

C.A.R.E. to C.R.A.M.

Clinical Analysis and Reporting Environment to Clinical Rules and Metadata.



Greg Fuller

GIANCE

20/09/2013





شركة أبوظبي للعمليات البترولية البرية (أدكو)
Abu Dhabi Company for Onshore Oil Operations (ADCO)



wellcome trust



THE
POWER
TO KNOW.



Merck Serono



PPD

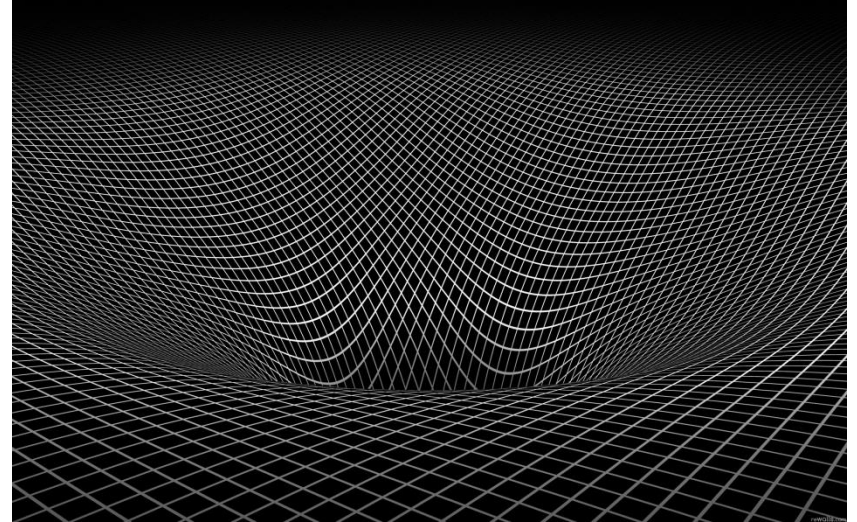


Stepping back

5 development projects 5 companies

Investment Prison

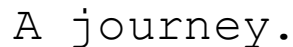
Cultural Bias





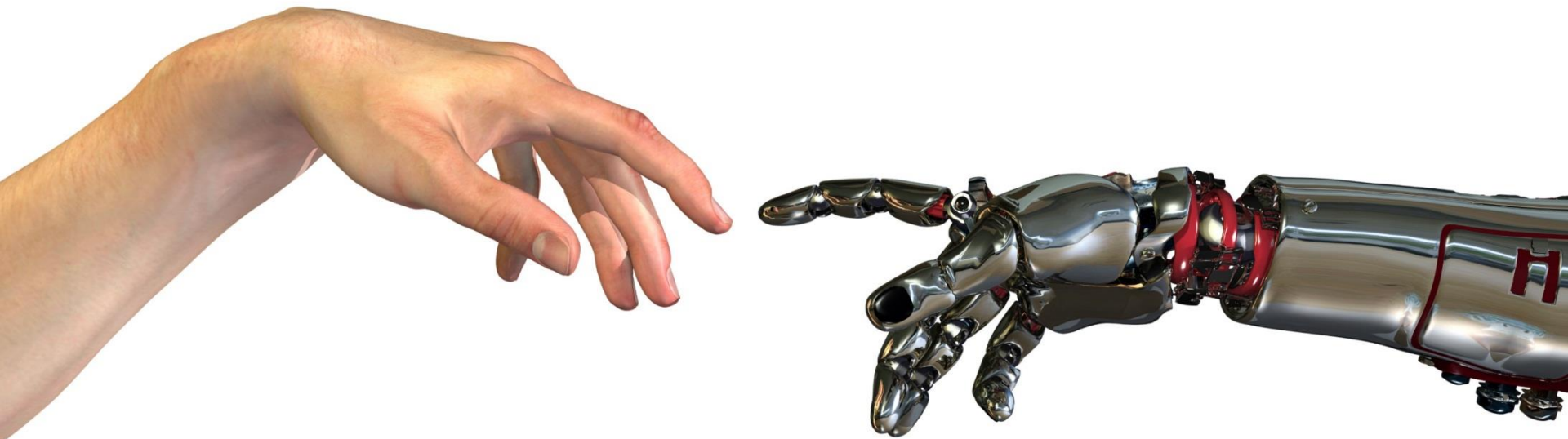
"You will not marginalise me, you will use Excel!"

