

CARE to CRaM

Greg Fuller, GIANCE SARL, Mulhouse, France

ABSTRACT

The Pharmaceutical industry today continues to face a number of challenges: the patent cliff, increasing competition from generics, the need to drive shareholder value, while continuing our mission to provide better solutions and medicines that improve people's lives. What this means to the process of clinical data analysis and reporting is that there is an imperative to adopt significantly more refined systemic solutions to dramatically improve cost efficiency, whilst retaining appropriately high standards of accuracy, repeatability, safety and compliance. This paper elaborates on the poster 'Clinical Analysis and Reporting Environment (CARE) – The Industrialisation of Clinical Trial Analysis and Reporting', presented at the 2013 FDA/PhUSE conference, taking a deeper look into the critical success factors involved in the architecting and building the Clinical Analysis and Reporting Environment (C.A.R.E.) framework. A technical view of the essential design of the foundation components of the Clinical Rules and Metadata (C.R.a.M.) is provided.

Introduction

"We cannot solve our problems with the same thinking we used when we created them." - Albert Einstein
"When you come to a fork in the road, take it!" -Yogi Berra

The initial approach to this paper was to present the details of my work with Sparx systems EA and C++ implementations in the development and implementation of the core components and operation of the C.A.R.E. Framework. However, given the nature of the forum for this paper and intended audience. I concluded that it would be more appropriate to step back and present details of the key influencers to the C.A.R.E. development trajectory. To name a few, Protégé, BOSCH, Intentional Software and last but by no means least Coherent Knowledge Systems. The technical specifications and resources will be made available at the GOOGLE+ C.A.R.E. community site. **(COMMUNITY, 2013)**

I have worked rather oxymoronically, as an independent SAS Application Developer on five different Clinical Analysis and Reporting applications for various Pharmaceutical companies. In each and every case, I came to understand that these applications have in many ways failed to deliver sufficient or significant improvements. There are a number of reasons for this, probably the most significant being an over reliance on tool sets that have no place as application development components. With many solutions relying upon large and complicated SAS/Macro libraries, as the cornerstone components of these solutions.

What are some of the reasons for the observation that the basis of Clinical Statistical Analysis and Reporting has remained pretty static for so many years, despite the rather significant corporate financial investments? Sebastian and Völter (Markus Voelter, 2013) nicely express some of the challenges being faced in this area, in particular the idea of the "Investment Prison". "The more you invest in reusable artifacts, the more productive you become. However, you may also get locked into a particular way of doing things. Radically altering the way we approach the C.A.R.E. domain may appear unattractive – given that we have become relatively efficient with our current approach. It becomes expensive to "move outside the box". Additionally, the idea of a there being a "Cultural Bias" (Manns, 2004) is another factor. Specifically, calling into question the strategy of relying upon Statistician or SAS Statistical Programmer being appropriately educated or experienced for the technical IT design and implementation tasks of these projects. From my experience, it does explain why SAS Scripts, Microsoft Excel, Microsoft PowerPoint, Microsoft Word, PDF files figure heavily in most current industry solutions in this domain.

This paper will work present a high level case for the strategically essential move to an Automated Knowledge perspective and examine some modern technology approaches that support such a move, outline some key premises of the C.A.R.E solution below and to present in detail the work in the area of Clinical Rules and Metadata(C.R.a.M.). An overarching concept behind the C.A.R.E. framework is to implement an automated

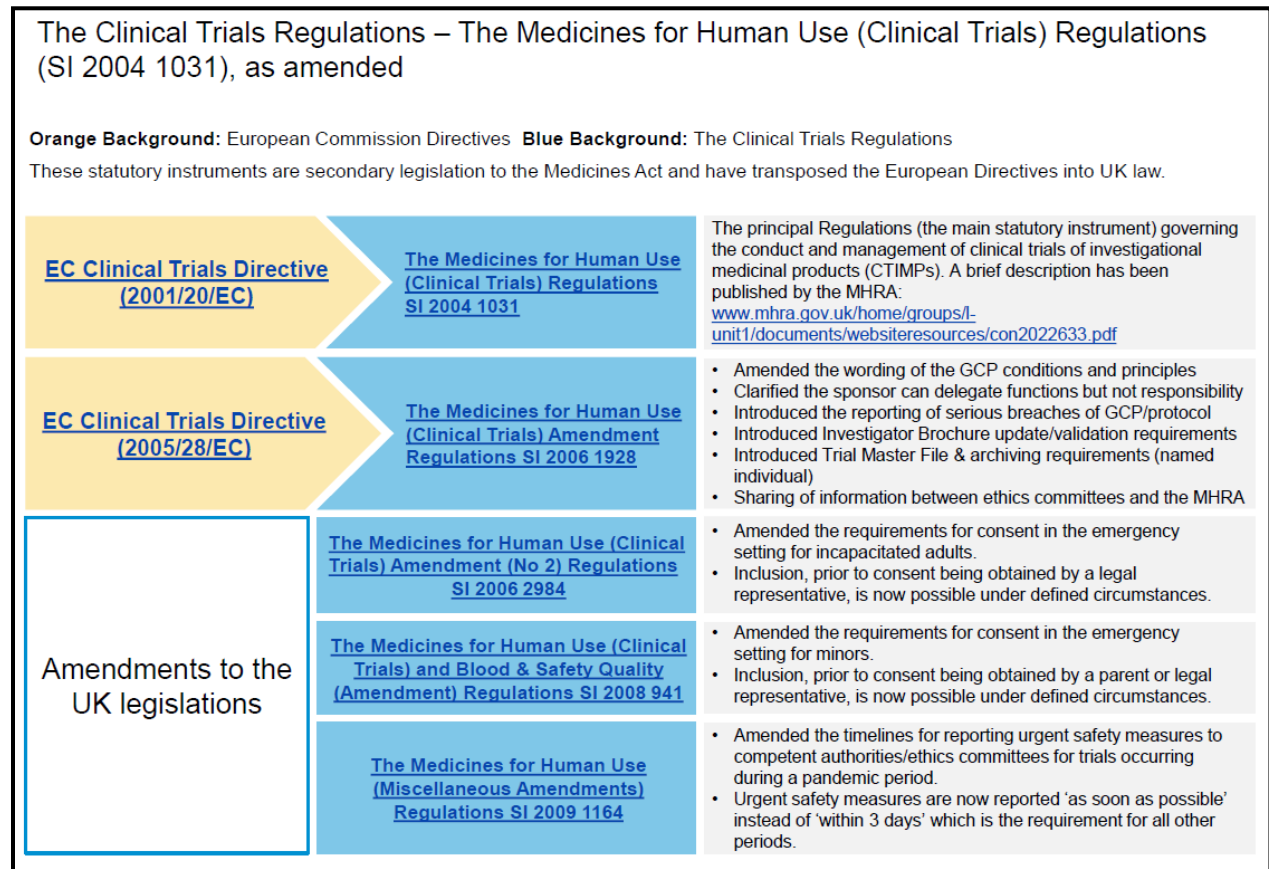
PhUSE 2013

knowledge based platform for Clinical Trial Analysis and Reporting. Considering the nature of the C.A.R.E. domain, specifically, statistical programming tasks the following make the premise of a C.A.R.E. approach viable (Rudolph, 2008):

1. there is a significant level of conditional branching within the clinical analysis process;
2. there is the possibility for complex decision making process;
3. the clinical logic/rules are relatively static;
4. clinical logic/rules can be applied across trials; and
5. there is the possibility for trial simulation and prototyping.

A CASE FOR AN AUTOMATED KNOWLEDGE APPROACH.

So what is the case for adopting a Knowledge based approach? Taking a step back, the Pharmaceutical industry and more specifically the Clinical Trial Analysis and Reporting process is highly regulated. By way of example, we have European wide legislation that is usually augmented by each of the member states into local legislation.



