

Cerebra®

UK CJIT Semantic Technologies Seminar

Joining Together IT for Automated Real-time Processing of Criminal Data; An Introduction to Data Semantics

Speaker Profile

- VP Technology, Cerebra
- Author: Adaptive Information
  - John Wiley and Sons: Business and Systems Management Series Interscience
- Consultant: US National & State Agencies
  - US Department of Homeland Security
  - Customs and Border Protection
  - California Anti-Terrorism Information Center
  - Various Fortune 500 and US IC Agencies
- Software Engineering & Architecture:
  - Chief Technology Officer, Modulant. Principal Engineer, Modem Media. Senior Consultant, Ernst & Young CTE, Software Engineer, Amgen

Cerebra®

Prelude

- **Presentation Goals**
  - Straight Talk; No Hype!
  - Education – what, why, how…
  - Explain Value-drivers & ROI
- **Presentation Takeaways**
  - CJIT challenges are information-centric
  - It's your information (not the s/w vendor's)
  - Use open and global standards
  - There are no silver bullets, BUT the Enterprise Semantic Web is better than anything before it – it will allow you to evolve your data long after the software has been replaced…

Cerebra®

CJIT: Same, but Different

- CJIT shares many of the same challenges as private enterprise, but CJIT faces these challenges in different ways & different scales
- Some of the special characteristics include:
  - Highly decentralized technology management
  - Sheer size, scope and scale of infrastructure
  - Vastly different community cultures
  - Progress comes from negotiation, not mandate
  - Security needs, privacy rights, watchdog groups
  - Public safety is at stake – not corporate profits
- These constraints are real – not hypothetical or easily overcome by executive mandates, more staff, or funding increases

Cerebra®

Firm Requirements

- CJIT's special challenges for information sharing and systems architecture requires key technical capabilities to succeed:
  - Massively federated information
    - No central data warehouse
  - Loosely-coupled IT system interfaces
  - Automated terminology/data reconciliation
  - Evolve-able data structures – not static
  - Secure data at multiple levels, guaranteed
  - Accuracy and correctness of algorithms
- Individually, these requirements have been accomplished – but rarely in combination
  - ..and one more thing: the overall approach should leverage open and global standards

Cerebra®

Horizontal Data Fusion

- To make information fluid and dynamic…
  - Wanted Persons, Alerts, Court Records
  - Firearms Databases, Prison Records
  - DVLA, Plates, Licenses, Address, Vehicle
  - Travelers – Airline, Trains, Bus
  - Transport – Cargo and Shipments
  - Vehicles – Aircraft, Boat, Train, Truck, Auto
  - Criminal histories, Agency Records, Prison Records
  - Biometrics Data and Private Records (credit etc.)
- Data from anywhere must be supported
  - Unstructured data – HTML, PDF, Word, Excel, Email…
  - Semi-structured data – EDI, delimited, tokenized, indexed, positional, ministry exchange formats
  - Structured – Relational, XML, Objects, Legacy/Mainframe

Cerebra®

What Hasn't Worked

- Hyped-up vendor/market solutions provide some, but not all of the pieces of the puzzle
- Consider the CJIT information sharing conundrum and the technologies that WILL NOT provide significant relief by themselves:
  - Data Warehouses
  - Service-Oriented Architectures
  - COTS Middleware / Process Management
  - Portals or Composite Applications
  - Programming Frameworks – J2EE/.NET etc.
  - Industry Vocabulary Standards
- Itemizing all the reasons none of these technology approaches will solve CJIT issues would be exhaustive,
  - Suffice to say here that none of them provide for enough agility at the information (data+metadata) layer – by UNLOCKING DATA FROM SOFTWARE

Cerebra®



Core Tenets of Change

- Information is power
  - As always, for nations and businesses
- A change is required
  - Collection rate outstrips understanding
- The change must be fundamental
  - OO, Java, RDB, XML, EAI (etc) aren't enough
- YOU own your information
  - Commit to a strategy that keeps it that way (not vendor or SI ownership)

Cerebra®



Remaining Agenda

- Semantics Backgrounder
  - What are semantics, inference?
  - Why is any of this important?
  - Should you be skeptical?
- Enterprise Semantics Vision
  - Logical "glue" for the enterprise
  - Complexity management
  - Automated decision making
- Technology Overview
  - Use Cases, Services, Architectures
- Conclusions
  - No silver bullets!
  - Train, Train, Train!

Cerebra®



Cerebra®

Semantics Backgrounder

Data is the Network



Information Overload

Q: In General, why are semantics important?

- Maintenance and Support of "Data Meaning" is the largest, most expensive, and time-consuming portion of IT budgets
  - Mars Climate Observer – Feet vs. Meters
  - Y2K – "01" equals 1901 or 2001; What will happen on 1/1/00?
  - Product Hierarchies – Is my PDA a phone or a camera or both?
- Over $200B is spent each year on integration issues involving the meaning of data
  - Classic "n-squared" problem is more severe at the data layer
  - Data-to-data interfaces outnumber "pipes"
  - Tightly-coupled = Brittle
- Information growth is accelerating – FAST!
  - 2002-2005 – more new data than all of history
  - 5 exabytes of new digital data created in 2002 – enough for .5 million new Libraries of Congress

Cerebra®

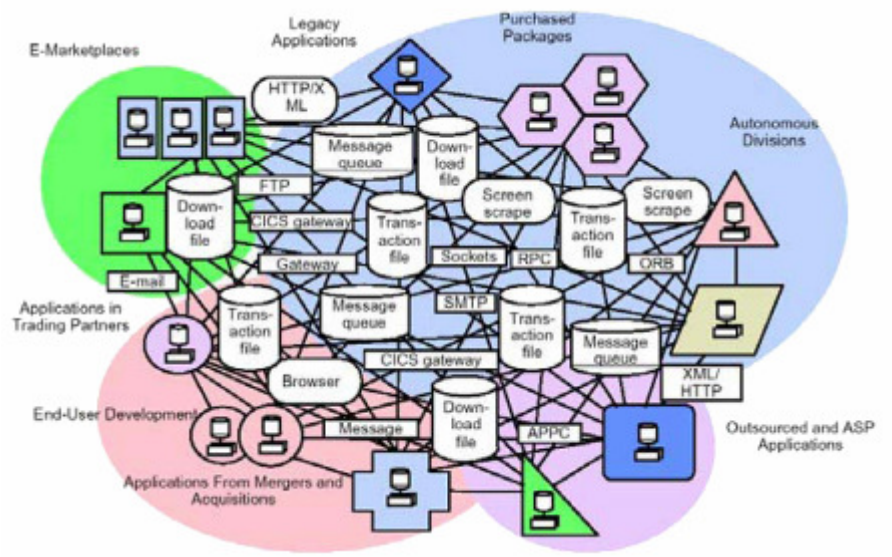
Infrastructure Complexity

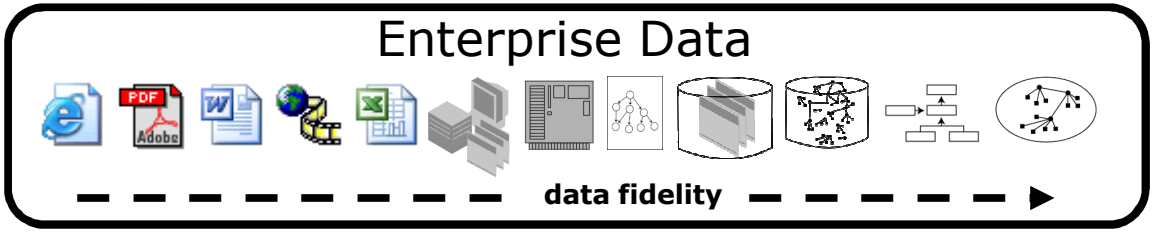
Q: For business, why are semantics important?

