The Power of Deep Reasoning with Large Graph Data

Benjamin Grosof and Janine Bloomfield

Coherent Knowledge*

Presentation (60-min.) at

Seattle Smart Data and Semantic Technology Meetup**

October 28, 2015

Hosted at Microsoft, Redmond, Washington, USA

Version of October 28, 2015

* Web: http://coherentknowledge.com

Email: *firstname.lastname*@coherentknowledge.com

** Web: http://www.meetup.com/Semantically-Webbed-Seattle-Meetup-Group/

Smart Rules for Smart Data

Coherent Knowledge: Company Overview

- Leverages over a decade of major government and privately funded research advances in artificial intelligence (AI) and semantic technologies. Founded 7/2013.
- Company offers: platform software product Ergo Suite™ + custom development / services
  - Capabilities: smart rules with deep reasoning
  - Apps: policy/regulatory compliance for finance, defense intelligence analysis, e-commerce
- World-class founder team: created many industry-leading logic systems & standards
  - XSB Prolog, RuleML, W3C RIF, W3C OWL-RL, IBM Common Rules, SWRL, SweetRules
  - Extensive experience applying logic systems to numerous domains in govt. and business

Michael Kifer, PhD
Principal Engineer
Creator, Flora.
Co-Architect, W3C RIF.
Prof., Stonybrook Univ.

Benjamin Grosof, PhD
CTO & CEO
Creator, IBM Common Rules.
Co-Architect, RuleML.
Prof., MIT.
Advanced AI Prog. Mger. for Paul Allen.

Theresa Swift, PhD
Principal Engineer
Co-Architect, XSB Prolog.
Consultant, US CBP.

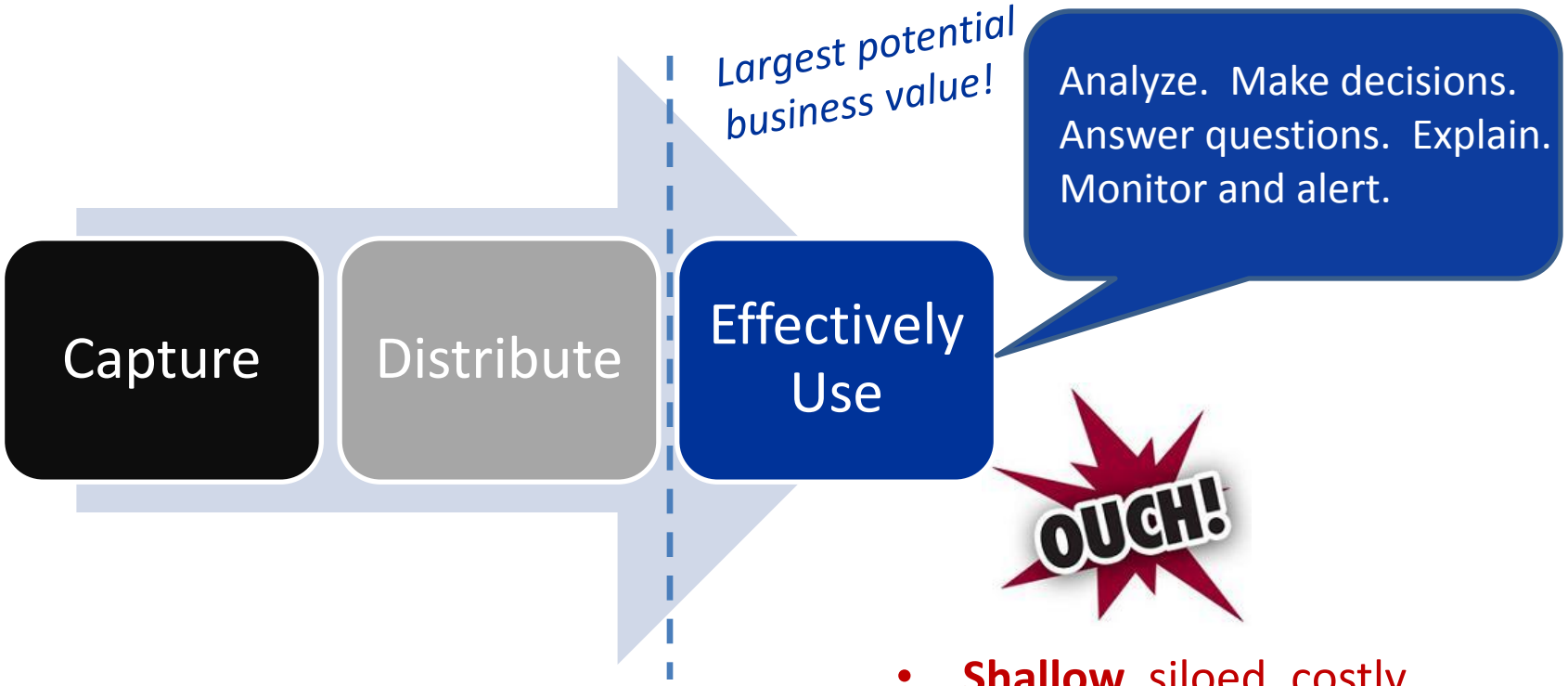
Paul Fodor, PhD
Senior Engineer
IBM Watson team.
Prof., Stonybrook Univ.

Janine Bloomfield, PhD
Dir., Marketing & Operations
Sr. Scientist, Climate Change, Environmental Defense Fund.
Mindexplorekids.org.

Outline

- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Problem: Knowledge Management

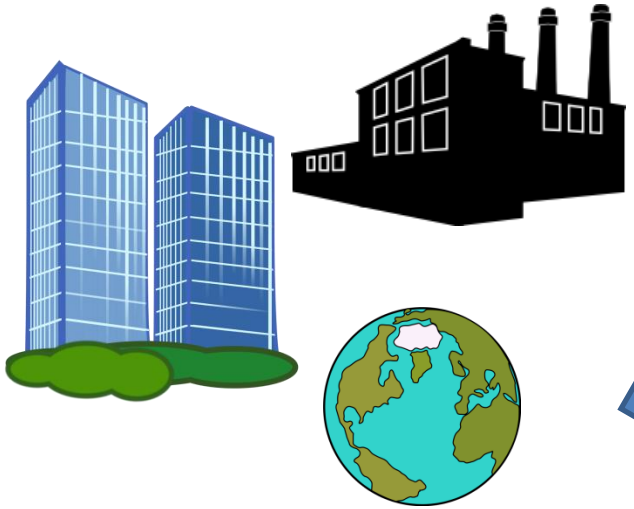


Existing Technologies

- Content Management
- Search
- Business Rules
- Databases

- **Shallow**, siloed, costly
- Not actionable
- Lacks automation, accuracy
- Lacks transparency
- End users and subject matter experts are not empowered

Evolution of Enterprise Knowledge Management

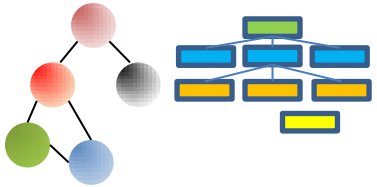


Data integration with flexible schemas, more meta-data. E.g., RDF, SPARQL, OWL.

