</i>	<i>Rattus</i>	<u><i>OMIMI: XXX</i></u>
6	<i>abc1</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Mus</i>	<u><i>Paper UID:45643</i></u>
7	<i>abc1</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Homo</i>	<u><i>Paper UID:2143</i></u>
8	<i>abc1</i>	<i>Co-occurs</i>	<i>Diabetes</i>	<i>Mus</i>	<u><i>Paper UID:1204</i></u>

Diabetes genes in SESL prototype

Gene symbols	Primary data sources			SESL project			
	UNIPROT Involvement in diabetes type 2?	OMIM in diabetes?	ArrayExpress Atlas?	SESL?	SESL co-occurrence in Full Text literature?	SESL OMIM diabetes?	SESL Arrayexpress Atlas?
ABCC8	Y - NIDDM	Y	Y	Y	Y	Y	N
ADIPOQ	Y - NIDDM	Y	Y	Y	Y	N	N
ALMS1	Y - NIDDM	N	Y	Y	Y	N	N
BLK	Y - MODY11	Y	Y	Y	Y	N	N
CAPN10	Y - NIDDM	Y	Y	Y	Y	Y	N
CEL	Y - MODY8	Y	Y	Y	Y	Y	N
ENPP1	Y - NIDDM	Y	Y	Y	Y	Y	N
GCK	Y - MODY2	Y	Y	Y	Y	Y	N
HNF1A	Y - MODY3	Y	Y	Y	Y	Y	N
HNF1B	Y - MODY5	Y	Y	Y	Y	Y	N
HNF4A	Y - MODY1	Y	Y	Y	Y	Y	N
INPPL1	Y - NIDDM	N	Y	N			
INS	Y - MODY10	Y	Y	Y	Y	N	N
IRS1	Y - NIDDM	Y	Y	Y	Y	N	N
KCNJ11	Y - NIDDM	Y	Y	N			
KLF11	Y - MODY7	Y	Y	Y	Y	Y	N
LIPC	Y - NIDDM	Y	Y	Y	Y	Y	N
MAPK8IP1	Y - NIDDM	Y	Y	Y	Y	Y	N
MCF2L2	Y - NIDDM	N	Y	N			
NEUROD1	Y - MODY6	Y	N	Y	Y	Y	N
PAX4	Y - MODY9	Y	Y	Y	Y	Y	N
PDX1	Y - MODY4	Y	Y	Y	Y	N	N
PPARG	Y - NIDDM	Y	Y	Y	Y	Y	N
PPP1R3A	Y - NIDDM	Y	Y	Y	Y	N	Y
SLC30A8	Y - NIDDM	Y	Y	Y	N	Y	N
TCF7L2	Y - NIDDM	Y	Y	Y	Y	Y	N
UCP1	Y - NIDDM	N	Y	Y	Y	N	N
WFS1	Y - NIDDM	Y	Y	Y	Y	Y	N

Diabetes genes in SESL prototype



Diseases cooccurring by a gene

Umls	Documents
Cystic Fibrosis (C0010674)	41
Hereditary Diseases (C0019247)	5
Knock-out (C1522225)	2
Pancreatitis (C0030305)	
Azoospermia (C0004509)	
Pancreatitis, Chronic (C0149521)	
Pancreatitis, Alcoholic (C0376670)	
Diabetes Mellitus (C0011849)	
Infertility (C0021359)	
Infection (C0021311)	
hypercholesterolemia (C0020443)	
Metabolic syndrome (C0948265)	
Aggressive behavior (C0001807)	
Bronchiectasis (C0006267)	
Lung diseases (C0024115)	
Bicuspid aortic valve (C0149630)	
Male infertility (C0021364)	
Osteoporosis (C0029456)	

Differentially expressed diseases by a gene

Umls	ArrayExpress
Never Married	E-GEOD-3189

Diseases by a gene

Umls	OMIM id
Cystic Fibrosis	219700
Congenital bilateral aplasia of vas deferens	277180

Interactions by a gene

Gene/Protein I	Gene/Protein II
Q99942	RNF5
Q5T2W1	PDZK1
Q9BUN8	DERL1
Q9HD26	GOPC
P13569	CFTR
Q15599	SLC9A3R2
Q14745	SLC9A3R1
P51572	BCAP31
Q9HBW0	LPAR2