EAI? **<XML>?** **Warehouse?**



BAM? **EII?** **ETL?**





Cerebra®



Graphs and Ontology

Q: How can the Semantic Web help?

- "Small Pieces Loosely Joined"
  - Small Pieces: RDF Triples – data graphs
  - Loosely Joined: OWL Models – concept collections **defined with precision**
- Fundamentally Different Than…
  - XML – taxonomic tree structures
  - Java – object structures (frames)
  - RDB – two-dimensional/tabular structures
- From "Trees to Piles of Leafs"
  - RDF Triples, collections of atomic data
  - Ontologies are "views" into the whole
  - "Strict" trees (XML) are not required
    - Ontologies are multi-dimensional, with formal, logically-sound, data semantics

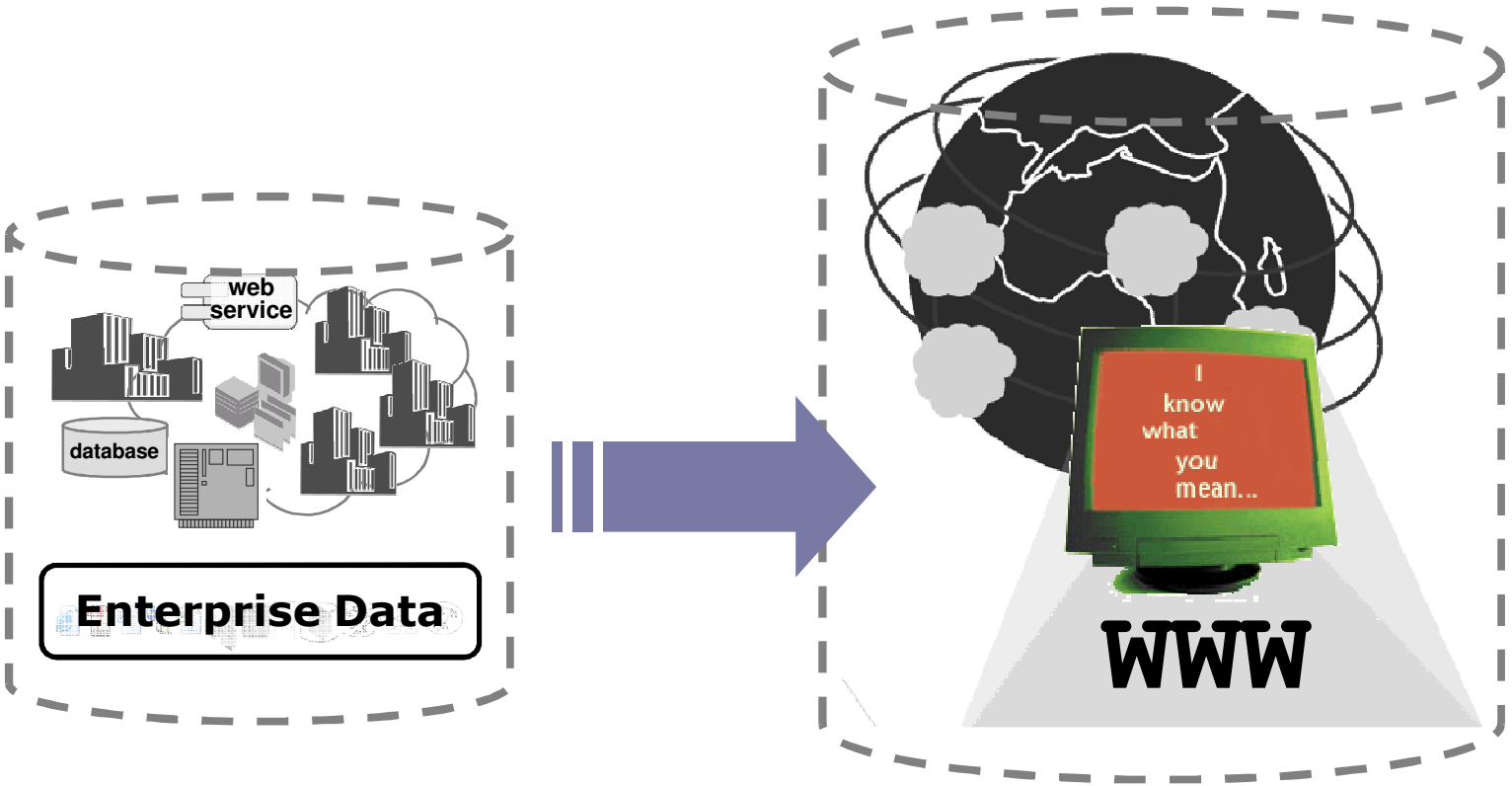
Some Definitions

Semantics
Taxonomy
Schema
Ontology
Inference

Cerebra®

It's Business, not Social

Q: What is the Enterprise Semantic Web?