Augment relational/traditional data stores.

Smart Data

Graph/Linked Data bases



Smart Rules

Decisions & Analysis

Deep reasoning, more complex knowledge. E.g., Textual Rulelog.

Leverage Smart Data investment.

Smart Rules with Deep Reasoning

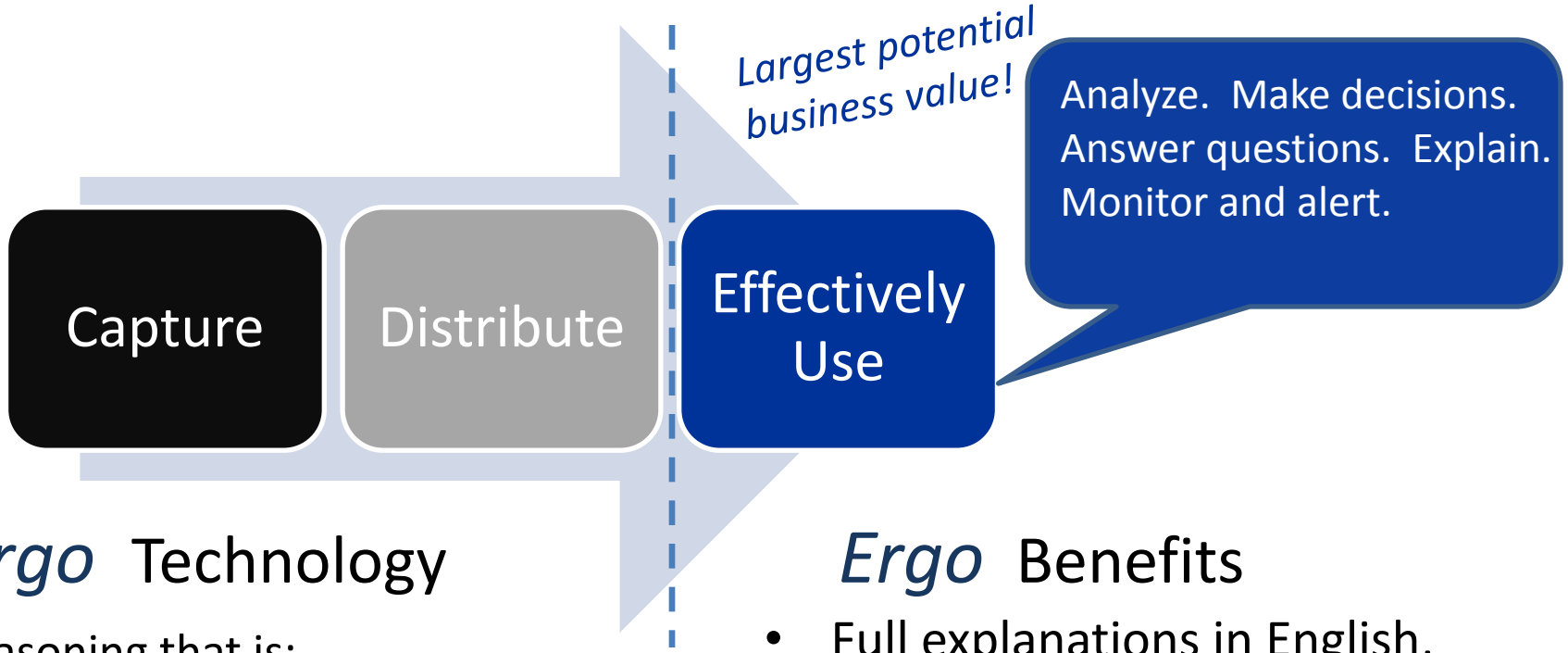
- Provide critical technical capabilities to unlock business value
  - More complex analytics
  - Context and mappings for data and system integration
  - Represent more complex knowledge: policies, regulations, science, …
  - Capture subject matter experts' (SMEs') insights directly, semantically

- Case studies in this presentation
  - Financial services regulatory/policy compliance
  - National intelligence analysis
  - E-commerce marketing

  - *Many other applications: health treatment guidance, info access, tutoring, …*

Smart Rules for Smart Data™

Solution: Deep Reasoning in Ergo™



Ergo Technology

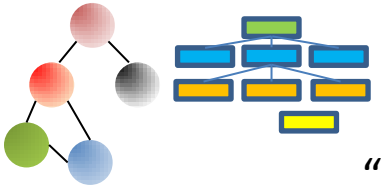
- Reasoning that is:
  deep, accurate, transparent
- Knowledge that is:
  flexible, complex
- Scalable. Draft standard.
- Recent research breakthrough in theory & algorithms

Ergo Benefits

- Full explanations in English, navigable in detail
- End users and subject matter experts are empowered
- Lower labor & cost. More agile.
- More automated, accurate
- Greater integration

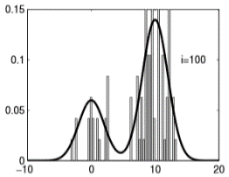
Actively Reason over Today's Gamut of Knowledge

Graph DB
& Semantic tech



Relational DB

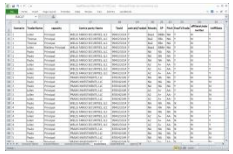
Machine
Learning



$\text{Prob}_N(x) = \Sigma_{i=1..N}\, \delta_{si}(x) / N$

Domain Apps
& Legacy

Spreadsheets



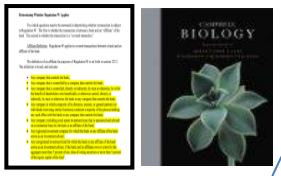
"Business Rules"

tier(X,1) ∧ supply(Y,X)
⇒ tier(Y,2)

Probabilistic engines

$p(H|C) = \Pi\, p(x_i|Y_j) / \Sigma_C\, \Pi\, p(x_i|Y_j)$

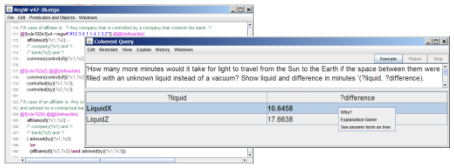
Text & Natural
Language processing



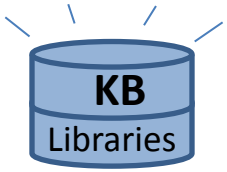
**External Info
& Services**

**Application
Actions**

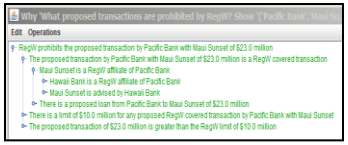
**Queries
Assertions
Edits**



ERGO



**Answers, Views
Decisions, Alerts
Explanations**



Outline

- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Next Generation Logic Methods

- **Rulelog** – based on major research advances
  - The logic itself is new, in addition to the innovative reasoning techniques
  - Very high expressiveness/reasoning power: higher-order, exceptions, probabilistic
  - Scalable computationally: bounded rationality ("restraint"), spirit of database logic
  - Orchestrates well: external queries & actions, standardization, high expressiveness
  - Open, adaptive, modular
- **Textual** Rulelog: combines closely with Natural Language Processing
  - Logic-based mappings between logic and English
    - English into logic for: assertions, queries
    - Logic into English for: answers, explanations
  - Very flexible in what one can say, and easy to change. Strong meta expressiveness.
  - Much easier to understand knowledge & reasoning: explanations, provenance