1. Overview of Clinical Trial Regulations UK (Archives)

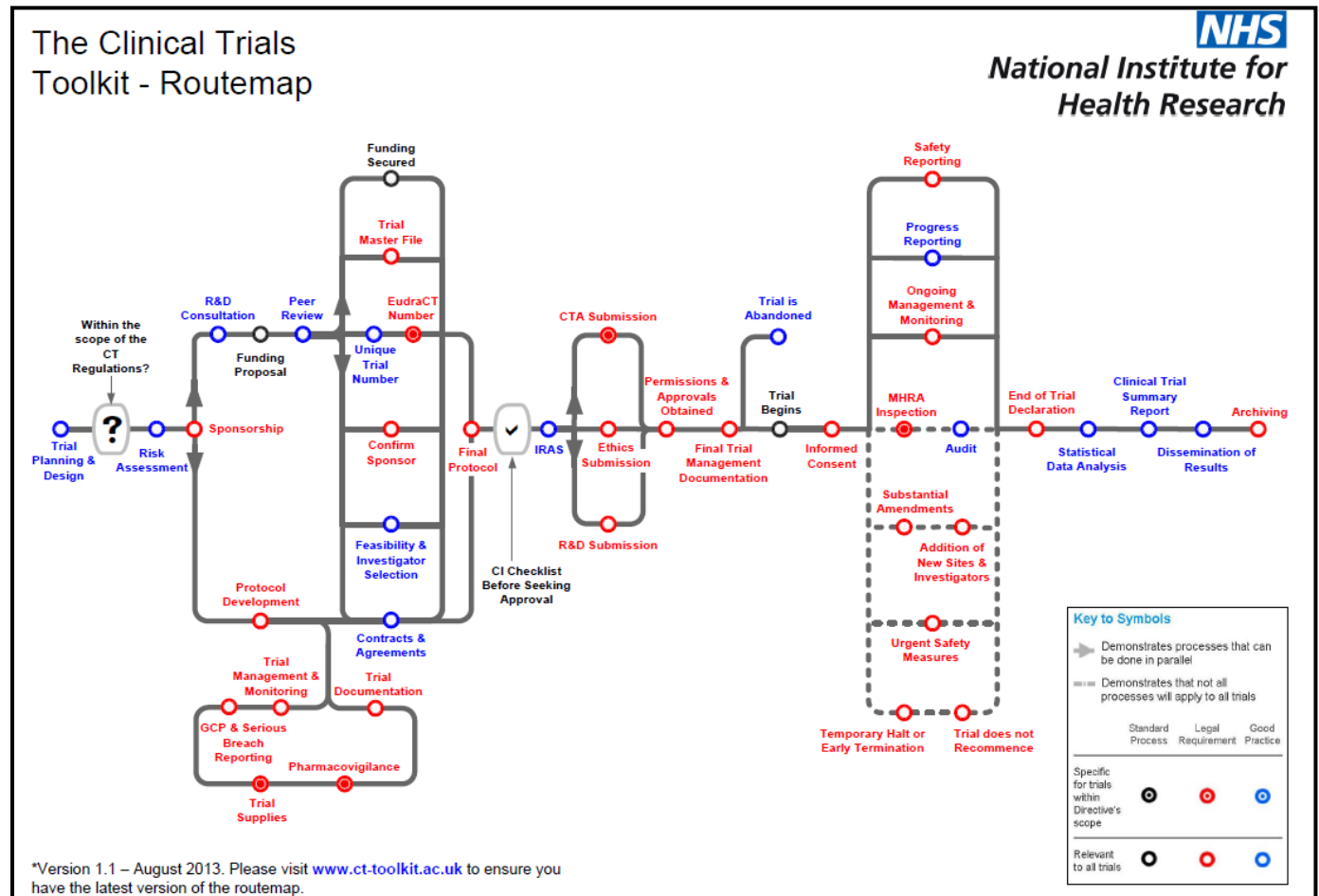
In thinking about the scope of the C.A.R.E. project. There is a clear requirement for traceability and transparency (see below C.A.R.E. Solution KEY Premises). In order to get a feeling for the size and depth, I looked at the issues from several perspectives. Namely those of patient, regulatory and pharmaceutical companies' perspectives.

Of most criticality, is the patient perspective, to get a sense of the size of this dimension I recommend the WHO Family of International Classifications:

1. The International Classification of Diseases (ICD) is the standard diagnostic tool for epidemiology, health management and clinical purposes. International Statistical Classification of Diseases and Related Health Problems 10th Revision (WHO, 2013) contains approximately 68,000 disease codes.
2. International Classification of Functioning, Disability and Health (ICF) is a classification of health and health-related domains. These domains are classified from body, individual and societal perspectives by means of two lists: a list of body functions and structure, and a list of domains of activity and participation. Since an individual's functioning and disability occurs in a context, the ICF also includes a list of environmental factors. (WHO)

- International Classification of Health Interventions (ICHI) The purpose of this classification is to provide Member States, health care service providers and organizers, and researchers with a common tool for reporting and analysing the distribution and evolution of health interventions for statistical purposes.

From a regulatory perspective, I chose to look to the site provided by the U.K. NHS National Institute for Health Research (National Institute for Health Research, 2013). The Toolkit is primarily focused on Clinical Trials of Investigational Medicinal Products (CTIMPs) and the regulatory environment and requirements associated with these. For the purpose of this paper it served very well to identify the links between the legislative requirements and clinical trials.



1. An example of a Clinical Trial Roadmap.

Drilling down into this route map reveals a consider body of information in the form of legislation, guidance, best practices, checklists and further references. Another very good source of information in this International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (ICH, 2013) whose mission "is to make recommendations towards achieving greater harmonisation in the interpretation and application of technical guidelines and requirements for pharmaceutical product registration, thereby reducing or obviating duplication of testing carried out during the research and development of new human medicines." (ICH, 2013)

From the Pharmaceutical Company perspective there are a set of Clinical Trials being conducted possibly in an international context across multiple indications at different phases, by a globally dispersed set of Clinicians, Medical Writers, Statisticians, Data Managers and Statistical Programmers. All of whom are fully trained and certified with the companies Policies, SOPs and Guides. As a conservative estimate at the very least there are over 300,000 pages¹ of relevant legislature and guidance's. Although, the overall technical approach for implementing a trial design

¹ A rough estimate based on the results from a modified web crawler (wikipedia, 2013).

PhUSE 2013

remains unchanged. There is an opportunity with the recent advances in the standardisation of clinical data, with the growing acceptance and use of CDISC Standards. The increase in sophistication in the approach to clinical data capture also bolsters this clinical trial analysis and reporting efficiency drive.