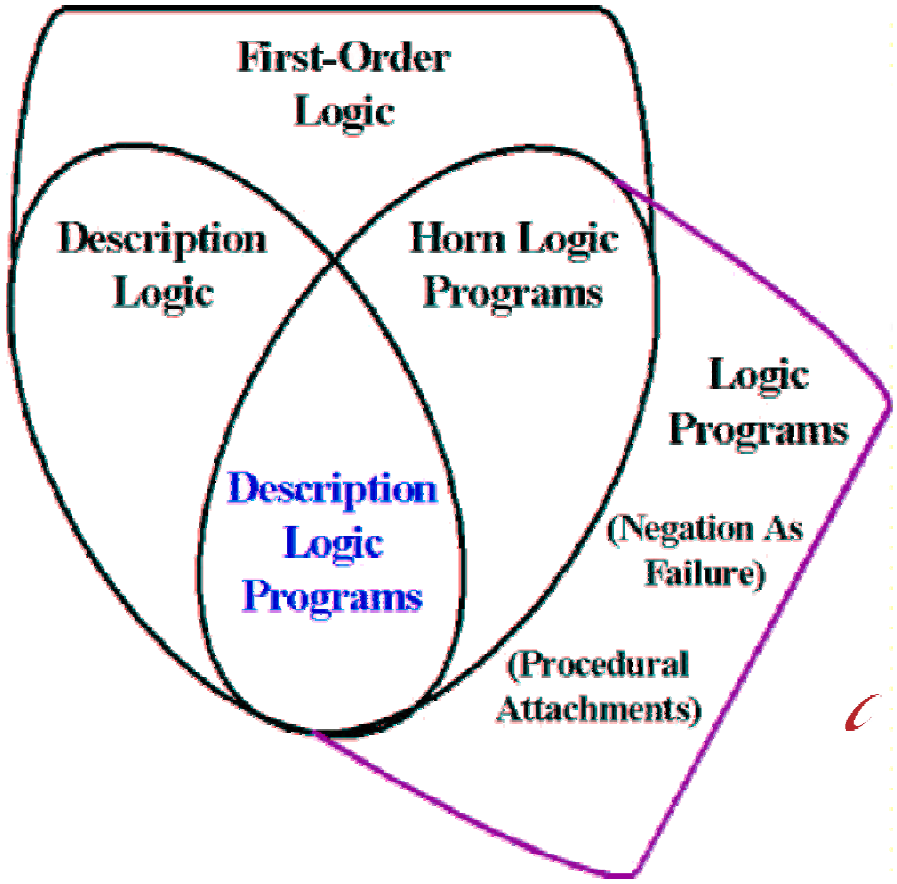


TODAY ***SOMEDAY***

Cerebra®

Math Matters

Q: Why are formal rules and logics crucial?



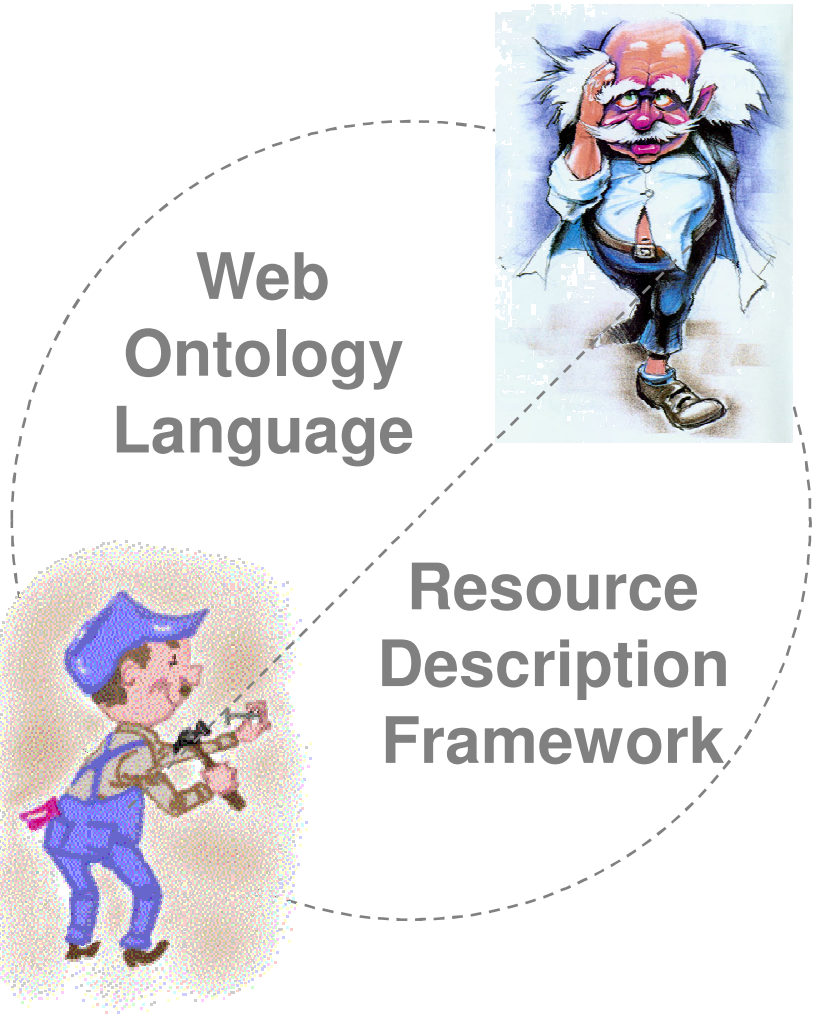
- What if your relational database gave unreliable query responses?
  - Can pigs fly?
  - How many units did you sell last quarter?
  - Is this annual audit report reliable?
- If the data or rule is there, the query ought to know of it…

Cerebra®

Two Sides of a Coin

Q: Why do scruffy/neat people matter?

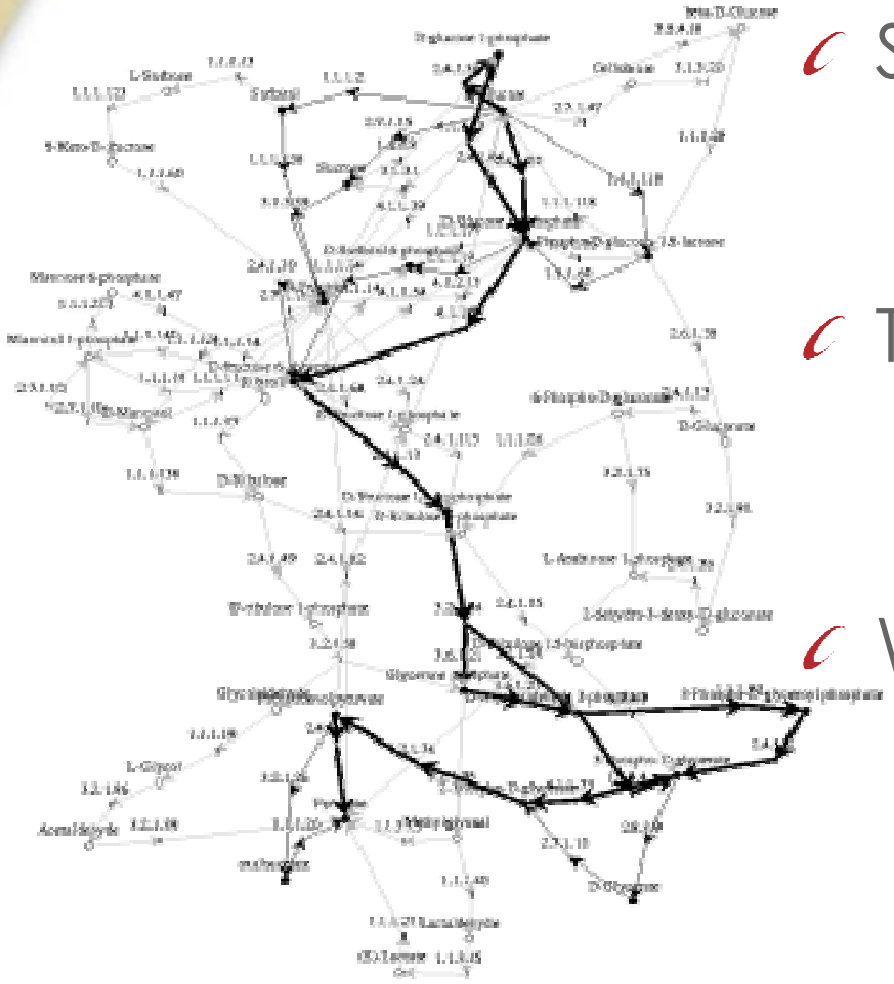
- Artificial Intelligence (AI) factions and levels
  - "**Neats**" – theoreticians, symbologists, proving truth
  - "**Scruffies**" – engineers, connectionists, getting a job done
- Theory and practice meeting in the middle…
  - Modern RDBMS
  - Nuclear Power
  - Stealth Technology
  - Steam Engine
  - …more great things