Series of Advances → Rulelog's Core Expressive Features

- Well-founded semantics; basic tabling algorithms
  - *Undefined* for paradox; smart cacheing; intuitionistic disjunction
- Higher-order syntax (Hilog); frame syntax
  - Associated optimizations of LP tabling etc. algorithms
- Statement id's for meta; argumentation meta-rules for defeasibility; provenance
- General formulas with all usual classical connectives and quantifiers (omniformity)
- Restraint bounded rationality
  - Use 3rd truth value *undefined* for "don't-care"
  - Radial, skipping; naf unsafety; external-query unsafety, unreturn

Additional Key Features on/in Rulelog

- Probabilistic uncertainty via evidential reasoning and distribution semantics
- Text generation using templates
- Explanations: extract justification graph, transform for presentation
- Text interpretation via templates, NLP parsers/tools

- Algorithms underlying, to: **transform**, compile, index, subgoal-reorder, dependency-aware update, externally query & import

KRR Features Comparison: Rulelog Shines

Feature \ System	Rulelog Rules - e.g., Ergo	Datalog Rules - e.g., Jena, SWRL, Ontobroker, SPIN	Production Rules - e.g., IBM, Oracle, Red Hat	Prolog - e.g., SICStus, SWI, XSB	FOL & OWL-DL - e.g., Vampire, Pellet, Prover9	ASP Solvers - e.g., DLV, CLASP
Semantic & on standardization path	✓	✓	restricted case	restricted case	✓	✓
Basic expressiveness						
• Datalog LP	✓	✓	✓	✓	✓	✓
• Logical functions	✓	✗	✗	✓	✓	✓
• General formulas	✓	✗	✗	✗	✓	✗
Full Meta expressiveness						
• Higher-order syntax, provenance	✓	✗	✗	✗	✗	✗
• Defeasibility & well founded negation	✓	✗	✗	✗	✗	✗
• Restraint bounded rationality	✓	✗	✗	✗	✗	✗
• Probabilistic	✓	✗	✗	✗	✗	✗
Efficiency						
• Goal-directed	✓	✗ (except Jena)	✗	✓	✓	✓
• Full LP tabling with dependency-aware updating	✓	✗	✗	✗ (except XSB)	✗	✗
• Polynomial time complexity	✓	✓	✓	✗	✗	✗

Notes on KRR Features Comparison

- "System" means system type / approach of logical knowledge representation and reasoning (KRR).
- "Semantic" means in the sense of KRR, i.e., fully declarative and having a model theory in the logical sense.
- "FOL" means First Order Logic. "ASP" means Answer Set Programs.
  - ASP is recently emerging. The tasks for which it's suitable are more similar to FOL than to the other systems here.
- "Standardization" here means industry standardization. "On path to" means in process of being, or already, standardized.
- "Restricted case" means for a syntactic/expressive subset.
- Event-condition-action rules in this context are similar to, and lumped in with, production rules.
- "LP" means declarative logic programs.
- Datalog means LP without logical functions. Usually this is restricted to Horn. But here we permit negation(-as-failure).
- OWL-RL is pretty much a restricted case of Datalog LP.
- "Higher-order syntax" means Hilog, which enables probabilistic – and also 1) fuzzy and 2) frame syntax cf. F-Logic.
- "Provenance" means provenance info about assertions, via properties of rule id's that are within the logical language / KRR.
- "Full" applies to all four of the meta expressiveness features.
- Defeasibility includes flexible argumentation theories.
- "General formulas" means classical-logic-like formulas, including with head existentials and with head disjunction.
- "LP tabling" includes sophisticated: cacheing of intermediate reasoning results, inference control, and indexing.
- "Dependency-aware updating" means that when assertions are added or deleted, saved inferences are only recomputed if they depend on the changes to the assertions.
- Polynomial time "complexity" means worst-case computational complexity, with constant-bounded number of variables per rule. Polynomial-time is similar to database querying, and is a.k.a. "tractable".

- Datalog X defeasibility: Ontobroker has full well founded negation.
- Prolog X defeasibility: XSB has full well founded negation.
- ASP X defeasibility: ASP has restricted defeasibility & well founded negation.
- Datalog X goal-directed: Jena has a backward engine as well as a forward engine.
- ASP X general formulas: ASP has head disjunction.
- FOL X full LP tabling with dependency-aware updating: Some FOL theorem-provers cache intermediate results in a way that is analogous to LP tabling, and some do dependency tracking but we're not sure how analogous or sophisticated.

Rulelog: Software Tools

- Rulelog Implementations with lots of expressiveness
  - Flora-2 (open source): Academic, not commercially supported
  - Ergo Suite (proprietary, from Coherent): Commercially supported
    - Ergo Suite greatly extends and enhances Flora-2

- Rulelog implementations with much smaller subsets of expressiveness
  - XSB Prolog (open source): Most of LP – with functions and well-founded negation.
  - Jena (open source): function-free negation-free LP, focused on RDF.
    - Similar: misc. other, e.g., that implement SWRL or SPIN