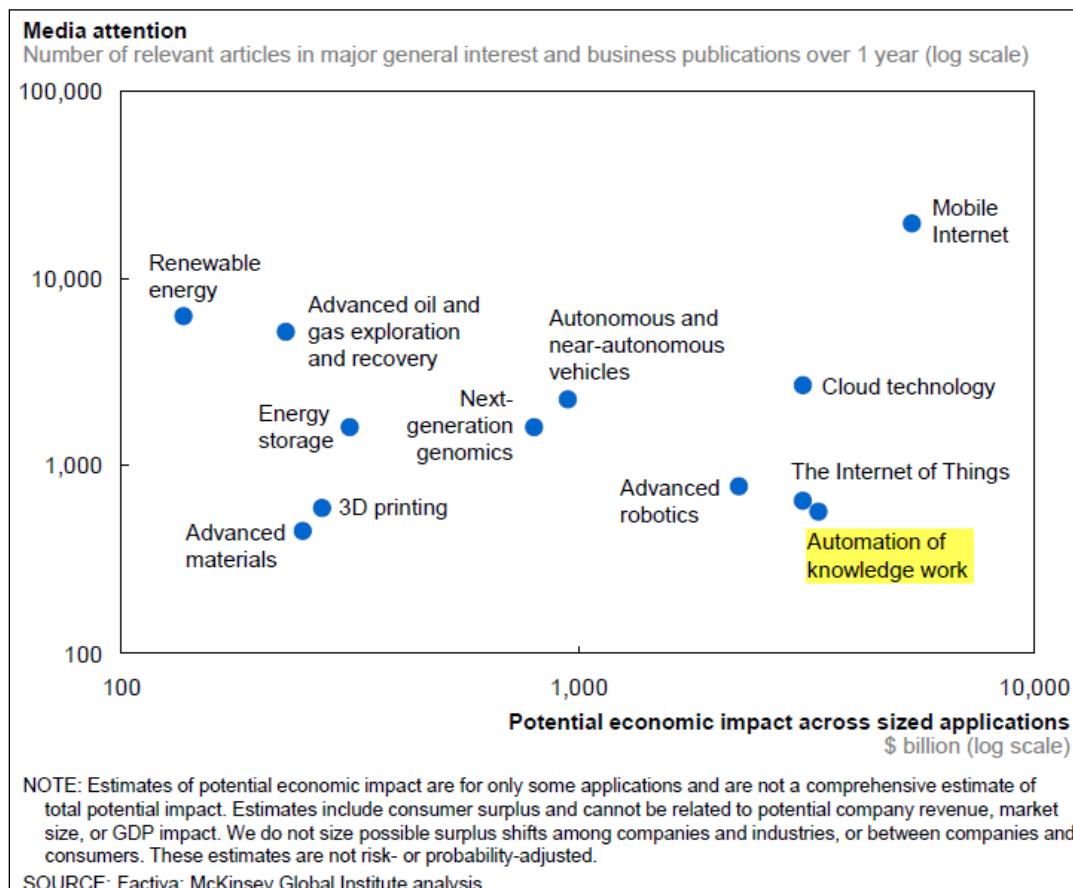
Looking at a report from McKinsey Global Institute “Disruptive technologies: Advances that will transform life, business, and the global economy” (James Manyika, 2013) which examines 12 technological areas with the potential for disruption.

“Technology is moving so quickly, and in so many directions, that it becomes challenging to even pay attention—we are victims of “next new thing” fatigue. 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. Business leaders and policy makers need to identify potentially disruptive technologies, and carefully consider their potential, before these technologies begin to exert their disruptive powers in the economy and society.” (James Manyika, 2013)

The McKinsey report attempts to draw a link between hype and potential. It makes the point that emerging technologies often receive a great deal of notice, whilst the majority goes mostly unnoticed.

“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 (James Manyika, Insights & Publications, 2013) 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.”

This estimate shows that, although the level of hype associated with knowledge automation is low the predicted benefit is extremely high.

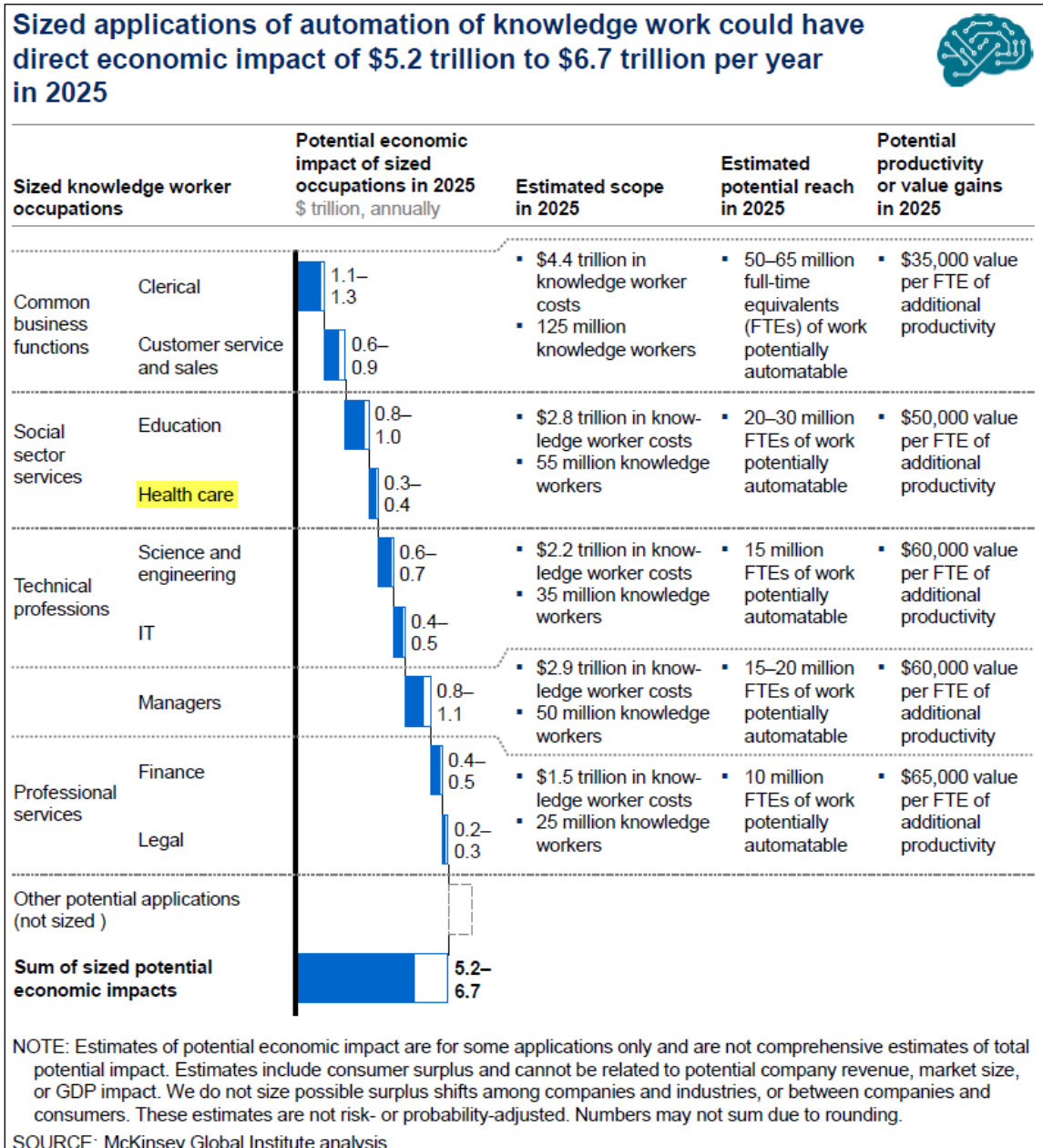


2. The relationship between hype about technology and its potential impact

PhUSE 2013

(James Manyika, Insights & Publications, 2013)

The report (James Manyika, Insights & Publications, 2013) also examines 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.



3. Industry Breakdown for Knowledge Automation. (James Manyika, Insights & Publications, 2013)

Taken together – the complexity and range of the C.A.R.E. backdrop of regulatory compliance – the size of the potential Pharmaceutical market and as we shall see the maturation of some key knowledge based approaches. I would conclude that it is vital to in our industry that we embrace a shift toward an automated knowledge based solution outlook.

C.A.R.E. Solution KEY Premises

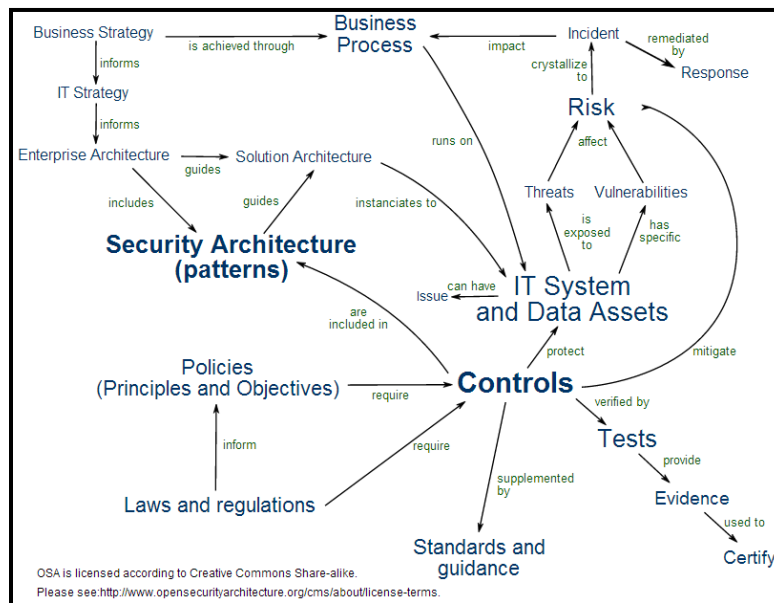
"If you don't know where you're going, chances are you will end up somewhere else." - Yogi Berra

Premise - A proposition upon which an argument is based or from which a conclusion is drawn.