Cerebra®

"Web-Effect" + Standards

Q: Isn't "Semantic Web" just AI all over again?



- Ssshhhhhh!
  - OWL/RDF are a type of AI "Knowledge Representation"
- Traditional AI
  - "Silos" of knowledge
  - Incompatible algorithms
  - Inconsistent "expertise"
- What's Different?
  - Everything has a URI
    - Models, classes, data
  - Predictable semantics
    - Std. Triples & Axioms

Cerebra®

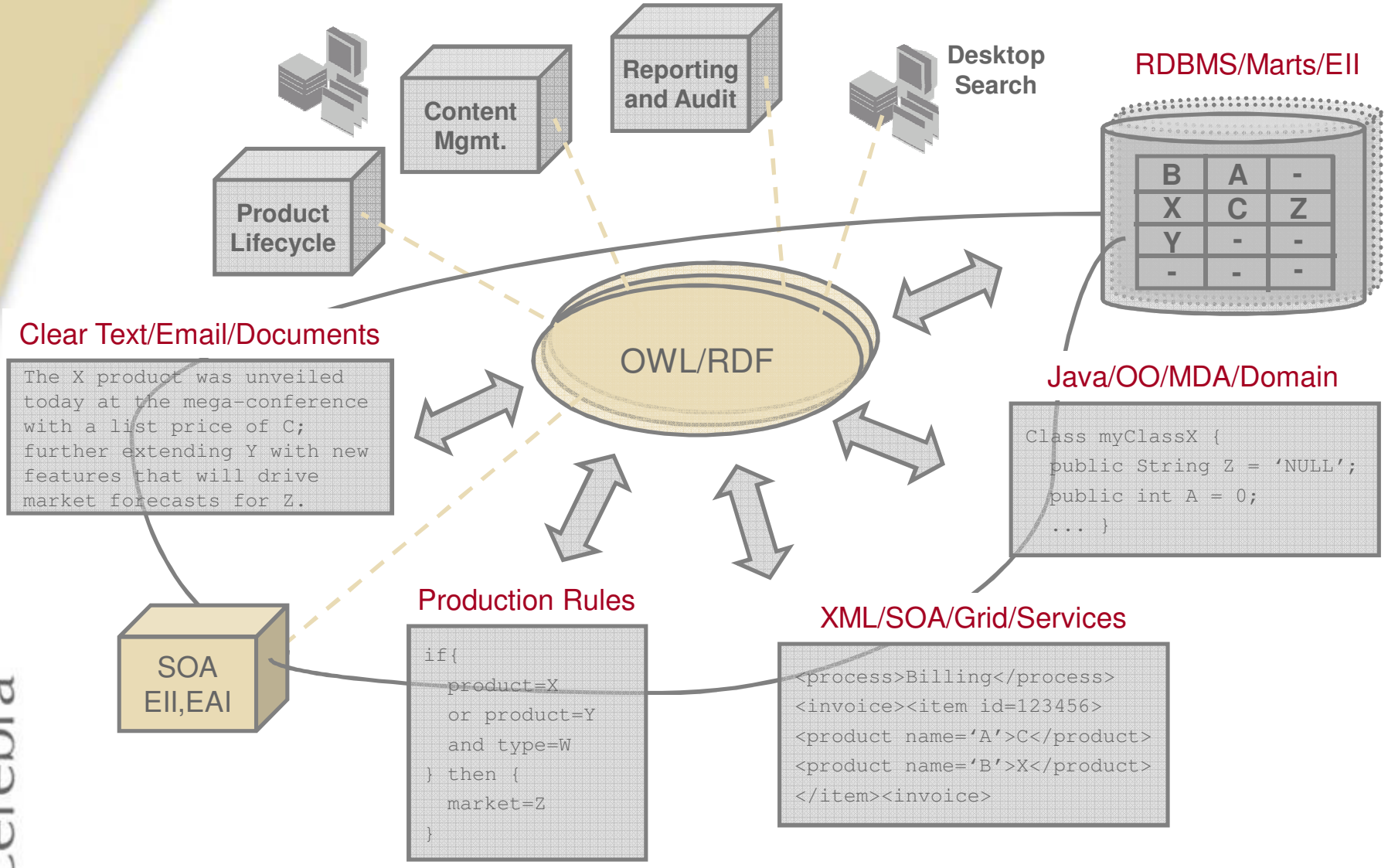


Cerebra®

Enterprise Semantic Web

Deployable Solutions Now

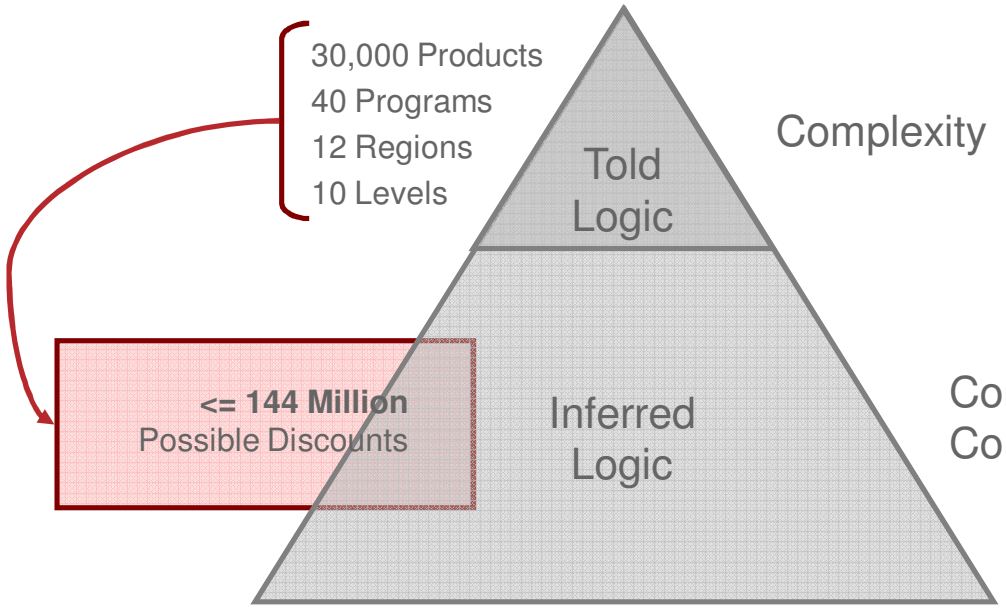
Logical Glue for Information

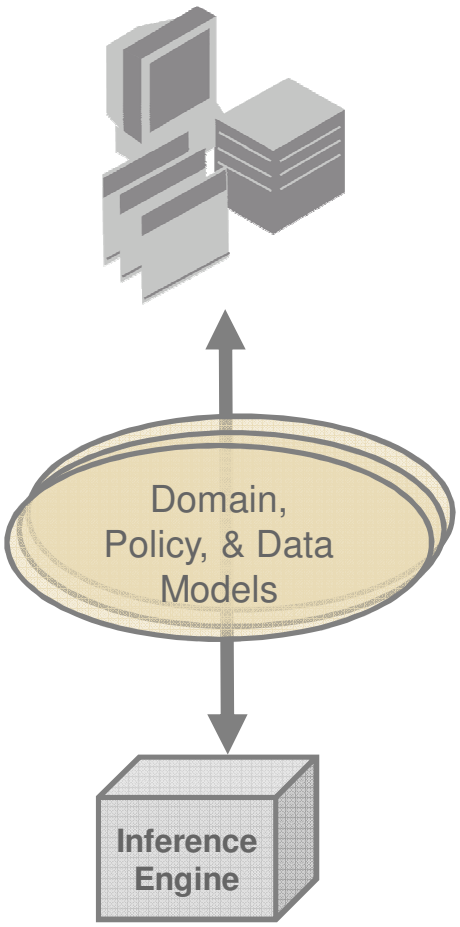


Cerebra®

Complexity Management

- Logical Inference
  - Infer the classification of data from models rather than encode it in brittle algorithms