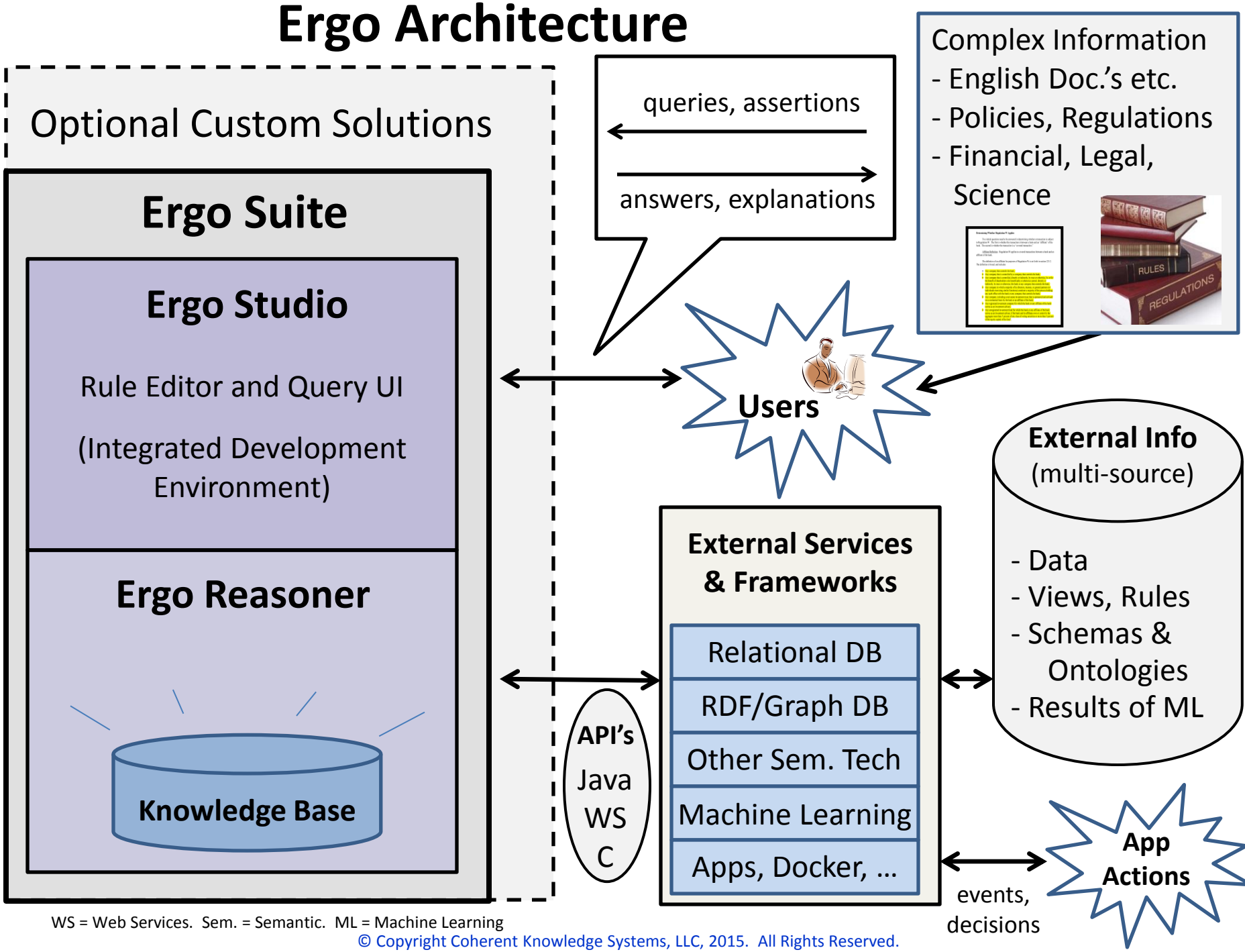
Ergo's Technological Approach

- Problem: high logical expressiveness is required for
  - Complex knowledge: policies, regulations, scientific
    - Often initially stated in natural language text
  - Data/knowledge integration mappings

- Solution: techniques in Coherent's Ergo platform for smart rules
  - *Rulelog* – fundamental logical knowledge representation
    - Full Meta expressiveness: higher-order, exceptions, probabilistic
  - *Textual* Rulelog – adds close relationship between text (English) and logic
  - *Explanations* – fully detailed, interactively navigable, in text (English)
    - Understandable to SMEs
  - *Optimizations* for large amounts of changing data

  ➢Address the 3 V's of big data, in combination with deep reasoning
    - Volume, velocity, variety

Ergo Architecture



WS = Web Services. Sem. = Semantic. ML = Machine Learning

Ergo Reasoner & Ergo Studio (IDE/UI)

- Textual Rulelog: Implementation of major research advances in logic (Rulelog) and how to map between logic and English (Textual Logic)
  - **Ergo is the most complete & highly optimized implementation available**
  - Rulelog significantly extends Datalog, the logic of databases, business rule systems (production/ECA/Prolog), semantic web ontologies, and earlier-generation semantic web rules cf. SWRL/RIF/RuleML
- Ergo Reasoner component – with sophisticated algorithms & data structures
  - Smart cacheing with dependency-aware updating. Leverages LP & DBMS techniques.
  - Transformation, compilation, reordering, indexing, modularization, dependency analysis, performance monitoring, virtual machine, programming kernel
  - Java API. Other interfaces: command line, web, C, knowledge interchange (below).
- Ergo Studio component – graphical Integrated Development Environment
  - Interactive editing, explanation, visualization of knowledge
  - Fast edit-test loop with award-winning toolset
- Knowledge interchange with leading and legacy systems
  - SQL, SPARQL, RDF, RDF-Schema, OWL. Others in dev or easy to add. Fully automatic.
- Open, standards-based approach. Builds on open source components (incl. XSB).
  - Rulelog is draft industry standard from RuleML (submission to W3C & Oasis)

Ergo Suite – Coherent Knowledge Management Platform

- Unprecedented flexibility in the kinds of complex info that can be stated as assertions, queries, and conclusions (highly expressive "knowledge" statements)
  - Almost anything you can say in English – concisely and directly
  - Just-in-time introduction of terminology
  - Statements about statements (meta knowledge) – contextualizes knowledge
  - State and view info at as fine a grain size as desired
- Probabilistic info combined in principled fashion, tightly combined with logical
  - Tears down the wall between probabilistic and non-probabilistic
- Unprecedented ease in updating knowledge
  - Map between terminologies as needed, including from multiple sources
- Conflict between statements is robustly handled (often arises during integration)
  - Resolved based on priority (e.g., authority), weighting, or else tolerated as an impasse
- Scalable and computationally well-behaved

Outline

- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Case Study 1: Automated Decision Support for Financial Regulatory/Policy Compliance

Problem: Current methods are expensive and unwieldy, often inaccurate

Solution Approach – using Textual Rulelog software technology:

- Encode regulations and related info as semantic rules and ontologies
- Fully, robustly automate run-time decisions and related querying
- Provide understandable full explanations in English
  - *Proof*: Electronic audit trail, with provenance
- Handles increasing complexity of real-world challenges
  - Data integration, system integration
  - Conflicting policies, special cases, exceptions
  - What-if scenarios to analyze impact of new regulations and policies

Business Benefits – compared to currently deployed methods:

- More Accurate
- More Cost Effective – less labor; subject matter experts in closer loop
- More Agile – faster to update
- More Overall Effectiveness: less exposure to risk of non-compliance

Demo of Ergo Suite for Compliance Automation: US Federal Reserve Regulation W

- EDM Council Financial Industry Consortium Proof of Concept – ***successful and touted pilot***
  - Enterprise Data Management Council (Trade Assoc.)
  - Coherent Knowledge Systems (USA, Technology)
  - SRI International (USA, Technology)
  - Wells Fargo (Financial Services)
  - Governance, Risk and Compliance Technology Centre (Ireland, Technology)
- Reg W regulates and limits $ amount of transactions that can occur between banks and their affiliates. Designed to limit risks to each bank and to financial system.
- Must answer 3 key aspects:

1. *Is the transaction's counterparty an <u>affiliate</u> of the bank?*
2. *Is the transaction contemplated a <u>covered transaction</u>?*
3. *Is the amount of the transaction <u>permitted</u> ?*

Determining Whether Regulation W Applies

Two initial questions need to be answered in determining whether a transaction is subject to Regulation W. The first is whether the transaction is between a bank and an "affiliate" of the bank. The second is whether the transaction is a "covered transaction."

<u>Affiliate Definition</u>. Regulation W applies to covered transactions between a bank and an affiliate of the bank.