Senior Manager - Output Governance meeting 2010



5





The case for an automated knowledge approach

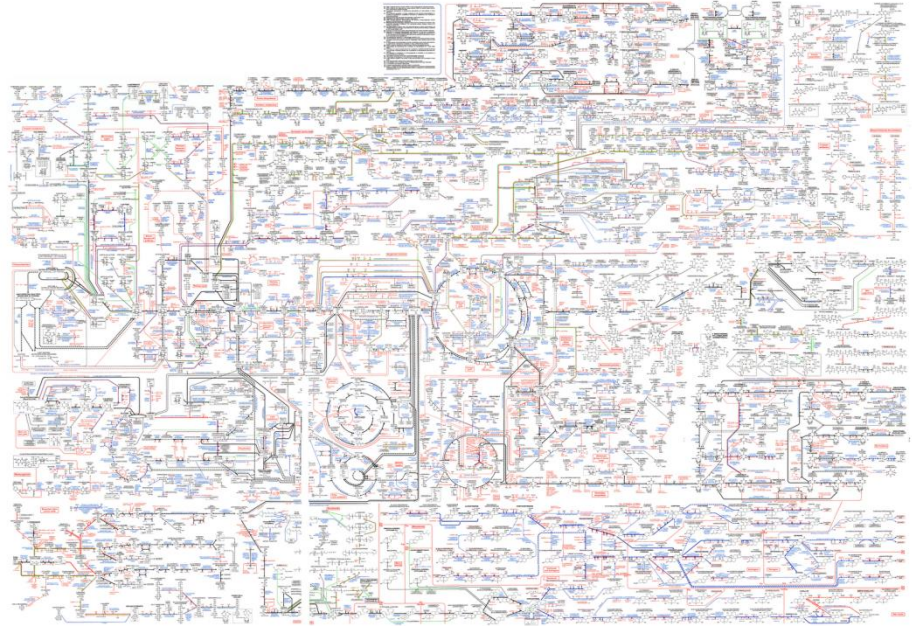
68000 disease codes

300,000 pages of legislation and guidance

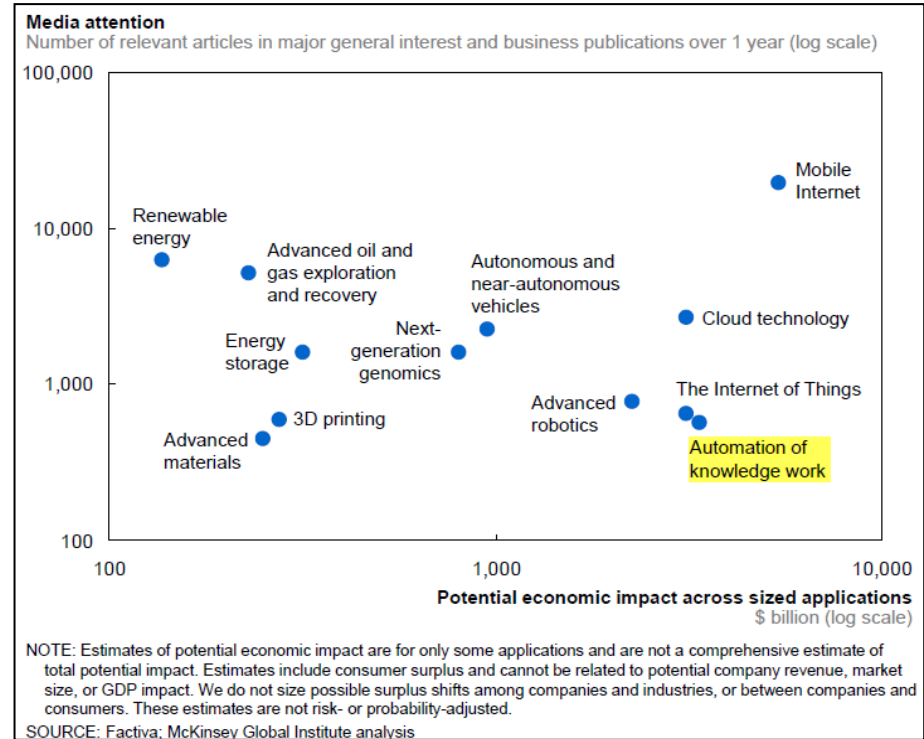
clinical data standardisation advances



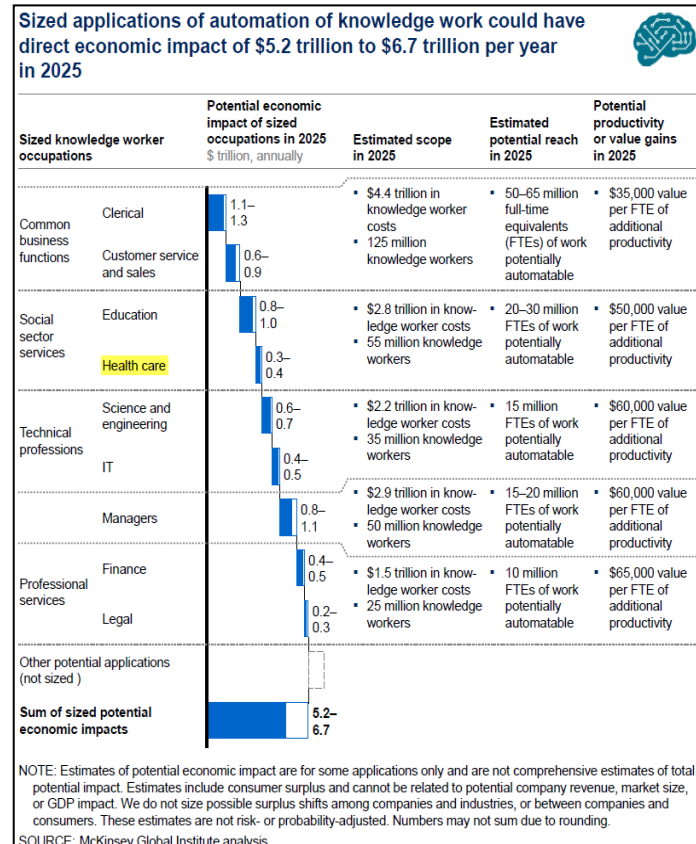
"Yet technology advancement continues to drive economic growth and, in some cases, unleash disruptive change. Economically disruptive technologies—like the semiconductor microchip, the Internet, or steam power in the Industrial Revolution—transform the way we live and work, enable new business models, and provide an opening for new players to upset the established order." (James Manyika, 2013)