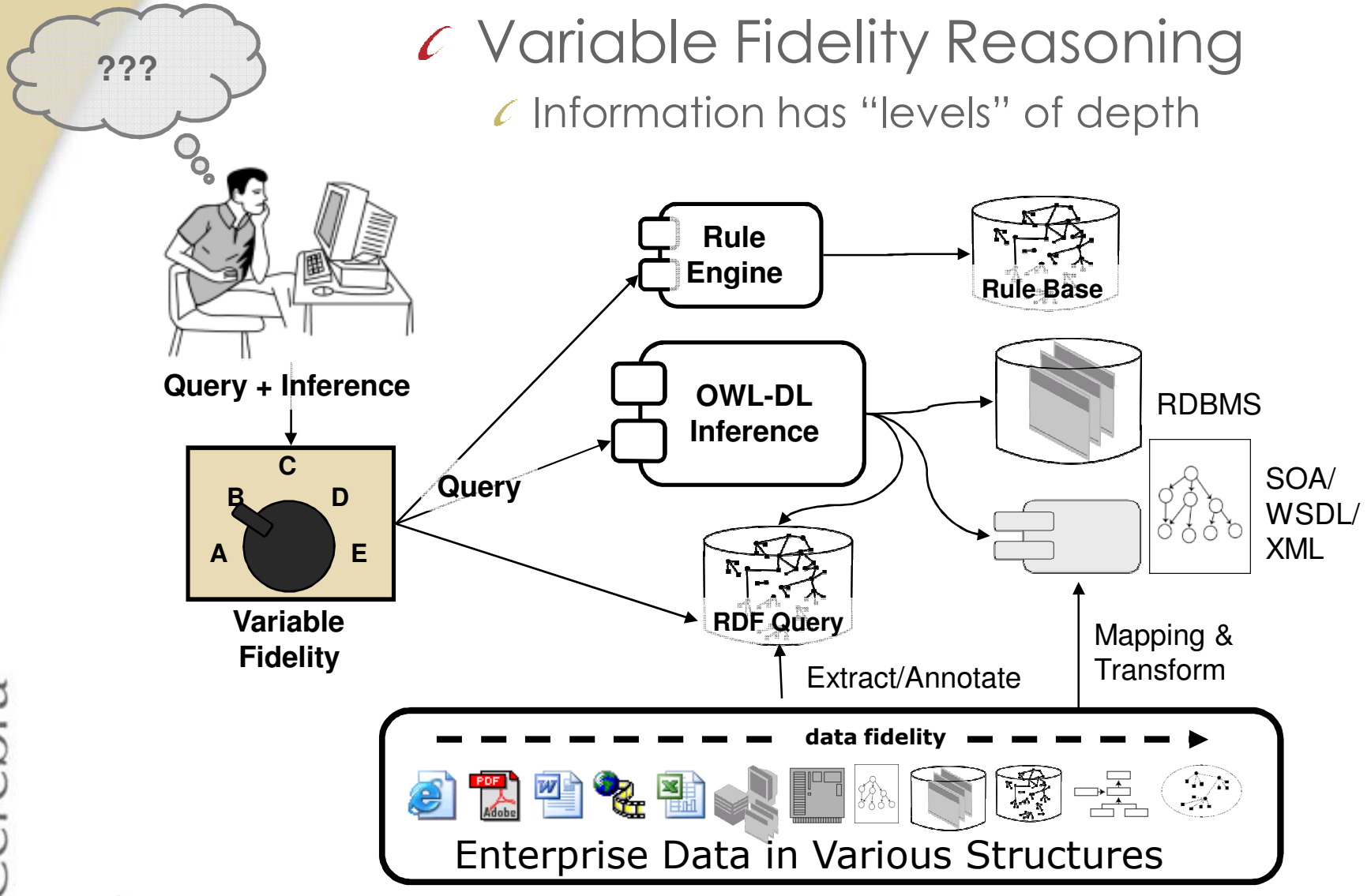




Cerebra®

Automated, Tuneable Decisions

- Variable Fidelity Reasoning
  - Information has "levels" of depth



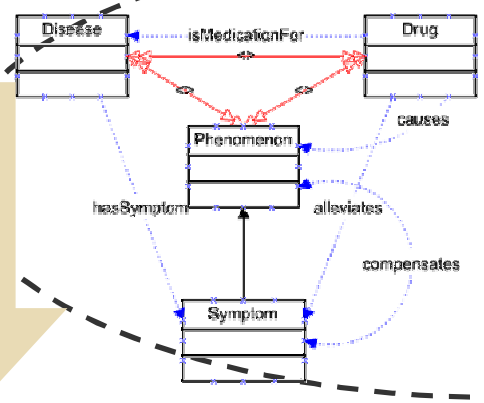
Cerebra®

Implement it Your Way

Top-down or Bottom-up



Model Ontologies

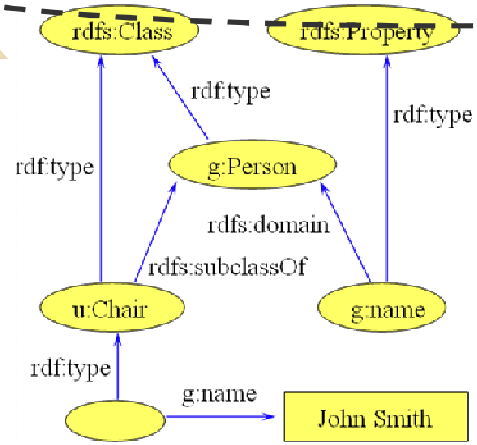



```
<?xml version="1.0" encoding="UTF-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#" x
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:ni
  xml:base="http://www.w3.org/2002/07/owl#" xmlns=
+ <owl:Ontology rdf:about="">
- <owl:Class rdf:ID="Disease">
  - <rdfs:subClassOf>
    - <owl:Restriction>
        <owl:onProperty rdf:resource="#hasSymptom" />
        <owl:allValuesFrom rdf:resource="#Symptom" />
      </owl:Restriction>
    </rdfs:subClassOf>
    <owl:disjointWith rdf:resource="#Phenomenon" />
  </owl:Class>
```

URI & Namespaces



Create RDF Triples




```
<rdfs:Property rdf:ID="name">
  <rdfs:domain rdf:resource="Person">
</rdfs:Property>

<rdfs:Class rdf:ID="Chair">
   <rdfs:subclassOf  rdf:resource=
      "http://schema.org/gen#Person">
</rdfs:Class>

<rdf:RDF
     xmlns:g="http://schema.org/gen"
     xmlns:u="http://schema.org/univ">
  <u:Chair rdf:ID="john">
    <g:name>John Smith</g:name>
  </u:Chair>
</rdf:RDF>
```

Cerebra®



Cerebra®

New Capabilities

Concept-based Query
Smart Query
Information Mediation
Taxonomy/Rule Alignment
Information View Management
Policy Audit and Security Compliance
Process Orchestration



Cerebra®

Service Utility Archetypes

Metadata Repository
Mediation Engine
Registry Service
Process Engine



Cerebra®

Component Architecture Patterns

RDF Database
RDF Facade
OWL Knowledge Base
OWL Facade
OWL Instance Store
Hybrid Fidelity Reasoning



Cerebra®

Govmt. Use Cases

QinetiQ – UK MoD, Tactical Ground Picture
LMCO – US Customs and Border Protection
BearingPoint – US Dept. Homeland Sec.
SAIC – US Intelligence Community
…



Cerebra®

Commercial Use Cases

Fortune 100 Electronics Manufacturer
Fortune 100 Software Company
Financial Compliance Services Provider
Clinical Healthcare Services Provider



Cerebra®

Conclusions

To Use or Not to Use?
Tradeoffs and Caution
Stay Tuned for a Wild Ride…

Choose Your Medicine

Pain Killers?

Oftentimes, there is a short-term need that predicates a fast, "stop the bleeding," type approach. Usually custom, one-off solutions are chosen because they are known quantities – at the cost of significant long-term issues like TCO and inflexibility.

Penicillin?

Certain industries are experiencing decades long systemic pain that software "pain-killers" can no longer help with. These industries are in need of solutions that solve the root of the problem – not the surface "bleeding."

- Government Agencies
  - Military – Network-Centric Enterprise Services
  - Security – Criminal Systems & Homeland Security
- Manufacturing / Product