The following is by no means exhaustive, “am Expressio unius est exclusio alterius” does not apply.

TRANSPARENCY

An essential aspect of the C.A.R.E. framework is the need to address security, traceability and audit. Of particular utility is the contribution by the Open Security Architecture Group OSA. “OSA is licensed in accordance with Creative Commons Share-alike. We believe that Open Source principles result in more secure systems, and want the computing architectures that we depend on for our daily lives to be as secure and reliable as possible” (OSA, 2013). Of particular use has been the OSA Taxonomy, shown below, which depicts the entities and relationships that are relevant for OSA. This taxonomy has been very helpful helps to understand how OSA can be applied within the C.A.R.E. framework. OSA is very much a living Taxonomy and there is an active support community.



4. OSA Taxonomy

For some operational aspects of the C.A.R.E. framework, for the event logging sub-system, the work of Extensible Event Stream(XES) (Verbeek, 2012) is of keen interest. XES is an XML-based standard for event logs. The purpose of XES is to define a standard format for the interchange of event log data between tools and application domains. Its primary purpose is for process mining, i.e. the analysis of operational processes based on their event logs.

RENDER

As soon as any information, whether it is a SOP, Clinical Output specification or Protocol/Trial Design, Clinical data is stored within an Excel, Word, PowerPoint, PDF document or SAS Script, then that information is mostly lost for further system processing. The link to the knowledge and associated derivation processing required to arrive at these documents becomes difficult and usually in a business context impossible to reverse engineer. Although it is an obvious requirement that these artifacts are produced, it is essential that these are rendered by the system from appropriate system internal data structures and processes. Think about the amount of task repetition and time spent hunting for information required in our current work. The current practice of writing SAS scripts should be thought of as an act of knowledge hiding.

XBRL -> XCRL

XBRL (eXtensible Business Reporting Language) is a freely available and global standard for exchanging business information. XBRL allows the expression of semantic meaning commonly required in business reporting. The language is XML-based and uses the XML syntax and related XML technologies such as XML Schema, XLink, XPath, and Namespaces. One use of XBRL is to define and exchange financial information, such as a financial statement. The XBRL Specification is developed and published by XBRL International, Inc. (XII). (Wikipedia, 2013) Within the

PhUSE 2013

financial sector there a number of the large foreign companies which use International Financial Reporting Standards (IFRS) and that will also be submitting their financial returns to the SEC using XBRL. (U.S. Securities and Exchange Commission, 2013). Although, I would suggest that within the Pharmaceutical Industry the Clinical Reporting requirements are not to be compared with those of the financial sector – the overall strategy is a good one offering clear benefits to both the industry and the regulators. To satisfy overall trial submission requirements any specification of a XCRL (Extensible Clinical Reporting Language) standard would be significantly different to what is seen in other sectors.

To get an idea of the structure of XBRL, there is a simple example available that uses an Excel spreadsheet to demonstrate the key concepts of the XBRL standard; you can find this “hello world” example at (Digital Financial Reporting, 2008).

USE FAST SCALABLE TECHNOLOGY.

SAS is a best of breed tool for Statistical Analysis and Reporting. The idea that the power of SAS, SAS macro and SAS SCL scripting is being used in application development is testimony to the products power and flexibility. This does not mean that this is the best choice of tool for the job. The C.A.R.E. framework is being built using C++ for benefits of speed and scalability. As a starting point the idea of using a desktop virtualization solution is common to most of us and for the moment sidesteps the issues of delivering a web based solution. I tend to shy away from Eclipse (Eclipse Org, 2013) plug in solutions, because although it has advantages in delivering applications quickly, it seems to be an iceberg solution, with most of the system dependencies and implementation issues hidden below the surface.

CAPTURE THE LEGAL COMPLIANCE CONTEXT

Natural Language Processing (NLP) is a field of computer science and engineering that has developed from the study of language and computational linguistics within the field of Artificial Intelligence. The goals of NLP are to design and build applications that facilitate human interaction with machines and other devices through the use of natural language.

The most common method for evaluating the performance of a NLP algorithm is to calculate how accurately it labels your dataset. This can be done by measuring the fraction of the results from the dataset that are labeled correctly using a standard technique of “relevance judgment” called the Precision and Recall metric.

For each label you are using to identify elements in the data, the dataset is divided into two subsets: one that is labeled “relevant” to the label, and one that is not relevant. Precision is a metric that is computed as the fraction of the correct instances from those that the algorithm labeled as being in the relevant subset. Recall is computed as the fraction of correct items among those that actually belong to the relevant subset. The following confusion matrix helps illustrate how this works:

	positive	negative
positive true	positive (tp)	false negative (fn)
negative false	positive (fp)	true negative (tn)

Precision is a metric that is computed as the fraction of the correct instances from those that the algorithm labeled as being in the relevant subset. Recall is computed as the fraction of correct items among those that actually belong to the relevant subset.

Precision	$P=tp/(tp+fp)$
Recall	$R=tp/(tp+fn)$
Accuracy	$A=(tp+tn)/(tp+tn+fp+fn)$

The values of Precision(P) and Recall(R) are combined into a single metric called the F-score, which is the harmonic mean of the two. $F=2*(P*R/(P+R))$ While the F-measure won't tell you specifically where your algorithm is producing incorrect results, it does provide a handy way for someone to tell at a glance how your algorithm is performing in general. This is why F-measure is commonly reported in computational linguistics papers: not because it's inherently useful to people trying to find ways to improve their algorithm, but because it's easier on the reader than processing an entire table of precision and recall scores for each tag used. (Pustejovsky & Stubbs, 2012)

The process of automatically generating input parsers using a Bayesian generative model reported F-Score of 80% against state of the art semantic parsing with F-Scores of 66.7% Tao Lei(et al.) (Tao Lei, 2013). Their C++ source is available at (Tao Lei F. L., 2013). Also, the work by, for example Estrella project whose objectives are to develop a Legal Knowledge Interchange Format (LKIF), building upon emerging XML-based standards of the Semantic Web, including RDF and OWL, and Application Programmer Interfaces (APIs) for interacting with legal knowledge-based systems. (University of Amsterdam, 2008)

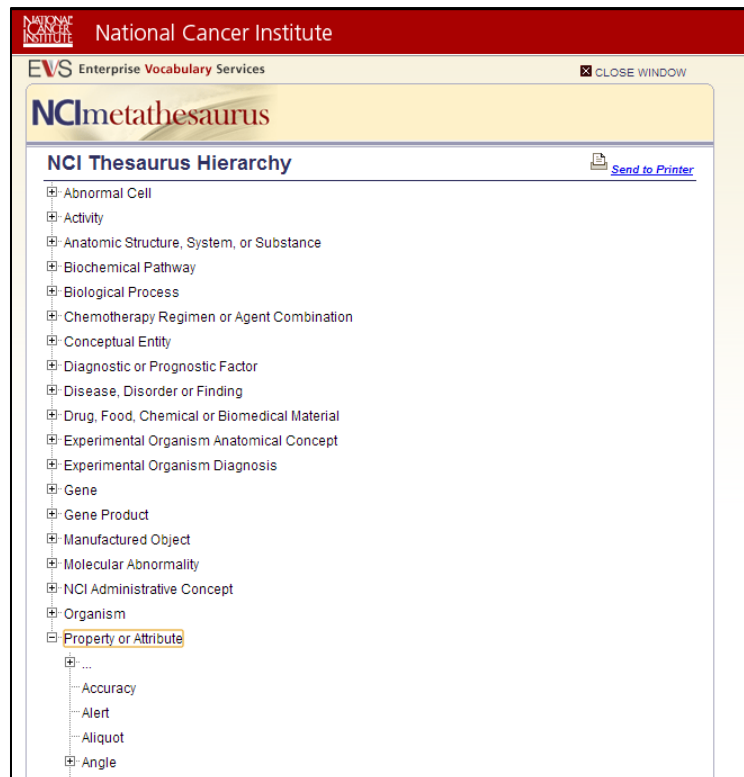
Would it be reasonable to use NLP to capture the Industries Global Regulatory Obligations and Guidance's in a form that could be consumed systemically by Pharmaceutical Companies?