GOA annotations for a gene

Id	Name(s)
0007585	respiratory gaseous exchange; breathing; respiration

Diabetes genes in SESL prototype



Diseases cooccurring by a gene

Umls	Documents
Cystic Fibrosis (C0010674)	41
Hereditary Diseases (C0019247)	5
Knock-out (C1522225)	3

Pancreatitis (C0030305)
Azoospermia (C0004509)
Pancreatitis, Chronic (C0149521)
Pancreatitis, Alcoholic (C0376670)
Diabetes Mellitus (C0011849)
Infertility (C0021359)
Infection (C0021311)
hypercholesterolemia (C0020443)
Metabolic syndrome (C0948265)
Aggressive behavior (C0001807)
Bronchiectasis (C0006267)
Lung diseases (C0024115)
Bicuspid aortic valve (C0149630)
Male infertility (C0021364)
Osteoporosis (C0029456)

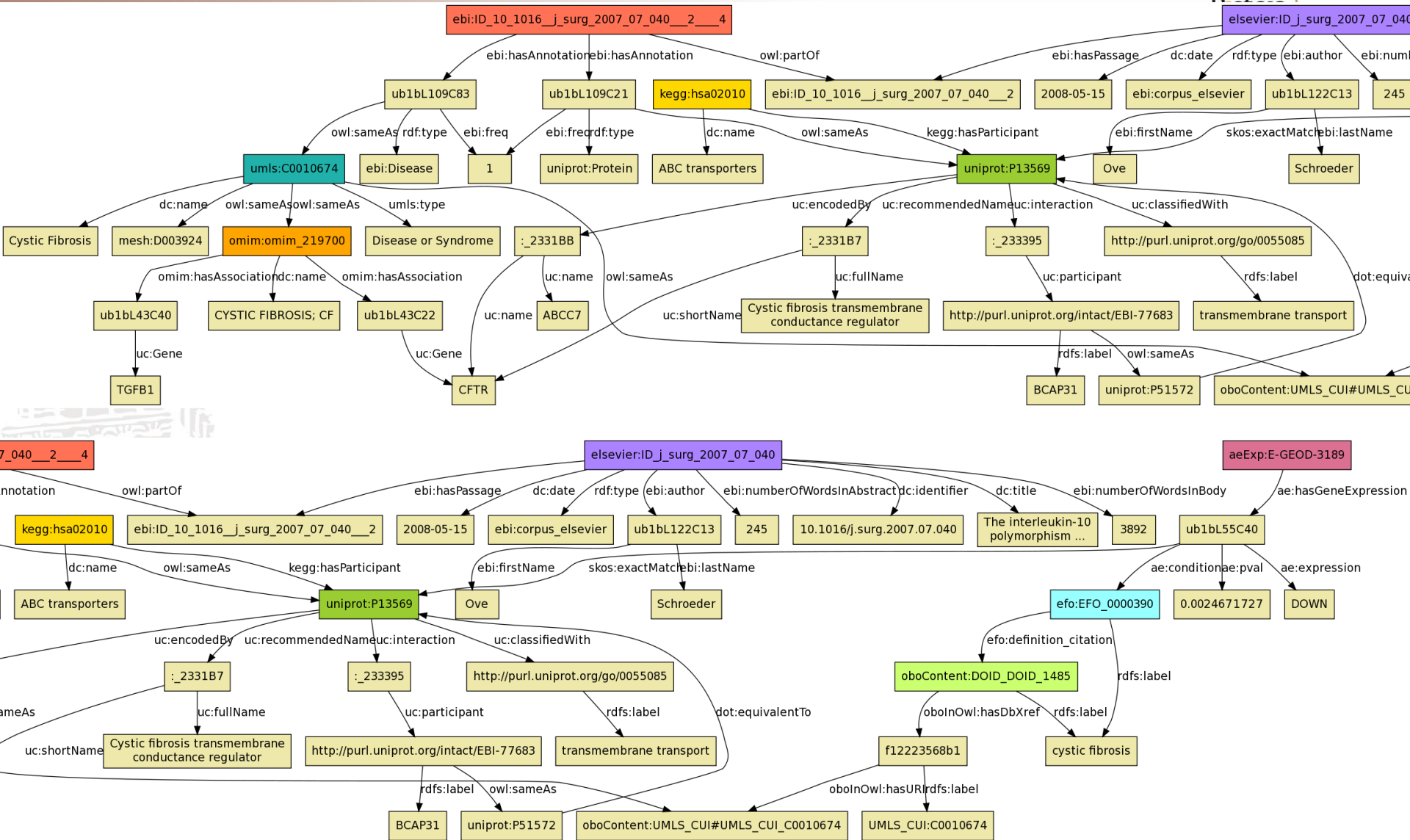
List of documents where 'CFTR' co-occur with 'Cystic Fibrosis'