  - Automotive Supply Chain - Electronics & CPG Supply Chain
- Life Sciences
  - Process Automation / Risk Analysis / Drug Discovery

Cerebra®

ROI Value Drivers

- New Information Capabilities
  - **What**: Doing something that you haven't done before to provide the business new value.
    Such As:
    - Linking annual risk strategies to live data inspection
    - Assembling SOA service integrations faster
    - Testing realtime "what-if" financial scenarios
    - Securing XML SOA payload with no code
- Total Cost of Ownership (reduction)
  - **What**: Doing what you are already doing, but doing it more efficiently at less operational cost.
  - **Such As**:
    - More productive integration maintenance staff
    - More productive naming/metadata staff

Cerebra®

Some Figures

- Caveat Emptor…
  - *"There's lies, damn lies, and ROI predictions"*
- Proven ROI Potential*
  - 20-90% reduction in labor hours, cycle time, inventory levels, operating cost, development time, and development cost
  - 2-50x improvements in quality, service levels and productivity
  - 20-80% reduction in total cost of ownership
  - 3-12 month positive return on investment
  - 30-300x positive ROI over 3 years
  - Mitigated risk, reduced vulnerability to fraud, liability or litigation

**TopQuadrant market research leveraging Gartner, IDC, Forrester and Giga*

Cerebra®



Uptake Momentum

- Commercial
  - Audi, General Motors, Cisco Systems, Fed-Ex, France Telecom, Nokia, Fujitsu, Hitachi, NEC, Unisys, Pfizer, General Electric…
- Federal
  - Federal CIO, SICoP, Navy, Army, DLA, EPA, NCI, DHS, CBP, NGA, MoD, NATO + other foreign federal
- Vendors/Systems Integrators
  - Oracle, IBM, Microsoft, Adobe, Hewlett Packard, WebMethods, CGEY, BearingPoint, SAIC, Lockheed Martin, SRI, MITRE, Booz Allen Hamilton
- Standards Bodies
  - W3C, OMG (MDA), OASIS (DCML, ebXML), ISO



No Free Lunch

THE BAD NEWS:

TRAIN YOUR PEOPLE!

THE GOOD NEWS:

DBA's may already know most of this stuff…
Object-Oriented analysts have a BIG head-start…

Tools are available:

Modeling – GUI driven, Visio-based, Rational-based, etc.
Inference – Java based, C++, App Server based, etc.
Rules – GUI driven, J2EE integration, .NET components, etc.
Database – RDF databases, open source, triples stores

Trainers are available:

Software Vendors (Network Inference, @Semantics, etc.)
TopQudrant – Dedicated RDF and OWL training

Cerebra®



Postscript

- **Presentation Goals**
  - Straight Talk; No Hype!
  - Education – what, why, how…
  - Explain Value-drivers & ROI
- **Presentation Takeaways**
  - CJIT challenges are information-centric
  - It's your information (not the s/w vendor's)
  - Use open and global standards
  - There are no silver bullets, BUT the Enterprise Semantic Web is better than anything before it – it will allow you to evolve your data long after the software has been replaced…

Cerebra®



Cerebra®

Questions?