CLINICAL RULES AND METADATA

Essential to the C.A.R.E. Framework is the necessity to tightly couple the clinical metadata and rules associated with the creation of validated statistical outputs. Many of the current corporate solutions include a Metadata Repository and some form of controlled terminology aspect. A significant piece of the solution that is missing is a comprehensive and integrated clinical rule repository. Currently, this considerable body of this knowledge is being actively locked away within SAS Scripts. The following is an overview of the some potential solution approaches, where each has interesting and compelling facets, which I would consider important for inclusion in any clinical analysis/reporting knowledge based solution.

PROTÉGÉ AND OWL

Protégé (Protégé, 2013) is a free, open source ontology editor and knowledge-based framework. The Protégé platform supports modeling ontologies via a web client or a desktop client. Protégé ontologies can be developed in a variety of formats including OWL, RDF(S), and XML Schema. By way of example OWL is used as part of the creation of the NCIthesaurus, NCI metathesaurus and SNOMED. (<http://www.cancer.gov/>, 2013)

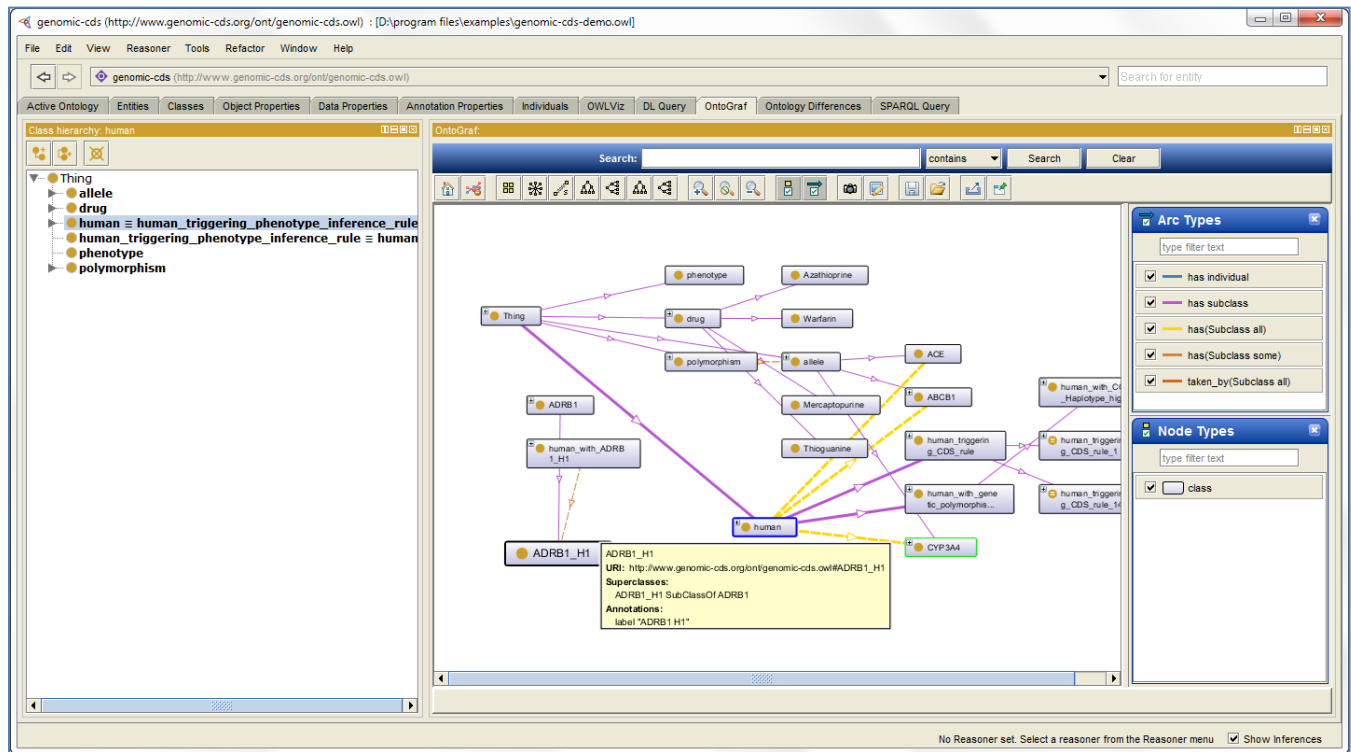


5. Owl is part of the workflow for NCI metathesaurus

The OWL ontology is a set of axioms, which provide explicit logical assertions about three types of things - classes, individuals and properties. By using a piece of software called a reasoner one can infer other facts which are implicitly contained in the ontology. The following example "Preview of OWL 2 ontology for clinical pharmacogenomics and decision support (now called 'Genomic CDS') can easily be loaded and explored with a local installation of Protégé. The Genomic CDS ontology that aims to unify several functionalities in a single resource, these being:

- a knowledge base for clinical pharmacogenomics that can be used for question;
- a rule base for clinical decision; and
- a tool for checking data consistency; (Samwald, 2012)

PhUSE 2013

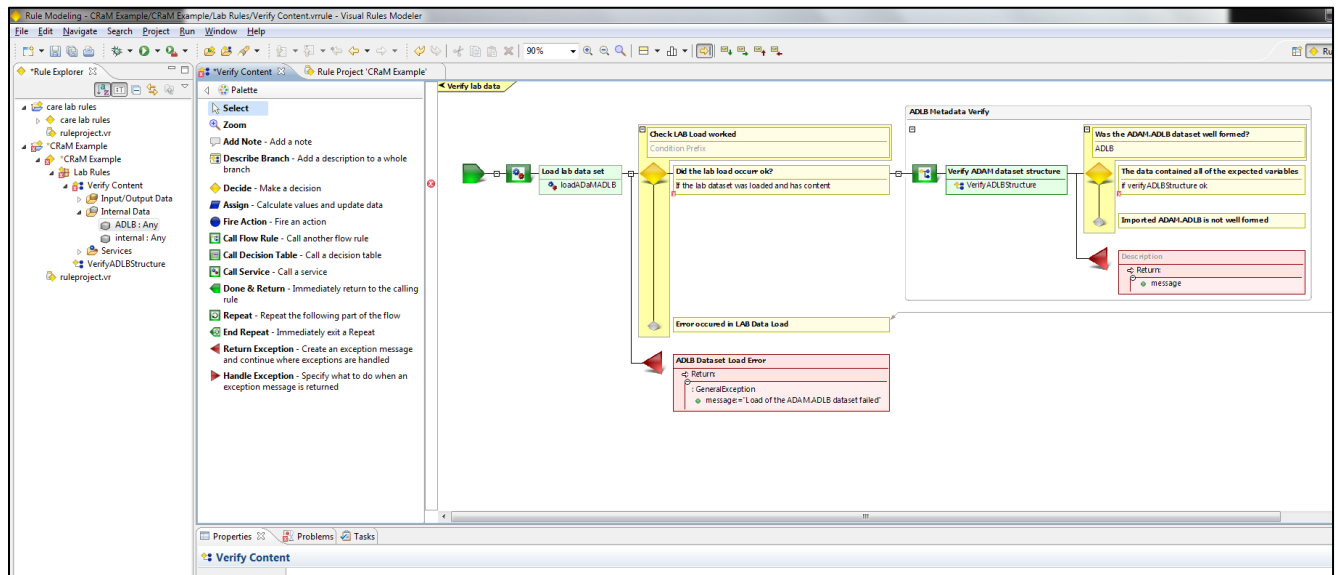


6. Import and exploration of 'Genomic CDS' into Protégé

Protégé (Protégé, 2013) is a feature rich and deep product. It figures in a number of organisations' implementation workflows. It does require some considerable efforts to get to grips with and at times it can be rather bug infested, at least my windows 7 installation. There are a number of very basic tutorials online – regarding Pizza ontologies. The products component of ontology management and representation is rather a compelling component.