The definition of an affiliate for purposes of Regulation W is set forth in section 223.2. The definition is broad, and includes:

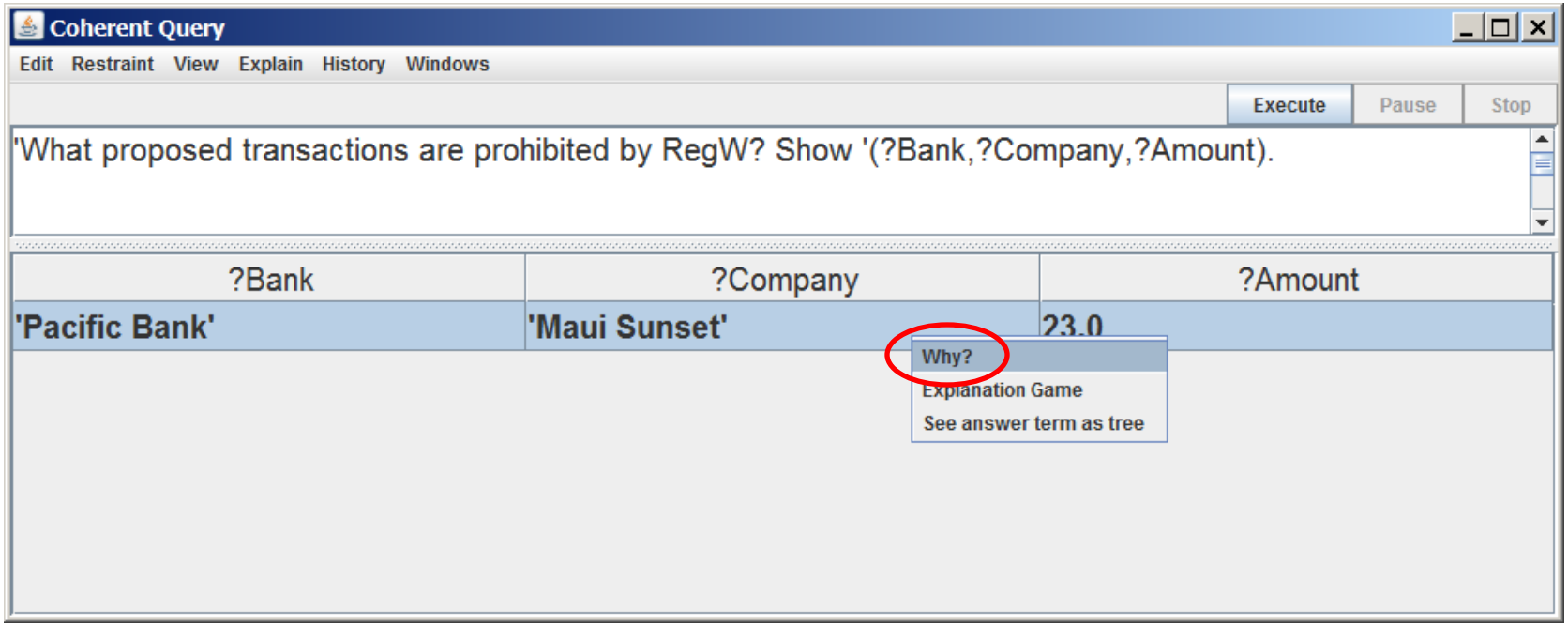
- Any company that controls the bank;
- Any company that is controlled by a company that controls the bank;
- Any company that is controlled, directly or indirectly, by trust or otherwise, by or for the benefit of shareholders who beneficially or otherwise control, directly or indirectly, by trust or otherwise, the bank or any company that controls the bank;
- Any company in which a majority of its directors, trustees, or general partners (or individuals exercising similar functions) constitute a majority of the persons holding any such office with the bank or any company that controls the bank;
- Any company, including a real estate investment trust, that is sponsored and advised on a contractual basis by the bank or an affiliate of the bank;
- Any registered investment company for which the bank or any affiliate of the bank serves as an investment adviser;
- Any unregistered investment fund for which the bank or any affiliate of the bank serves as an investment adviser, if the bank and its affiliates own or control in the aggregate more than 5 percent of any class of voting securities or more than 5 percent of the equity capital of the fund[1];

The Starting Point - Text of Regulation W

Demo goes here

- Note: The demo actually covers the next 10 slides that show explanation screenshots and executable rule assertions in Textual Rulelog. The deck includes those slides in order to be more self-contained. The demo also includes showing interactive syntax checking in the Ergo rule editor.

Query is asked in English



User Clicks the handles to expand the Explanations

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
    - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
    - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
      - Maui Sunset is a RegW affiliate of Pacific Bank
      - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
    - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why is the proposed transaction prohibited by Regulation W?

1. *Is the transaction's counterparty an "<u>affiliate</u>" of the bank?*

YES.

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - Maui Sunset is a RegW affiliate of Pacific Bank
      - Hawaii Bank is a RegW affiliate of Pacific Bank
        - There is common control of Hawaii Bank and Pacific Bank
          - Hawaii Bank is controlled by Americas Bank
            - Hawaii Bank is a subsidiary of Americas Bank
          - Pacific Bank is controlled by Americas Bank
            - Pacific Bank is a subsidiary of Americas Bank
      - Maui Sunset is advised by Hawaii Bank
    - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

And here's why …

Why is the proposed transaction prohibited by Regulation W?

2. *Is the transaction contemplated a "<u>covered transaction</u>"?*

YES.

And here's why …

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Su

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - Maui Sunset is a RegW affiliate of Pacific Bank
      - Hawaii Bank is a RegW affiliate of Pacific Bank
      - Maui Sunset is advised by Hawaii Bank
    - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why is the proposed transaction prohibited by Regulation W?

3. *Is the amount of the transaction <u>permitted</u>?*

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - There is an aggregated-affiliates limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with any affiliate
    - There is an individual-affiliate limit of $250.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The overall RegW limit of $10.0 million is the lesser of $10.0 million and $250.0 million
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

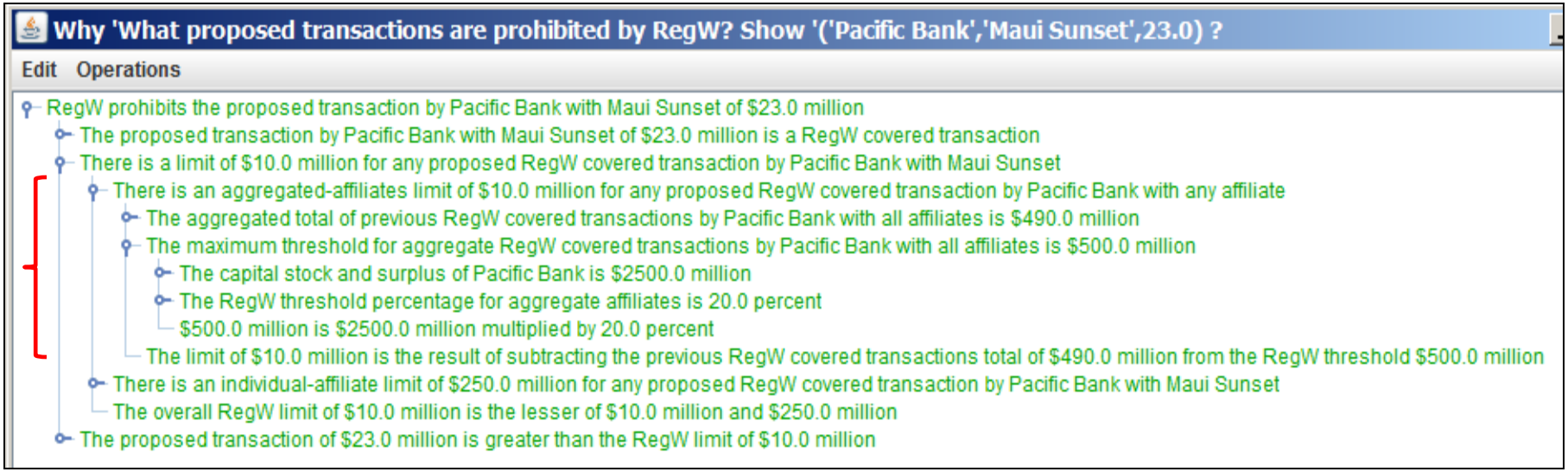
And here's why …

NO.