Jeff Pollock
Vice President, Technology
jeff.pollock@cerebra.com



Cerebra®

“Pocket” Slides:

Definitions and Examples

Inference?
Logic With Less Code

BACK

Deduce **implicit facts** on explicit data at runtime. Get more application knowledge with **less coding.** Keep your business knowledge in **models, not algorithms.** Maybe learn a few new things too…

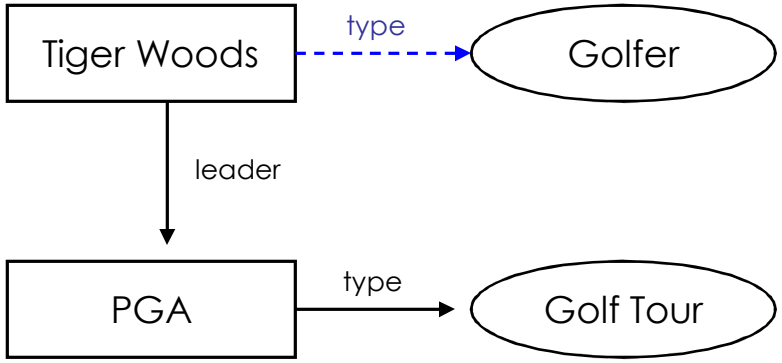
```
<owl:Property rdf:ID="leader">
  <rdfs:subPropertyOf
        rdf:resource="member" />
</owl:Property>

<owl:Class rdf:ID="Golfer">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty
            rdf:resource="member" />
      <owl:someValuesFrom
            rdf:resource="Golf Tour" />
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>
```

The **leader** of a **tour** is also a **member** of it

A **member** of a **golf tour** is a **golfer**

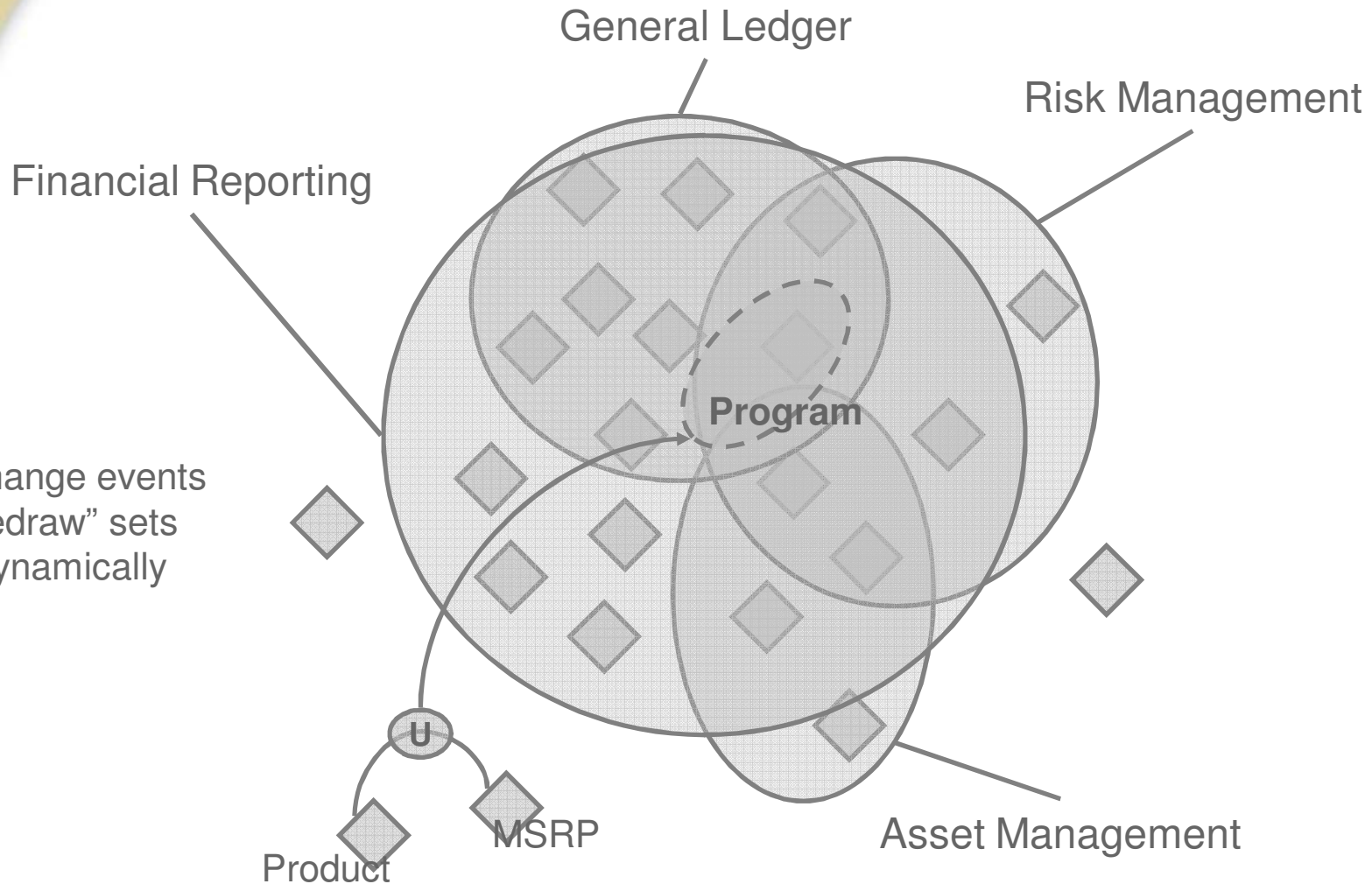
Therefore, the **leader** of a **golf tour** is a **golfer**



Cerebra®

Inference (2)