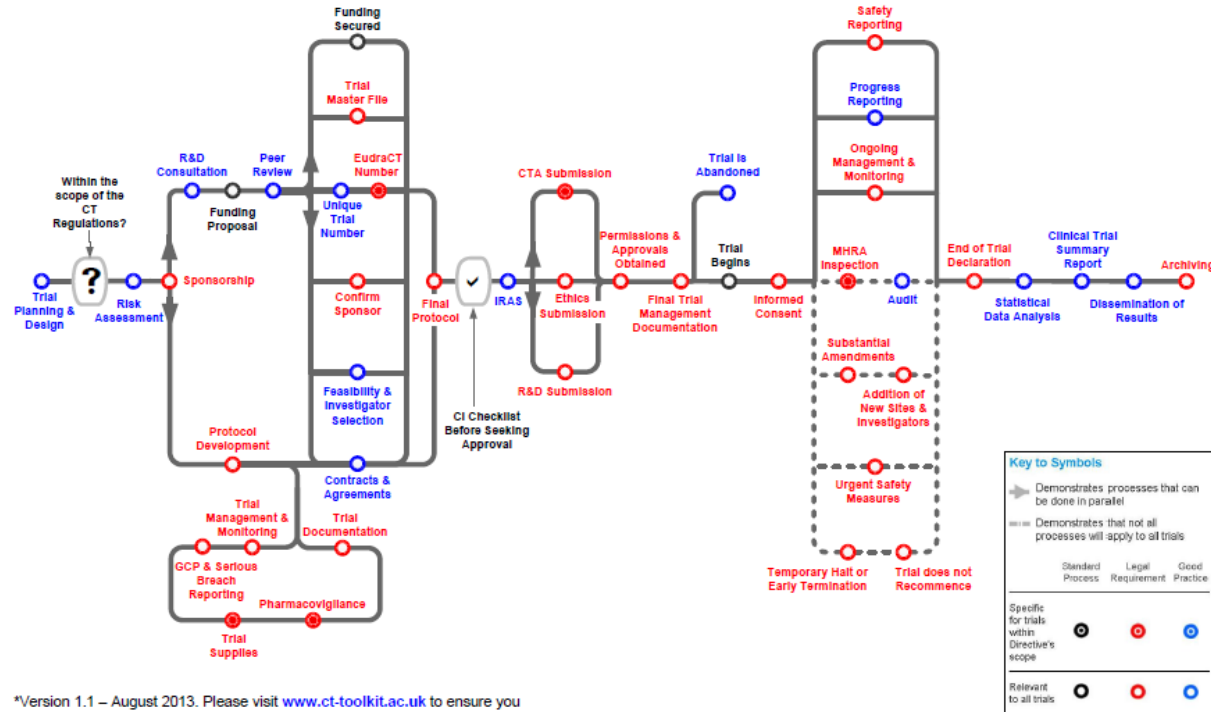
"The history of technology is littered with breathless stories of breakthroughs that never quite materialized. The hype machine can be equally misleading in what it chooses to ignore. As shows, with the exception of the mobile Internet, there is no clear relationship between the amount of talk a technology generates and its potential to create value. The lesson for leaders is to make sure that they and their advisers have the knowledge to make their own assessments based on a structured analysis involving multiple scenarios of technology advancement and potential impact."



The potential economic impact of knowledge automation tools in the types to reach \$5.2 trillion to \$6.7 trillion per year by 2025 due to greater output per knowledge worker. Their prediction includes \$4.3 trillion to \$5.6 trillion could be generated in advanced economies where wage rates are higher. With the 'Health Care' projections being in the order of 0.3-0.4 trillion USD annually.



The Clinical Trials Toolkit - Routemap



*Version 1.1 – August 2013. Please visit www.ct-toolkit.ac.uk to ensure you have the latest version of the routemap.