Doi	Publisher	Title	Date
10.1016/j.tim.2009.06.006	ELSEVIER	Color me bad: microbial pigments as virulence factors	n/a
10.1016/j.bmc.2008.06.026	ELSEVIER	In vitro inhibition of glycogen-degrading enzymes and glycosidases by six-membered sugar mimics and their evaluation in cell cultures	2008-05-23

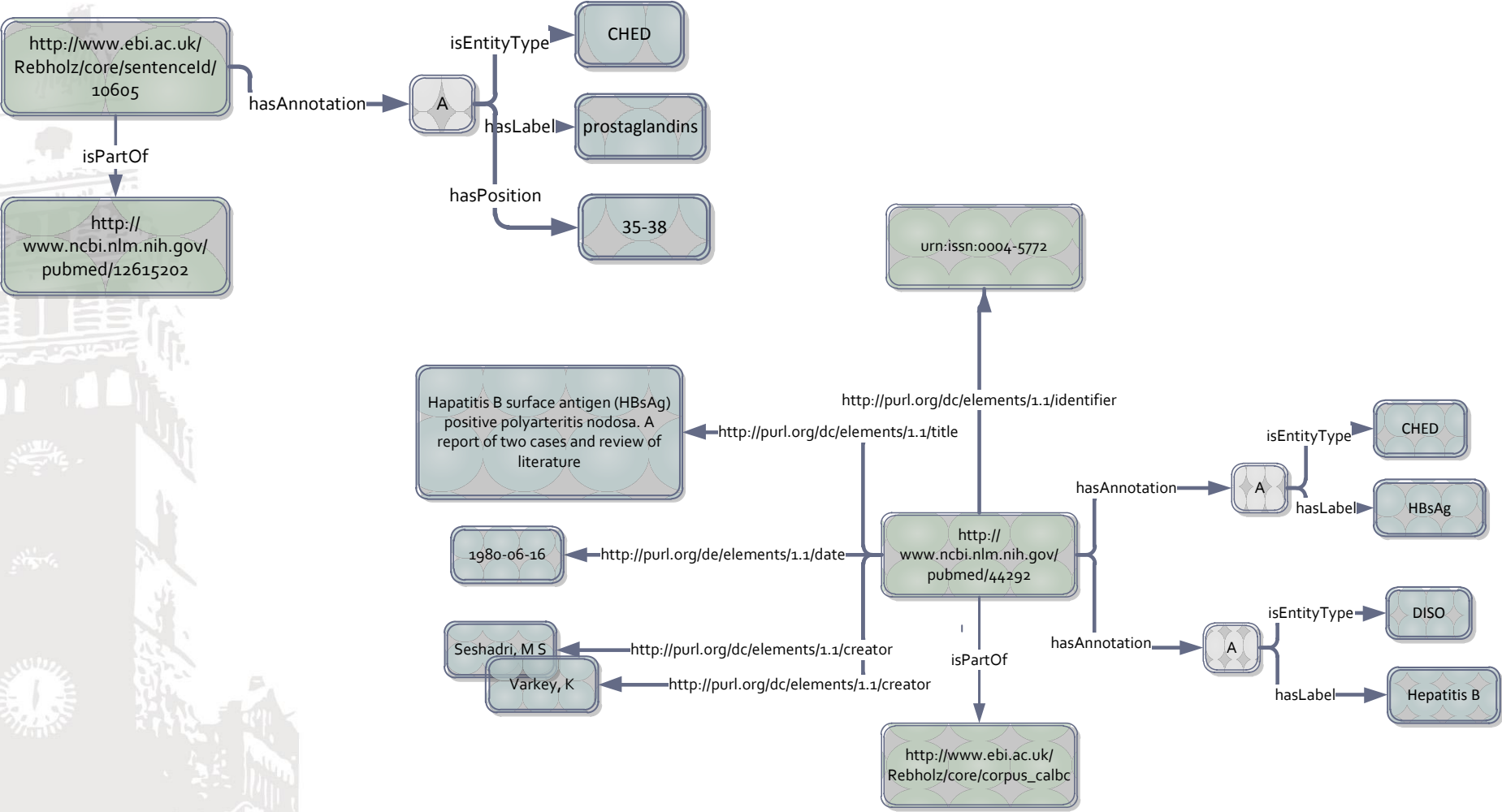
The fundamental ability of pyocyanin to alter the redox cycle and increase oxidative stress appears central to its diverse detrimental effects on host cells. For example, pyocyanin disrupts Ca^{2+} homeostasis in human airway epithelial cells by oxidant-dependent increases in inositol trisphosphate and the abnormal release of Ca^{2+} from intracellular stores. Because Ca^{2+} is important for regulating ion transport, mucus secretion and ciliary beat, these alterations probably have important ramification for *P. aeruginosa* lung infections [42]. The pathway of vacuolar ATPase vesicle transport and protein targeting appears particularly sensitive to pyocyanin action, as revealed in a yeast-mutant library screen [41]. Pyocyanin inhibition of ATPase could directly explain many of its toxicities including ciliary dysmotility [43], disruption of calcium homeostasis [42] and diminished apical membrane localisation of the **cystic fibrosis transmembrane conductance regulator** (CFTR) [44]. Other potentially toxic effects of pyocyanin include perturbation of cellular respiration, epidermal cell growth inhibition, prostacyclin release from lung endothelial cells and altered balance of protease-antiprotease activity in the **cystic fibrosis** lung [45].