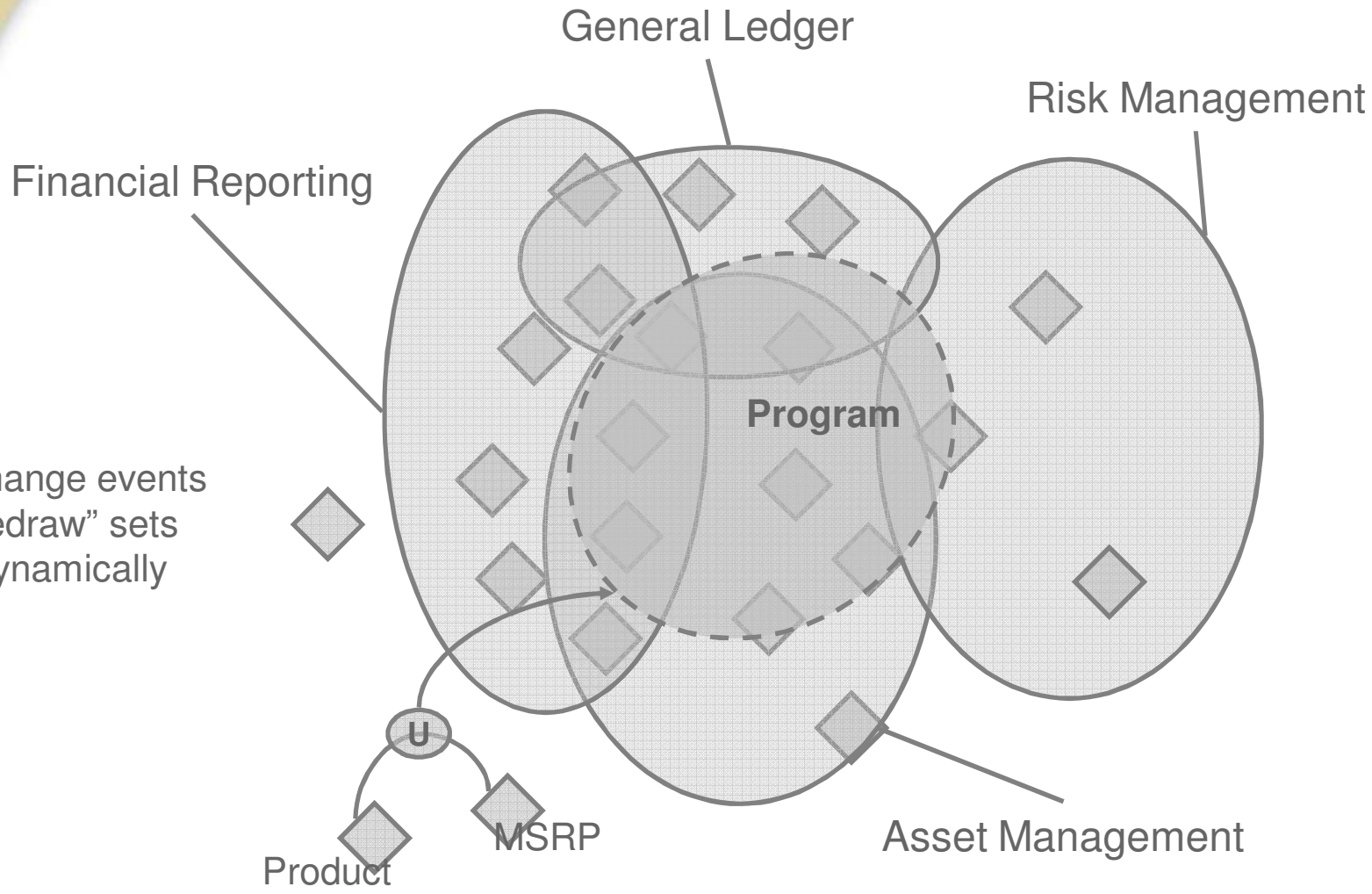
BACK



Cerebra®

Inference (3)

BACK



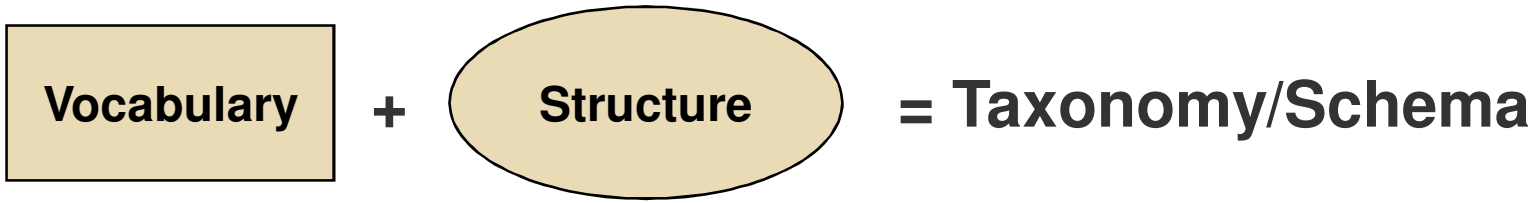
Cerebra®

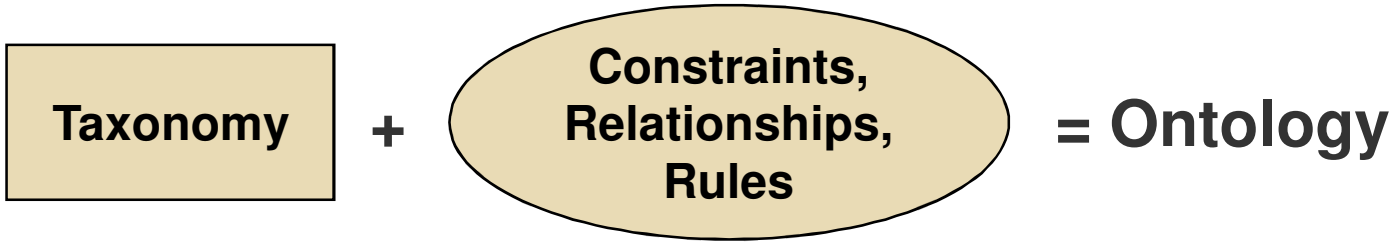


Ontology?

BACK

- Many definitions:
  - Vague – "a specification of a conceptualization"
  - Narrow – "the explicit meaning of terms in a vocabulary and relations between those terms"





Cerebra®

A Little Ironic…

Q: What are semantics?

BACK



"I don't know what you mean by 'glory,'" Alice said.

Humpty Dumpty smiled contemptuously. "Of course you don't—till I tell you. I meant 'there's a nice knock-down argument for you!'"

"But 'glory' doesn't mean 'a nice knock-down argument,'" objected Alice.

"When I use a word," Humpty Dumpty said, in a rather scornful tone, "**it means just what I chose it to mean—neither more nor less**."

Lewis Carroll, Through The Looking-Glass

Cerebra®



Semantics? (OWL Formal)

BACK

Constructor	DL Syntax	Example
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human ⊓ Male
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor ⊔ Lawyer
complementOf	$\neg C$	¬Male
oneOf	$\{x_1 \ldots x_n\}$	{john, mary}
allValuesFrom	$\forall P.C$	∀hasChild.Doctor
someValuesFrom	$\exists P.C$	∃hasChild.Lawyer
maxCardinality	$\leq nP$	≤1hasChild
minCardinali		

CLASS CONSTRUCTORS

AXIOMS

Axiom	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human ⊑ Animal ⊓ Biped
equivalentClass	$C_1 \equiv C_2$	Man ≡ Human ⊓ Male
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male ⊑ ¬Female
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{President_Bush} ≡ {G_W_Bush}
differentFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	{john} ⊑ ¬{peter}
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter ⊑ hasChild
equivalentProperty	$P_1 \equiv P_2$	cost ≡ price
inverseOf	$P_1 \equiv P_2^-$	hasChild ≡ hasParent$^-$
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+$ ⊑ ancestor
functionalProperty	$\top \sqsubseteq \leqslant 1P$	$\top \sqsubseteq \leqslant 1$hasMother
inverseFunctionalProperty	$\top \sqsubseteq \leqslant 1P^-$	$\top \sqsubseteq \leqslant 1$hasSSN$^-$

Cerebra®