BOSCH INVENTED FOR LIFE

“Visual Rules Modeller” (Bosch Sotare innovations GmbH, 2013) solution has a lot of functionality, that is of some interest and relevance. Putting aside the issue that it is built on top of eclipse, the learning curve when compare with tools such as Protégé is a less steep and offers a considerable amount of useful business functionality.



7. Example of Visual Rules interface.

So what aspects of the Visual Rules Modeller are relevant to the clinical analysis and reporting business domain?

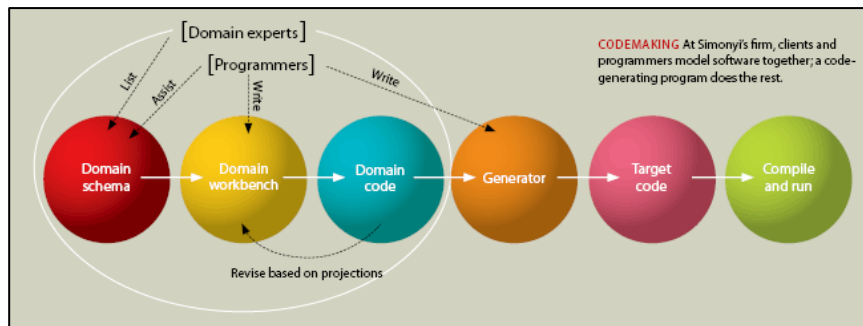
PhUSE 2013

The rule specification process is visual and clinical metadata is integrated and easily used during the clinical rule specification process. The application dynamically performs background checking as the rules are being constructed so that issues are highlighted in real time. Very much like a modern IDE with syntax checking. Significantly, rules can be created separately and imbedded in a larger rule flow. The product also demonstrates a JAVA code generation capability that allows the tool to be integrated into existing application software. Finally, the capability to manage how rules change over time, rule versioning is of utility.

INTENTIONAL SOFTWARE

In the article “Anything you can do I can do Meta” (Rosenberg, 2007) Charles Simonyi (Intentional Software, 2013) shares his dissatisfaction with software. “Software as we know it is the bottleneck on the digital horn of plenty,” he says. “It takes up tremendous resources in talent and time. It’s disappointing and hard to change. It blocks innovation in many organizations.” and “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.”

Simonyi argues that his approach solves several of software engineering’s most persistent problems. Programmers today, he often says, are “unwitting cryptographers”: they gather requirements and knowledge from their clients and then, literally, hide that valuable information in a mountain of implementation detail—that is, of code. He provides a useful and interesting overview at the Computer Science and Engineering COLLOQUIA – 2008 “The Intentional Domain Workbench” [<http://uwtv.org/series/17392605/watch/16205827/>]

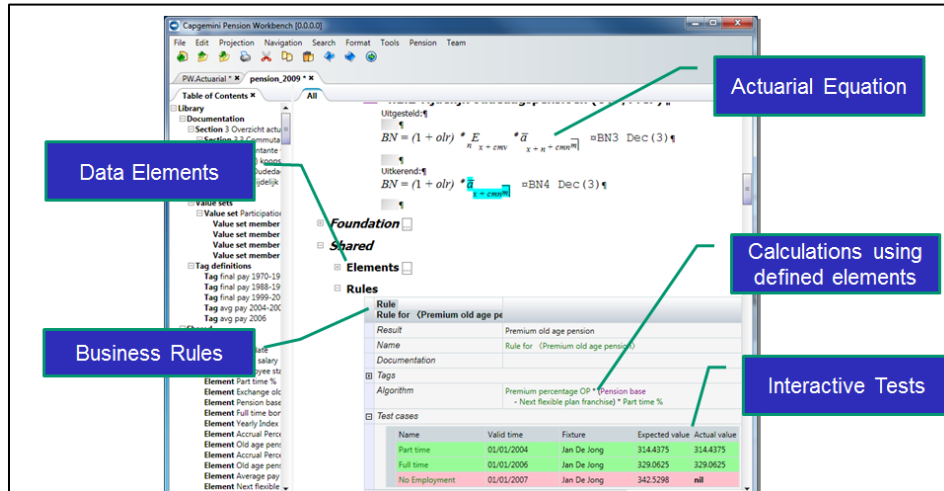


8. Overview of Simony's firm's approach (MIT Technology Review)

Intentional provides a case study presented on the use of the Intentional Software’s Domain Workbench applied to the Life Insurance and Pension planning.

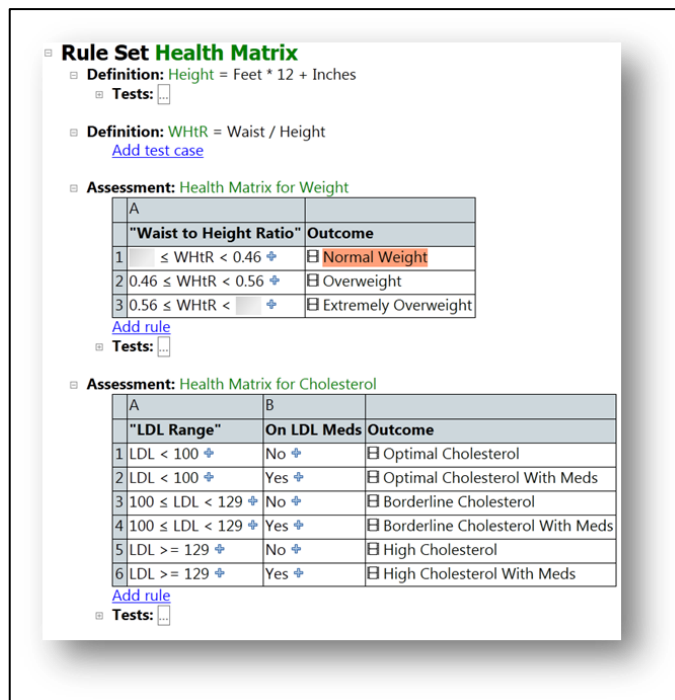
“The pension insurance domain is extremely complex with a typical insurer having multiple pension plan products, each having hundreds of plan specific rules with a stringent regulatory environment. Each plan might have between thousands and millions of insured participants, where each insured has up to 40 years of personal employment data history.” The business knowledge about Pension Plan administration is encoded in the Pension Workbench, and specific Pension Plans are created as versioned and audited documents using the workbench. From the Pension Plan document, calculation engines for Pension Fund administration can be generated.

PhUSE 2013



9. The Capgemini Pension workbench (Intentional Software, 2013)

The domain experts express their knowledge using the domain languages. The workbench makes use of spreadsheets shows that when we give business professionals a structured and effective domain language for computation.



10. Sample of Domain Experts knowledge capture using Excel.

Altimeter Tests

Principles of operation and testing procedures for three pointer and counter drum pointer altimeters are described in the following paragraphs.

▸ Prerequisites

▾ Scale error test

The altimeter shall be subjected successively to pressures corresponding to the altitudes specified in Scale Error (Three Pointer) up to the maximum altitude.

Prerequisites:

- Barometric pressure at 31.02 in (Hg).
- The reduction in pressure shall be made at a rate not in excess of 20000 ft/min to within 150 ft of the test point.
- The test point shall be approached at a rate compatible with the test equipment.
- The altimeter shall be kept at the pressure corresponding to each test point for at least 1 min, but not more than 10 min, before a reading is taken.
- The error at all test points must not exceed the tolerances specified in Scale Error (Three Pointer).