Naturally occurring iminosugars DNJ (1) and fagomine (2) were isolated from the bark of *M. alba* (Moraceae),²⁶ and HNJ (3) and Glc-HNJ (4) were prepared from the roots of *S. tuberosa* (Stemonaceae),²⁷ according to the literature. Miglitol (5) was chemically prepared by heating DNJ with 2-bromoethanol in the presence of anhydrous K_2CO_3 in DMF according to the literature.²⁸ 1,5-Dideoxy-1,5-imino-d-xylitol (6) was purchased from Industrial Research Limited (Lower Hutt, New Zealand). Both d-IFG (7) and l-IFG (8) were enantiospecifically synthesized according to the literature.^{29,30} Valiolamine (9) was isolated from a fermentation broth of *Streptomyces hygroscopicus* var. *limoneus*, which is a producer of the antifungal agent validamycin,³¹ and voglibose (10) was obtained by reductive amination of valiolamine with dihydroxyacetone,³² according to the literature. Acarbose (11) was prepared from commercially available Glucobay Tab 100mg. A tablet was dissolved in water and the filtrate was applied to a short column of Dowex 50W-X2 (H^+ form). The column was washed with water and eluted with 0.5M NH_4OH . The eluate was concentrated and lyophilized to give a powder of acarbose.

Full triple store schema



Access to document / sentence content (Medline & LexEBI)