C.A.R.E. PREMISES

TRANSPARENCY

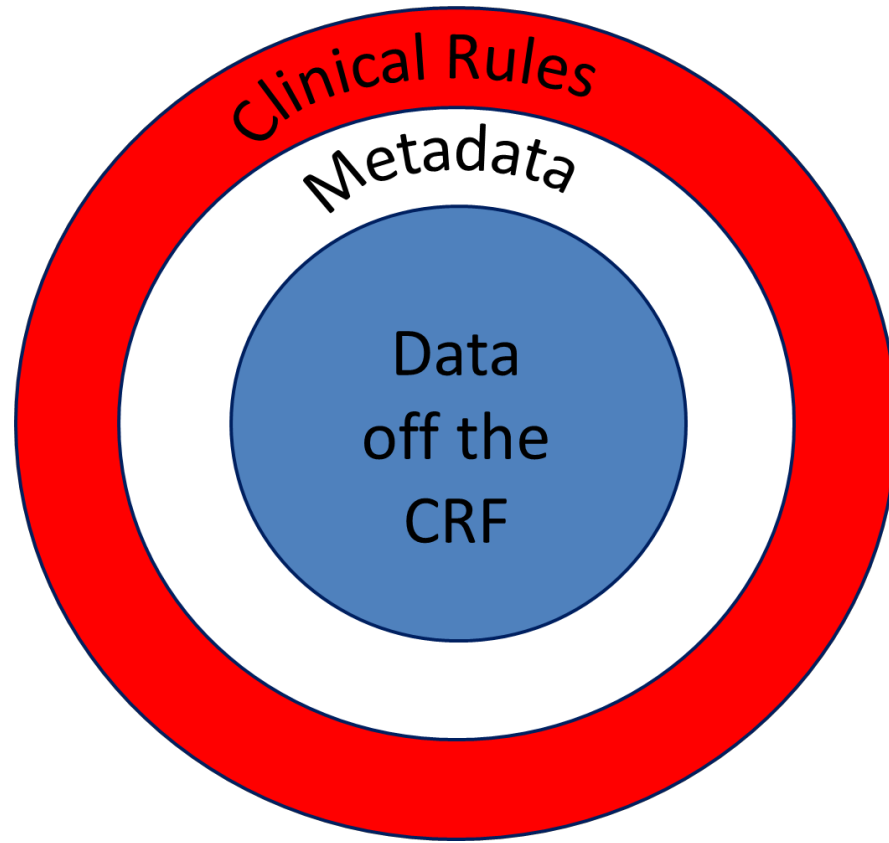
RENDER

XBRL -> XCRL

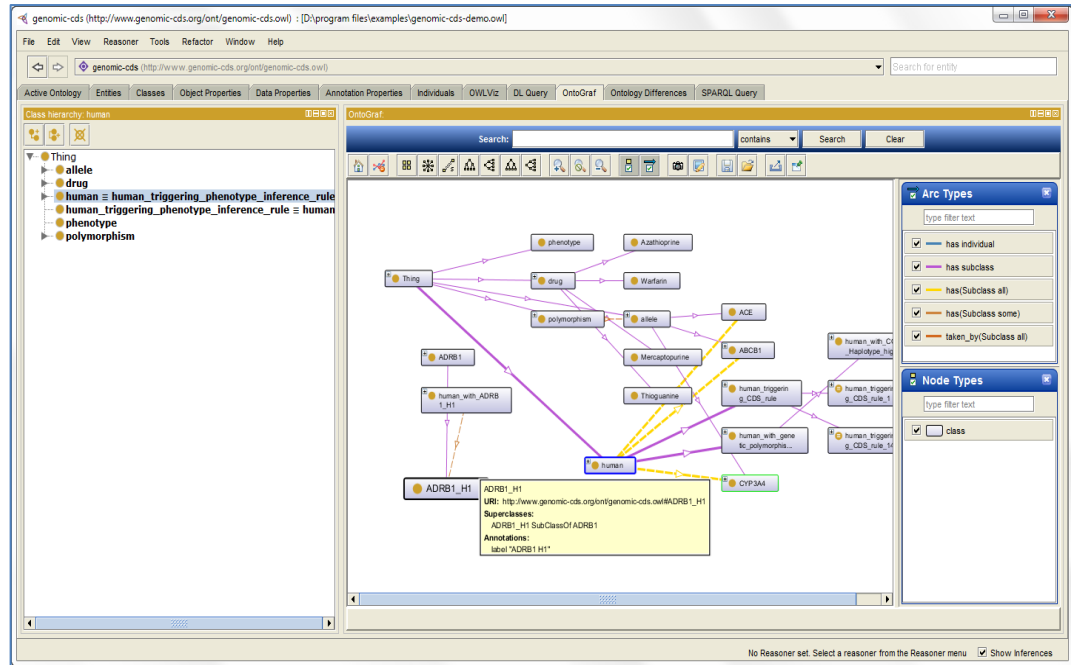
USE FAST SCALABLE TECHNOLOGY

CAPTURE THE LEGAL COMPLIANCE CONTEXT

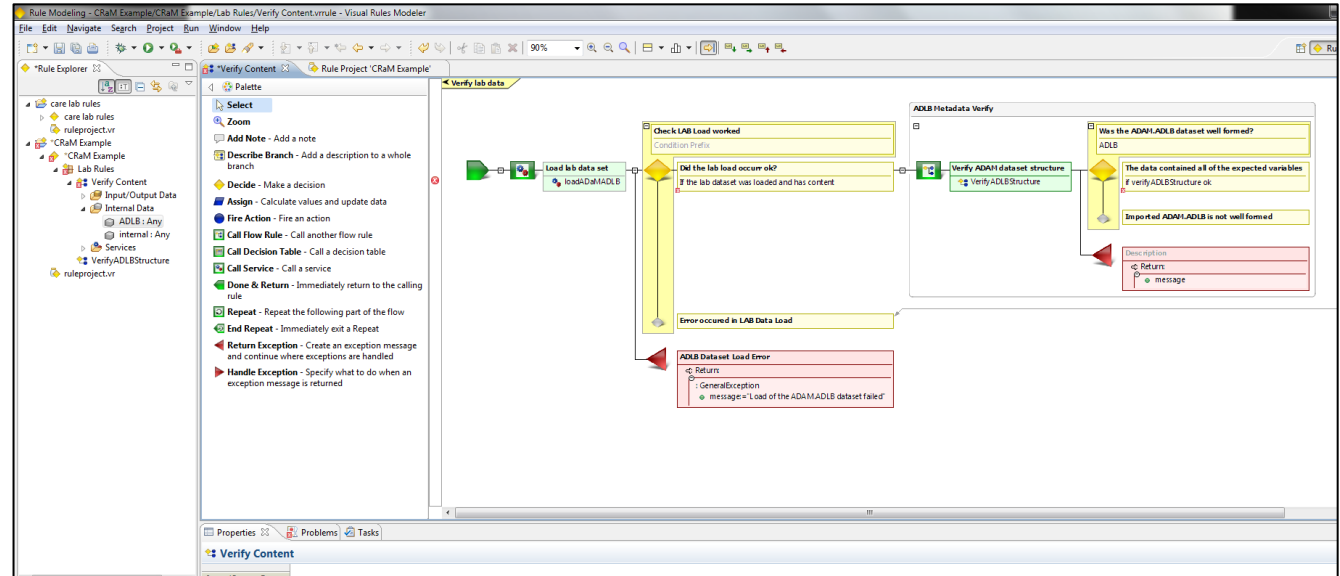




Protégé



Bosch



Intentional Software

“Businesses invest a great deal of time and expense developing software and other work products. But all too often the knowledge and insights gained disappear into the details of the code or at best only exist in documents with slender ties to the actual software.”

Charles Simonyi

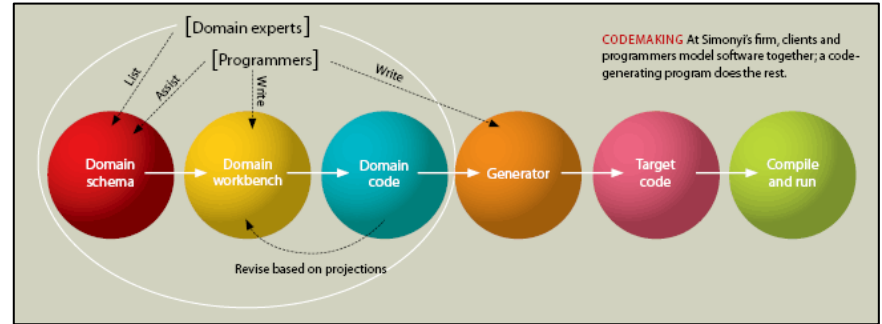
The screenshot shows the Capgemini Pension Workbench (0.0.0.0) interface. It features a 'Table of Contents' on the left, a central workspace with mathematical formulas and code, and a 'Test cases' table at the bottom. Annotations with green arrows point to specific parts of the interface:

- Data Elements**: Points to the 'Table of Contents' on the left.
- Business Rules**: Points to the 'Rules' section in the central workspace.
- Actuarial Equation**: Points to the formula $BV = (1 + olr) \cdot E \cdot \frac{a}{x - csw}$ in the central workspace.
- Calculations using defined elements**: Points to the formula $BV = (1 + olr) \cdot \frac{a}{x - csw}$ in the central workspace.
- Interactive Tests**: Points to the 'Test cases' table at the bottom.

Name	Valid time	Future	Expected value	Actual value
Part time	01/01/2004	Jan De Jong	314.4375	314.4375
Full time	01/01/2006	Jan De Jong	329.0625	329.0625
No Employment	01/01/2007	Jan De Jong	342.5298	nil



"Anything you can do I can do meta."
Charles Simonyi



Summary of Effort & Results:

Captured 3,000+ sentences concerning cellular biology from Campbell's Biology textbook