It went over the limit.

Why is the proposed transaction prohibited by Regulation W?

3. *(continued) Why is the aggregate-affiliates limit $10 million?*



Examples of the Underlying Textual Rulelog Executable **Fact** Assertions

- subsidiary(of)('Pacific Bank','Americas Bank').
- advised(by)('Maui Sunset','Hawaii Bank').
- bank('Hawaii Bank').
- company('Maui Sunset').
- capital(stock(and(surplus)))('Pacific Bank',2500.0).
- proposed(loan) (from('Pacific Bank'))(to('Maui Sunset')) (of(amount(23.0))) (having(id(1101))).
- previous(loan)(from('Pacific Bank'))(to('Hawaii Bank')) (of(amount(145.0))) (having(id(1001))).
- proposed(asset(purchase))(by('Pacific Bank')) (of(asset(common(stock)(of('Flixado'))))) (from('Maui Sunset')) (of(amount(90.0)))(having(id(1202))).

Executable Assertions: non-fact **Rules**

```
/* A company is controlled by another company when the first company
    is a subsidiary of a subsidiary of the second company. */
@!{rule103b} /* declares rule id */
@@{defeasible} /* indicates the rule can have exceptions */
 controlled(by)(?x1,?x2)
 :- /* if */
   subsidiary(of)(?x1,?x3) \and
   subsidiary(of)(?x3,?x2).

/*A case of an affiliate is: Any company that is advised on a contractual basis by
   the bank or an affiliate of the bank. */
@!{rule102b} @@{defeasible}
 affiliate(of)(?x1,?x2) :-
  ( advised(by)(?x1,?x2)
     \or
    (affiliate(of)(?x3,?x2) \and advised(by)(?x1,?x3))).
```

Executable Assertions: **Exception** Rule

```
@!{rule104e}
@{‘ready market exemption case for covered transaction'} /* tag  for prioritizing */
\neg covered(transaction)(by(?x1))(with(?x2))
          (of(amount(?x3)))(having(id(?Id))) :-
  affiliate(of)(?x2,?x1) \and
  asset(purchase)(by(?x1))(of(asset(?x6)))(from(?x2))(of(amount(?x3)))
     (having(id(?Id))) \and
  asset(?x6)(has(ready(market))).

/* prioritization info, specified as one tag being higher than another */
\overrides(‘ready market exemption case for covered transaction',
                 'general case of covered transaction').

/* If a company is listed on the New York Stock Exchange (NYSE), then the
    common stock of that company has a ready market. */
@!{rule201} @@{defeasible}
asset(common(stock)(of(?Company)))(has(ready(market))) :-
     exchange(listed(company))(?Company)(on('NYSE')).
```

Executable Assertions: **Import of OWL**

```
:- iriprefix fibof =  /* declares an abbreviation */
    "http://www.omg.org/spec/FIBO/FIBO-Foundation/20120501/ontology/".

/* Imported OWL knowledge:  from Financial Business Industry Ontology (FIBO) */
rdfs#subClassOf(fibob#BankingAffiliate, fibob#BodyCorporate).
rdfs#range(fibob#wholly OwnedAndControlledBy, fibob#FormalOrganization).
owl#disjointWith(edmc#Broad_Based_Index_Credit_Default_Swap_Contract,
                 edmc#Narrow_Based_Index_Credit_Default_Swap_Contract).

/* Ontology Mappings between textual terminology and FIBO OWL vocabulary */
company(?co) :- fibob#BodyCorporate(?co).
fibob#wholly OwnedAndControlledBy(?sub,?parent) :- subsidiary(of)(?sub,?parent).

/* Semantics of OWL - specified as general Rulelog axioms */
?r(?y) :- rdfs#range(?p,?r), ?p(?x,?y).
?p(?x,?y) :- owl#subPropertyOf(?q,?p), ?q(?x,?y).
```

Outline

- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Textual terminology

- Leverage Rulelog to much more simply and closely map between natural language and logic

- English phrase ↔ logical term in Rulelog
- English word ↔ logical functor in Rulelog