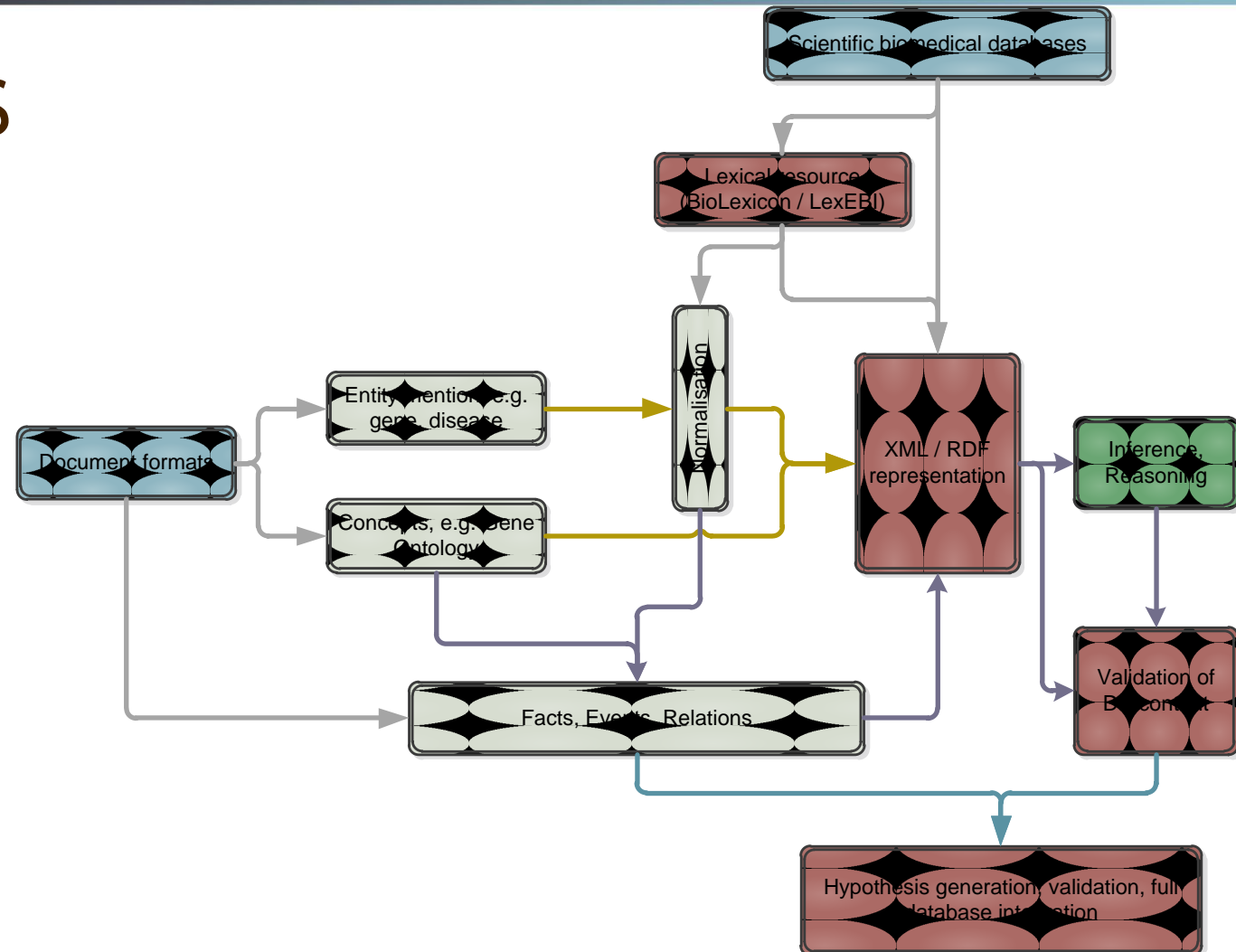


Content of the triple store

description	# triples
ArrayExpress homebrew	182,840
Experimental Factor Ontology (ArrayExpress)	49,026
UMLS homebrew	6,906,735
Disease Ontology	1,863,664
Gene Ontology	495,595
UniProt filtered for human	12,552,239
Overall triples on meta data from FT documents	3,485,212
Triples with gene annotations in FT documents	2,373,584
Triples with disease annotations in FT documents	4,983,788
Triples of GO annotations in FT documents	3,870,834

Content of the triple store

- SWAT4LS
2010
- Croset
et al.



Summary of accomplishments



- Significant progress towards realising the technical goal of knowledge brokering
- A unique consortium from three cultures: industry, publishers and academia
 - Working together - sharing efforts and risks
- Business opportunities and concerns
 - For data providers and consumers?

Acknowledgements



Industry

Ian Harrow
Wendy Filsell
Mike Westaway
Ian Stott
Peter Woollard
Nigel Wilkinson
Catherine Marshall
Ian Dix
Nick Lynch
Michael Braxenthaler
Ashley George

Publishers

Claire Bird - OUP
Richard O'Bierne - OUP
Jabe Wilson - Elsevier
Bradley Allen - Elsevier
Colin Batchelor - RSC
Richard Kidd - RSC
David Hoole - NPG
Alf Eaton - NGP

EMBL-EBI

Christoph Grabmueller
Silvestras Kavaliauskas
Roderigo Lopez
Dominic Clark
Jo McEntyre
Janet Thornton
Cath Brooksbank