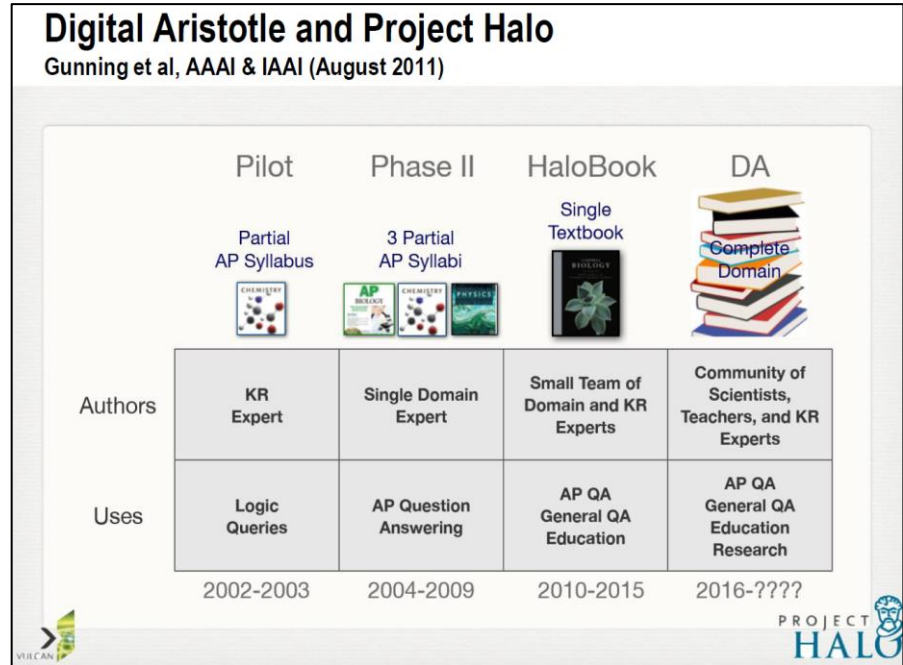
Sentence length averaged 10 words up to 25 words

hundreds of parses common, > 30 per sentence on average

the correct parse is typically not the parse ranked best by statistical NLP

sentences disambiguated and formalized into logic in very few minutes on average

resulting logic is typically more sophisticated than skilled logicians typically produce



Sentences (2322) Relations		all types	axiomatic	all editors	7 / 1/2012
Text	Axiom				
The environment of a cell is the solution surrounding it.	$\forall (x8) \text{cell}(x8) \Rightarrow \forall (x6) \text{environment}(\text{of}(x8))(x6) \Rightarrow \text{solution}(x6) \wedge \text{surround}(x6, x8)$				
Enzymes are produced by cells.	$\forall (x5) \text{enzyme}(x5) \Rightarrow \exists (x8) (\text{cell}(x8) \wedge \text{produce}(x8, x5))$				
An enzyme is a complex protein.	$\forall (x6) \text{enzyme}(x6) \Rightarrow \text{complex}(\text{protein})(x6)$				
The endoplasmic reticulum is an organelle of cells in eukaryotic organisms.	$\forall (x6) \text{endoplasmic}(\text{reticulum})(x6) \Rightarrow \exists (x19) (\text{eukaryotic}(\text{organism})(x19) \wedge \exists (x14) (\text{cell}(\text{in}(x19))(x14) \wedge \text{organism}(x19, x14) \wedge \text{organelle}(x6, x14)))$				
A eukaryotic cell is not a prokaryotic cell.	$\neg (\exists (x6) (\text{eukaryotic}(\text{cell})(x6) \wedge \text{prokaryotic}(\text{cell})(x6)))$				
Diffusion is a result of the constant motion of molecules.	$\forall (x8) \text{molecule}(x8) \Rightarrow \forall (x5) \text{diffusion}(\text{of}(x5, x8)) \Rightarrow \forall (x18) \text{constant}(\text{vibration})(\text{of}(x8)) (x18) \Rightarrow \text{result}(\text{of}(x5, x8), x18)$				
Cholesterol is a steroid.	$x5 = \text{cholesterol} \rightarrow x5 = \text{steroid}$				

Coherent Knowledge Systems



SAS sas7bdat Consumer class in C++, based upon
the work of Matt Shotwell in R.



Matthew Shotwell, Ph.D.



Benjamin Grosf, PhD

Rulelog implementation in the
Clinical Analysis and Reporting
context.