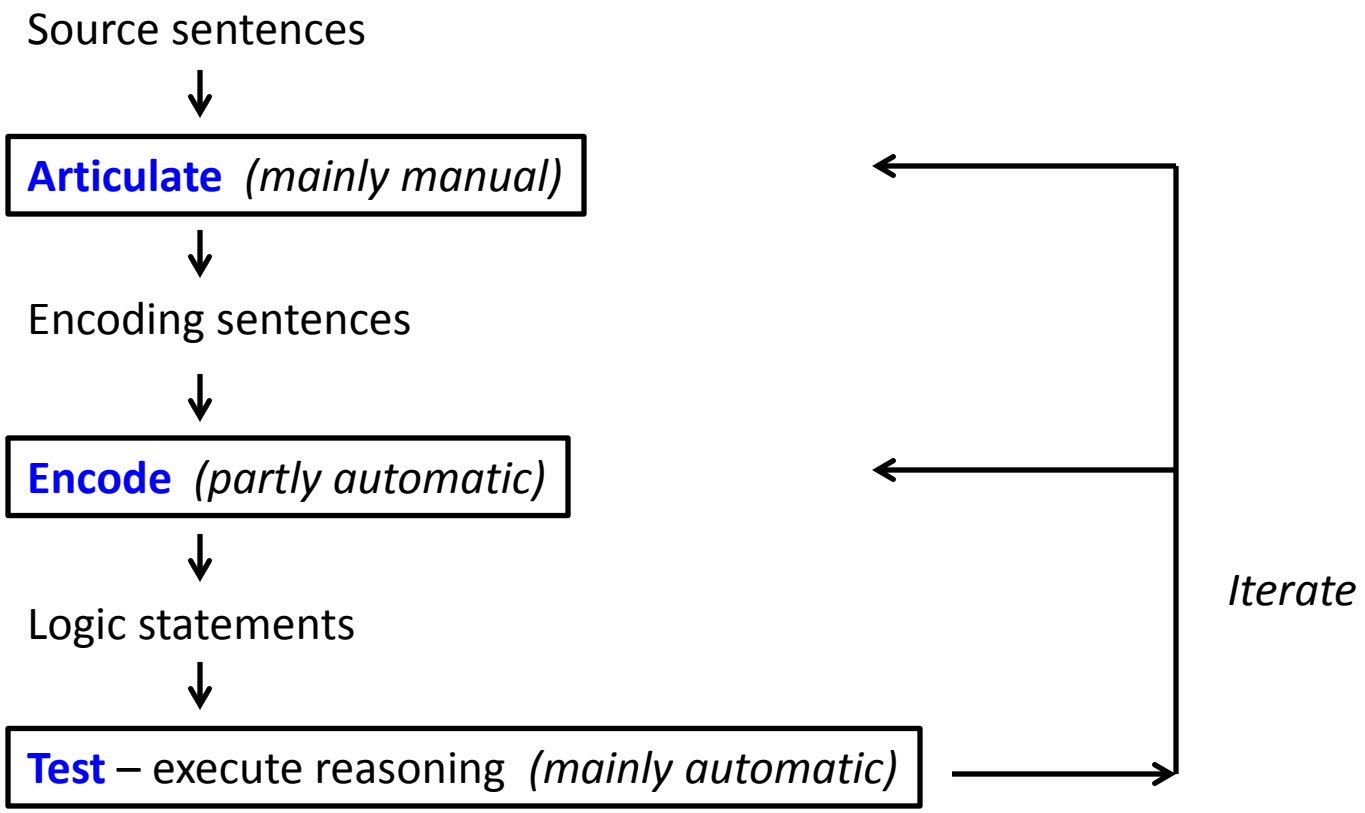
- Basis for textual templates

Knowledge Authoring Process using Ergo Suite

- Start with source text in English – e.g., textbook or policy guide
  - A sentence/statement can be an assertion or a query
- Articulate: create <u>encoding</u> sentences (text) in English.

  As necessary:
  - Clarify & simplify – be prosaic and grammatical, explicit and self-contained
  - State relevant background knowledge – that's not stated directly in the source text
- Encode: create executable logic statements
  - Each encoding text sentence results in one executable logic statement ("rules")
  - Ergo Suite has tools and methodology
- Test and debug, iteratively
  - Execute reasoning to answer queries, get explanations, perform other actions
  - Find and enter missing knowledge
  - Find and fix incorrect knowledge
  - Optionally: further optimize reasoning performance, where critical

Knowledge Authoring Steps using Ergo Suite



In-development: methods to greatly increase the degree of automation in encoding

Concept: Humagic Knowledge

- Humagic = human-machine logic
- A humagic KB consists of a set of linked sentences
  - Assertions, queries, conclusions (answers & explanations)
- NL-syntax sentence may have 1 or more logic-syntax sentences associated with it
  - E.g., that encode it, or give its provenance
- Logic-syntax sentence may have 1 or more NL-syntax sentences associated with it
  - E.g., that result from text generation on it
- Other sentences can be in a mix of NL-syntax and logic-syntax
  - ErgoText: templates used for text interpretation and text generation

Rulelog enriches Text Extraction

- Leverage Rulelog's high expressiveness and flexibility
- Mappings between multiple terminologies or ontologies

"moving a bomb" —*implies*→ "transporting weaponized material"

Outline

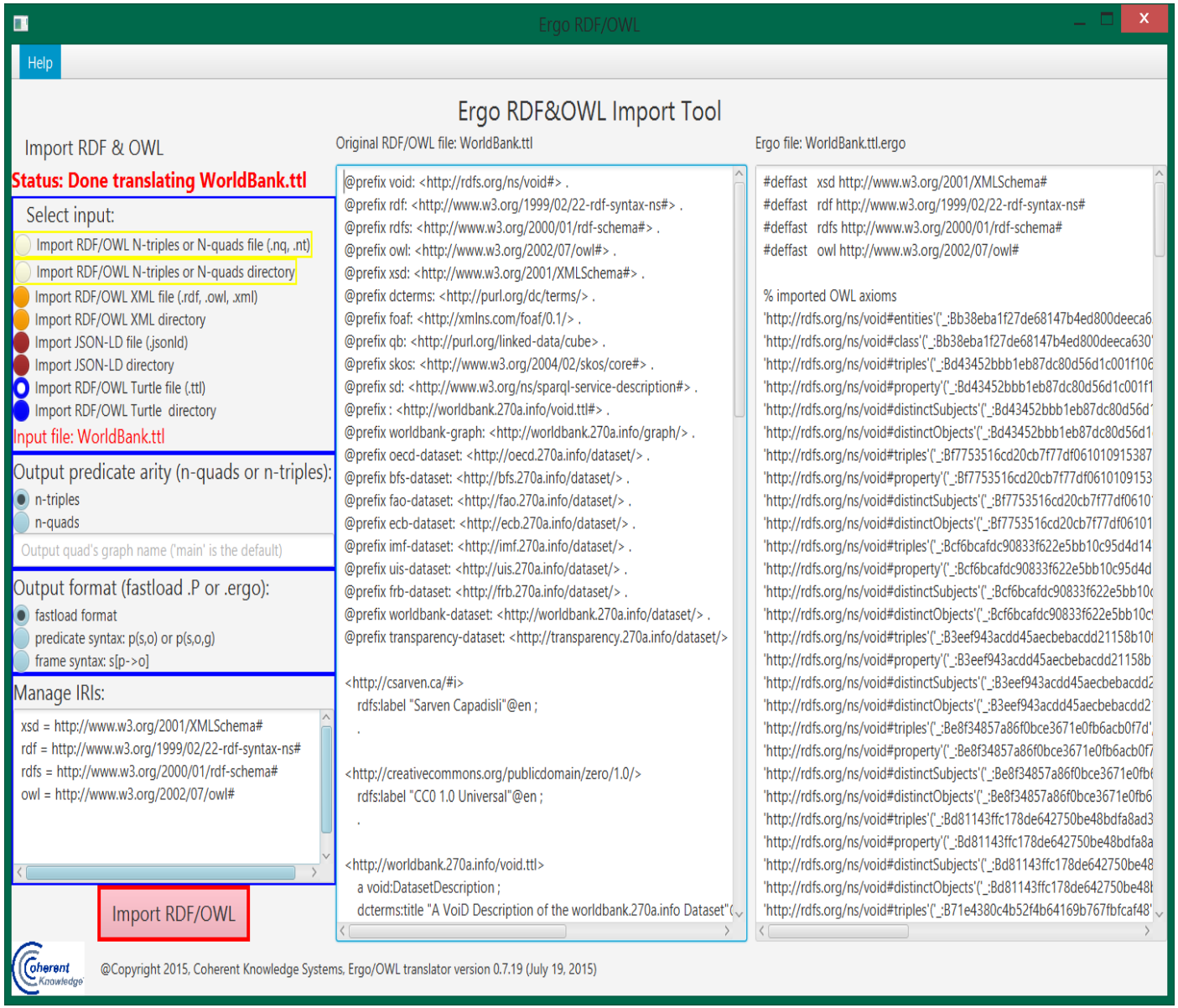
- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Importing Large Amounts of Data

- Problem – long time (many minutes) taken to load into the Rulelog in-memory knowledge base for reasoning, when there are many rules
  - A fact is a special case of a rule
  - But often there are LOTS of facts: e.g., Millions of RDF triples
- Solution – fastloader optimization for scaling
  - Streamlines processing of facts
  - >50X speedup: seconds not minutes

Importing RDF & OWL knowledge into Ergo

Screenshot of Ergo OWL connector part of Ergo Studio



Translates RDF & OWL to Ergo

Define IRIs in Ergo Studio

N-triples and N-quads

RDF/OWL XML, JSON-LD, or Turtle as input. Predicate or Frame syntax output.