Altitude test pressure (ft)	Equivalent atmospheric pressure (mm (Hg))	Equivalent atmospheric pressure (in (Hg))	Tolerance at 25°C (ft)	Test Value
0	760	29.921	30	23
500	746.37	29.385	30	514
1000	732.93	28.856	30	1066
1500	719.7	28.335	30	1523

▸ Hysteresis test

▸ After-effect test

▸ Friction test

11. Electronic Work Documents (Intentional Software, 2013)

The example of dynamic electronic work documents is certainly relevant. For example the inclusion of Protocol Representation and Study design domain information means that there is the opportunity to render this knowledge in a standard human readable form. Many of the components of a Protocol Design document are standard across Trials.

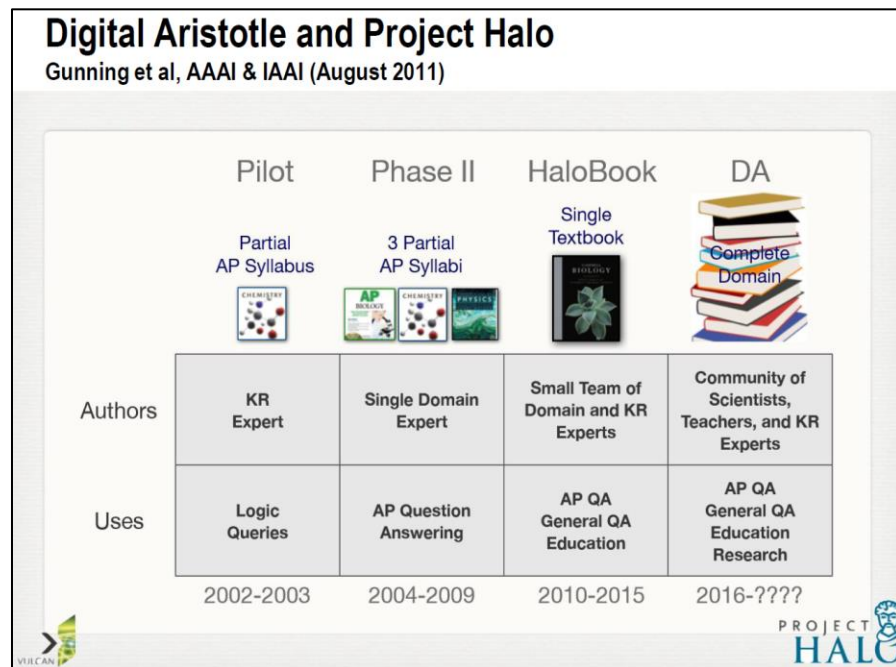
PhUSE 2013

RULELOG

The body of work presented by Benjamin Grosf (Grosf, 2013) and Michael Kifer (Kifer, 2013) in the RULELOG specification “Rulelog is the logic underlying knowledge representation languages such as FLORA-2 and SILK. It combines much of the latest work in logic programming, non-monotonic reasoning, business rules, and the Semantic Web. It is designed to be appropriately expressive for supporting knowledge representation in complex domains, such as sciences and law, and yet to be efficiently implementable.” (Kifer G. a., 2013) Their recent presentation at Universal Health Exchange Language Workshop provides an impressive and appropriate body of work for the clinical analysis and reporting Domain. (Grosf, Rulelog as Theoretical Foundation for Universal Health Exchange Language, 2013)

The team at Coherent Knowledge Systems (Coherent Knowledge Systems, 2013) and the depth of their approach that enables rapid semi-automatic acquisition of rich logical knowledge from text

“In this TL1 KA experiment, about 2,500 English encoding sentences were axiomatized. These included hundreds of questions. A number of questions, some of them sophisticated, answered successfully using Rulelog inferencing (in SILK) on the axioms. However, due to resource limitations of the study, only relatively limited tests of question-answering (QA) were conducted. The focus of the experiment was on KA productivity, primarily, and KA coverage, secondarily. Encoding sentence length averaged 10 words and ranged up to 25 words. One main defeasible axiom in Rulelog (SILK syntax) resulted from each sentence. On average, each such main axiom transformed into over 5 rules in normal (unextended) LP. It took less than 10 minutes (of KE labor) on average per sentence to: author, disambiguate, formalize, review, and revise a sentence.”



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

PhUSE 2013

Sentences (2322) Relations	all types	axiomatic	all editors	7/ 1/2012
Text	Axiom			
The environment of a cell is the solution surrounding it.	$\forall (x8) \text{cell}(x8) \Rightarrow \exists (x6) \text{environment}(\text{of}(x8))(x6) \Rightarrow \text{solution}(x6) \wedge \text{surround}(x6, x8)$			
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.	$\forall (x6) \text{endoplasmic}(\text{reticulum})(x6) \Rightarrow \exists (x19) (\text{eukaryotic}(\text{organism})(x19) \wedge \exists (x14) (\text{cell}(\text{in}(x19))(x14) \wedge \text{organism}(\text{in}(x14))(x6)))$			
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.	$\forall (x8) \text{molecule}(x8) \Rightarrow \forall (x5) \text{diffusion}(\text{of}(x5, x8)) \Rightarrow \forall (x18) \text{constant}(\text{vibration})(\text{of}(x8)) (x18) \Rightarrow \text{result}(\text{of}(x8), x5)$			
Cholesterol is a steroid.	$x5 = \text{cholesterol} \rightarrow x5 = \text{steroid}$			

12 Example of Text and Axiom logic.

A reported a costing of approximately \$3-4/word or \$500-1500/page. (Grosf P. H., 2013). Based upon my loose analysis of approximately 300,000 pages of legislation and guidance's. This amounts to a cost of Between \$150m-\$450m, using today's estimates and technology.

CONCLUSION

I hope that this rather brief overview has been of use. Currently, finalizing the development and testing of a SAS dataset consumer class in C++, based upon the work by Matt Shotwell (Shotwell, 2013), with the source R read.sas7bdat code (Shotwell, SAS7BDAT R Code, 2013) and documentation (Shotwell, SAS7BDAT Database Binary Format, 2013) being available online. Strategically, this makes sense as it would allow for the building of a solution that leverages existing SAS based solutions. Additionally, work continues on the issue of a viable solution for SAS script generation using Clinical Metadata/Rules is another core software capability. It was, actually the implementation of a SAS/Macro base derivation engine in a recent project – with all its brittleness, that lead to some of realisations that form the basis of this paper. With respect to the agile manifesto, the creation of some clinical rule and metadata solution that can deliver value to existing clinical reporting frameworks is a reasonable goal. The solutions offered by RULELOG standard (Grosf B. , Rulelog as Theoretical Foundation for Universal Health Exchange Language, 2013) has been of incredible value in thinking about the C.A.R.E. framework.