Externally querying SPARQL, tightly integrating with endpoints

- Problem – How to leverage data and processing available from existing RDF triple stores, take advantage of their persistence and transactional etc. robustness
  - Often very large scale (e.g., Billions of triples)
- Solution: Ergo connector that
  - Goes out from Ergo to *dynamically* query via SPARQL to triple stores (interleaved during Ergo reasoning)
  - Translates results into Ergo and keeps reasoning
  - Uses Apache Jena libraries for translation, querying and integration of SPARQL endpoints (multiple distributed services that accept SPARQL queries and return results)

External SPARQL query within Ergo rule

```
db1(?Film, ?Title, ?Gross) :-
        sparqlOpen(?ConnectionID, ‘http://dbpedia.org/sparql’),
        sparqlQuery(?ConnectionID, ‘prefix : <http://dbpedia.org/ontology/>
                prefix prop: <http://dbpedia.org/property/>
                prefix xsd: <http://www.w3.org/2001/XMLSchema#>
                select ?film ?title ?gross where {
                        ?film a :Film;
                        rdfs:label ?title;
                        prop:gross ?gross
                        filter(langMatches(lang(?title), "EN"))
                        filter (datatype(?gross) = ‘http://dbpedia.org/datatype/usDollar')
                        }
                        order by desc(?gross)
                        limit 100’,
                        [?Film, ?Title, ?Gross]).

?- db1(?Film, ?Title, ?Gross).    /* Test query.  Ergo outputs the answer below */
        ?Film = ‘https://en.wikipedia.org/wiki/Harry_Potter_(film_series)’
        ?Title = ‘Harry Potter (film series)’
        ?Gross = $7,723,431,572
```

Updating of Inferred Data

- Problem: When the *asserted* rules or facts data change, stored previously *inferred* facts may no longer be valid. But recomputing a large set of inferred data may take a long time.
- Solution: dependency-aware updating via "incremental tabling" extension of tabled logic programming reasoning algorithm
  - Fast edit-test loop during knowledge authoring
  - Near-real-time decision automation

Outline

- Introduction: knowledge management, smart rules, deep reasoning, variety of knowledge
- Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Technology drill downs:
  - Human-machine logic: combining English and logic
  - Large changing data
- Discussion and Conclusions
  - More case studies. Machine learning.

Case Study 2: E-commerce marketing: Configuring Complex Products

- Very large manufacturing company
- Product catalog: highly technical, complex
- Over 1 Million products, in over 50 countries
- Problem: Creating configurators and updating catalog too slow & costly. Thus siloed by sales regions.
- Solution (with smart data partner) on $multi-Billion segment
  - Developed product ontology: compatible with legacy and standards
  - Converted all product data into live triple store
  - Added configuration rules, in Rulelog, on top of that catalog
  - Leveraged Ergo expressiveness and RDF import capability
- Benefits: faster, cheaper, agile updating, reusable globally
  - Enable greatly enhanced up-sell and cross-sell, across product lines

Case Study 3: Defense Intelligence Analysis

- Challenge: managing, accessing, integrating knowledge
  - Huge amounts of information as text, RDF, triple stores
- Current text extraction methods:
  - Noisy/inaccurate, shallow, patchy
  - Lacks contextualization, e.g., is a date past, present, or future?
  - Events are more complex than Entities, and are only shallowly treated

Lessons Learned from Case Studies

Financial, E-commerce, and Defense customers benefited from:

- Agility: Flexibility and ease of authoring, fast updating
- High accuracy and transparency
  - Explanations and provenance
  - Lower risk of non-compliance or confusion
- More Cost Effectiveness – less labor, SMEs in closer loop
- Improved volume, velocity, variety of data – with deep reasoning
  - Optimized loading of millions of facts (e.g., triples) as input assertions
  - Optimized dependency-aware updating of millions of inferred facts
  - External dynamic querying of triple stores
  - Mapping text ↔ logic using Rulelog methodology, e.g., for terminology
- Leveraging investment in Smart Data tech: RDF, SPARQL, OWL

“Smart Rules for Smart Data”, unpacked

- The smart rules work with smart data as **input** assertions
- The smart rules **infer** smart data
- The smart rules are fully declarative/semantic and themselves **become** rich data not procedural code
- The smart rules overall **leverage** investments in smart data

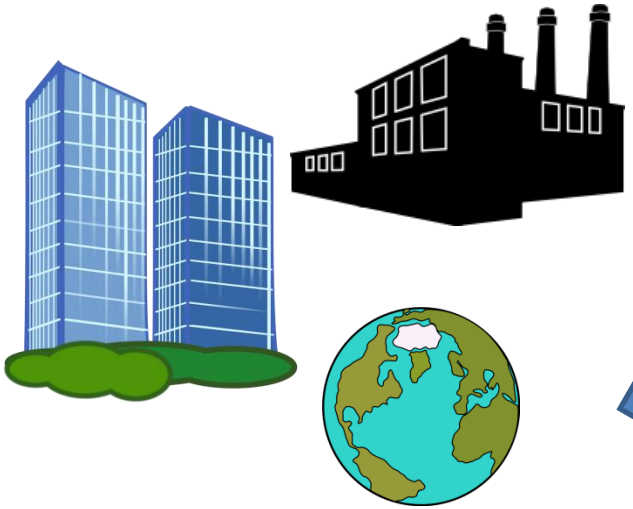
- Coherent helps customers develop and deploy smart rules KBs and applications, for a range of tasks and domains
  - Via its Ergo platform **product** capabilities
  - Via its professional **services**

Next Step: Combining ML and KRR

- Core AI = KRR + ML. KRR is required for Cognitive Computing.
- The prediction step of ML requires reasoning
- The target of ML is a representation
- Getting value from ML requires reasoning for analysis and decisions
- KRR is required to combine results of ML, accumulate knowledge coherently, and explain knowledge
  - Weaknesses of ML today
- Reasoning to supply derived facts for ML to chew on
- Reasoning to focus ML's tasks and conjecture schemas
  - e.g., sets of relevant features, important questions, to drive ML

Evolution of Enterprise Knowledge Management

(repeat slide)

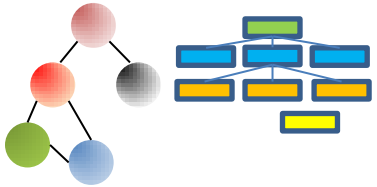


Data integration with flexible schemas, more meta-data. E.g., RDF, SPARQL, OWL.

Augment relational/traditional data stores.

Smart Data

Graph/Linked Data bases



Smart Rules

Decisions & Analysis

Deep reasoning, more complex knowledge. E.g., Textual Rulelog.

Leverage Smart Data investment.

Thank You.

Smart Rules for Smart Data™