REFERENCES

- Archives, T. N. (kein Datum). *The Medicines for Human Use (Clinical Trials) Regulations 2004*. Von <http://www.legislation.gov.uk/ukxi/2004/1031/contents/made> abgerufen
- Bosch Sotare innovations GmbH. (01. 01 2013). *Visual Rules for Business Rule Management*. Abgerufen am 09. 08 2013 von Software Innovations : <http://www.bosch-si.com/technology/business-rules-management-brm/visual-rules.html>
- Coherent Knowledge Systems. (01. 01 2013). *Home*. Abgerufen am 09. 08 2013 von Coherent Knowledge Systems: <http://coherentknowledge.com/>
- Community, C. (09. 08 2013). *C.A.R.E.* Abgerufen am 08 2013 von Clinical Analysis and Reporting Environment: <https://plus.google.com/communities/103940563393464453082>
- Digital Financial Reporting. (08. 12 2008). *Hello World! XBRL Example*. Abgerufen am 09. 08 2013 von <http://xbrl.squarespace.com/>: <http://xbrl.squarespace.com/JOURNAL/2008/12/18/HELLO-WORLD-XBRL-EXAMPLE.HTML>
- Eclipse Org. (01. 01 2013). *Eclipse Home Page*. Abgerufen am 09. 08 2013 von Eclipse: <http://www.eclipse.org/>
- Grosf, B. (16. 06 2013). *Benjamin Grosf, PhD*. Abgerufen am 09. 08 2013 von Benjamin Grosf, PhD: <http://www.mit.edu/~bgrosf/>
- Grosf, B. (25. 06 2013). *Rulelog as Theoretical Foundation for Universal Health Exchange Language*. Abgerufen am 09. 08 2013 von MIT: <http://www.mit.edu/~bgrosf/paps/talk-health-exchange-lang-2013.pdf>
- Grosf, P. H. (20. 06 2013). *Rulelog: Highly Expressive yet Scalable Logic enabling Cost-Efficient Ontologies*. Abgerufen am 09. 08 2013 von Coherent Knowledge: <http://coherentknowledge.com/wp-content/uploads/2013/05/talk-rulelog-ontolog-forum-20130620.pdf>
- <http://www.cancer.gov/>. (09. 08 2013). *Home*. Abgerufen am 09. 08 2013 von National Cancer Institute: <http://www.cancer.gov/>
- ICH. (01. 01 2013). *ICH*. Abgerufen am 09. 08 2013 von ICH: <http://www.ich.org/>
- ICH. (01. 01 2013). *ICH Vision*. Abgerufen am 09. 08 2013 von ICH: <http://www.ich.org/about/vision.html>
- Intentional Software. (01. 01 2013). *Home*. Abgerufen am 09. 08 2013 von Intentional Software: <http://www.intentsoft.com/>
- James Manyika, M. C. (01. 05 2013). *Disruptive technologies: Advances that will transform life, business, and the global economy*. Abgerufen am 09. 08 2013 von McKinsey: http://www.mckinsey.com/insights/business_technology/disruptive_technologies
- James Manyika, M. C. (01. 05 2013). *Insights & Publications*. Abgerufen am 09. 08 2013 von McKinsey & Company:

PhUSE 2013

- http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Disruptive%20technologies/MGI_Disruptive_technologies_Full_report_May2013.ashx
- Kifer, G. a. (31. 05 2013). *Rulelog: Syntax and Semantics*. Abgerufen am 09. 08 2013 von Ruleml: <http://ruleml.org/rif/rulelog/spec/Rulelog.html>
- Kifer, M. (01. 01 2013). *Michael Kifer*. Abgerufen am 09. 08 2013 von Michael Kifer Home Page: <http://www.cs.stonybrook.edu/~kifer/>
- Manns, L. R. (2004). *Fearless Change Patterns for Introducing New Ideas*. Addison-Wesley Professional.
- Markus Voelter, C. D. (2013). *DSL Engineering - Designing, Implementing and Using Domain-Specific Languages*. dslbook.org.
- MIT Technology Review. (kein Datum). Von MIT Technology Review: <http://www2.technologyreview.com/player/07/01/programming/1.aspx> abgerufen
- National Institute for Health Research. (01. 01 2013). *Clinical Trials Toolkit*. Abgerufen am 09. 08 2013 von Clinical Trial Toolkit: <http://www.ct-toolkit.ac.uk/home>
- OSA. (01. 01 2013). <http://www.opensecurityarchitecture.org/cms/>. Abgerufen am 09. 08 2013 von <http://www.opensecurityarchitecture.org/cms/>: <http://www.opensecurityarchitecture.org/cms/>
- Protégé. (01. 01 2013). *Protégé Home*. Abgerufen am 09. 08 2013 von Protégé: <http://protege.stanford.edu/>
- Pustejovsky, J., & Stubbs, A. (2012). Natural Language Annotation for Machine Learning. In J. Pustejovsky, & A. Stubbs, *Natural Language Annotation for Machine Learning* (S. 342). O'Reilly Media, Inc.
- Rosenberg, S. (01. 01 2007). *Anything you can do I can do Meta*. Abgerufen am 09. 08 2013 von MIT Technological Review: <http://www.technologyreview.com/featuredstory/407091/anything-you-can-do-i-can-do-meta/>
- Rudolph, G. (01. 06 2008). *Some Guidelines For Deciding Whether To Use A Rules Engine*. Abgerufen am 09. 08 2013 von Jess information: <http://herzberg.ca.sandia.gov/guidelines.shtml>
- S. Modgil, P. H. (01. 01 2000). *Developing A Knowledge-based Tool for Authoring Clinical Trial Protocols (2000)*. Abgerufen am 09. 08 2013 von citeseer: <http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=FE7B0C8B3942E8E79A1E52E14D9F633D?doi=10.1.1.27.7205&type=ab>
- Samwald, M. (21. 11 2012). *Preview of OWL 2 ontology for clinical pharmacogenomics and decision support (now called 'Genomic CDS')*. Abgerufen am 09. 08 2013 von W3C Public Mailing List Archives: <http://lists.w3.org/Archives/Public/public-semweb-lifesci/2012Nov/0030.html>
- Shotwell, M. (05. 06 2013). *BioStatMatt*. Von BioStatMatt Blog: <http://biostatmatt.com/> abgerufen
- Shotwell, M. (09. 01 2013). *SAS7BDAT Database Binary Format*. Abgerufen am 09. 08 2013 von github: <https://github.com/BioStatMatt/sas7bdat/blob/master/inst/doc/sas7bdat.rst>
- Shotwell, M. (26. 04 2013). *SAS7BDAT R Code*. Abgerufen am 09. 08 2013 von GITHUB: <https://github.com/BioStatMatt/sas7bdat/blob/master/R/sas7bdat.R>
- standards.iso.org. (01. 01 2013). *Freely Available Standards*. Abgerufen am 09. 08 2013 von ISO Standards: <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- Tao Lei, F. L. (01. 01 2013). *From Natural Language Specifications to Program Input Parsers*. Abgerufen am 09. 08 2013 von <http://people.csail.mit.edu/>: <http://people.csail.mit.edu/taolei/papers/acl2013.pdf>
- Tao Lei, F. L. (29. 07 2013). *From Natural Language Specifications to Program Input Parsers*. Abgerufen am 09. 08 2013 von MIT Computer Science and Artificial Intelligence Laboratory: <http://groups.csail.mit.edu/rbg/code/nl2p/>
- U.S. Securities and Exchange Commission. (20. 05 2013). *Securities and Exchange Commission XBRL Portal Home Page*. Abgerufen am 09. 08 2013 von Securities and Exchange Commission: <http://xbrl.sec.gov/>
- University of Amsterdam. (01. 12 2008). *European project for Standardised Transparent Representations in order to Extend Legal Accessibility*. Abgerufen am 09. 08 2013 von Welcome to the Estrella project website: <http://www.estrellaproject.org/doc/D1.1-LKIF-Specification.pdf>
- Verbeek, C. G. (09. 11 2012). *XES Standards*. Abgerufen am 09. 08 2013 von XES Standards Organisation: <http://www.xes-standard.org/start>
- WHO. (09. 08 2013). *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. Abgerufen am 03. 08 2013 von International Statistical Classification of Diseases and Related Health Problems 10th Revision: [<http://apps.who.int/classifications/icd10/browse/2010/en>
- WHO. (kein Datum). *ICF Web Page*. Abgerufen am 09. 08 2013 von ICF Web Page: <http://apps.who.int/classifications/icfbrowser/>
- wikipedia. (03. 08 2013). *Web Crawler*. Abgerufen am 09. 08 2013 von wikipedia: http://en.wikipedia.org/wiki/Web_crawler
- Wikipedia. (13. 07 2013). *XBRL*. Abgerufen am 09. 08 2013 von wikipedia: <http://en.wikipedia.org/wiki/XBRL>

RECOMMENDED READING

1. S Modgil, P Hammond, JC Wyatt, H Potts *Developing A Knowledge-based Tool for Authoring Clinical Trial Protocols* (S. Modgil, 2000)
2. Freely available standards (standards.iso.org, 2013)
3. D Jurafsky, J H Martin *Speech and Language Processing 2nd edition*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Greg Fuller
GIANCE SARL
56 Rue du Chateau zu Rhein
68200 Mulhouse
France
Work Phone: 0033369194871
Mobile Phone: 0033667248661
Email: fuller.greg@gmail.com
Web: <https://plus.google.com/communities/103940563393464453082>

Brand and product names are trademarks of their respective companies.

