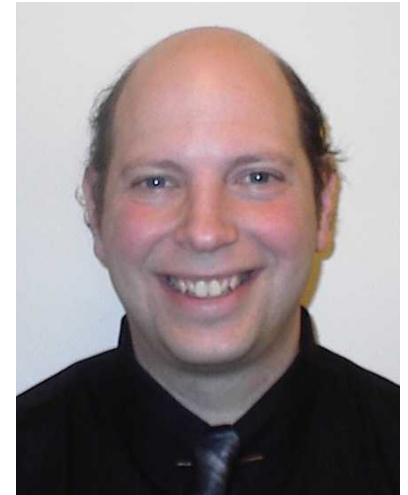# *Opportunities for Semantic Web knowledge representation to help XBRL*

by **Benjamin Grosof**

Senior Research Program Manager, Vulcan Inc.; and
Principal Consultant, Benjamin Grosof & Associates, LLC
http://www.mit.edu/~bgrosof

Workshop on Improving Access to Financial Data on the Web

http://www.w3.org/2009/03/xbrl/program.html
Co-organized by XBRL International and World Wide Web Consortium (W3C)
hosted by FDIC, Arlington, VA
October 5-6, 2009

Relevant Background of Presenter

- Involved with XBRL since 2001
  - Research on semantic rules for financial reporting and info integration
    - E.g., contextual mappings between ontologies incl. equational-form
  - Informal scientific adviser to XBRL Intl. and W3C during 2005-2007

- Senior Research Program Manager, Vulcan, Inc. (Paul G. Allen's co.)
  - Leads SILK and Advanced Research thrust in Vulcan's Project Halo
    - Game-changer rule-based knowledge representation language & system
  - Also advises Venture Capital arm (leading investor in the space)
- Principal, Benjamin Grosof & Associates (consulting – part time, since MIT days)
- Formerly MIT Sloan IT professor (2000-2007), IBM Research (1988-2000)
- Pioneer/inventor of semantic rules for web and enterprises. Basis for:
  - Main web industry standards
    - W3C RIF (Rule Interchange Format)
      - Also: the RuleML standards design it was based on
    - OWL RL (Web Ontology Language's Rules Profile)
  - Business applications pilots and strategy roadmaps
    - E-commerce; trust; finance; mobile; biomed; etc.
  - Oracle's and IBM's pioneering semantic rules products

Outline

1. Overview

2. Drill down:
   - Uses of Semantic Rules for XBRL
   - Knowledge Representation (KR) & Semantic Rules on Web
     (if there's time)

3. Wrap-up: Directions

History of Parallax

- XBRL and Semantic Web have historically evolved separately and in parallel with each other
  - Began about the same time
  - Communities mainly non-overlapping
- Convergent evolution, to a considerable degree
  - Data, ontologies, and rules – in XML markup

- *Now*: large opportunities for synergy
  - Leverage/share technical approaches
    … and application domains

What Semantic Web can offer XBRL

1. Sophistication in knowledge representation (KR)

   Overall: expressiveness; interoperability; performance optimization

   – Data representation

      Better for wide sharing than plain XML

      • RDF unordered directed-graph with Webized naming

   – Semantic rules

      Better for wide sharing than previous kinds of business rules

      • Many uses specific to business reporting and financial info integration

      • Can handle exceptions, change/updates, reformulations gracefully

         – "Defaults", "defeasibility", "logical nonmonotonicity", "prioritized conflict handling"

What Semantic Web can offer XBRL (cont.'d)

2. Sophistication in knowledge acquisition (KA)
   – Target "business users"/"subject matter experts", collaboration
     • UI & KR for authoring & explanation
       – Controlled natural language; tabular and graphical UI metaphors
       – Semantic wikis, e.g., Vulcan's Semantic MediaWiki+ (SMW+) – semantic enhancement of Wikipedia software
   – Via knowledge interchange (translation between KR formats)
     • E.g., legacy database (DBMS) and business rule systems (BRMS)

3. Related domain ontologies and knowledge bases (KBs)

4. "Virality" to more applications, methods, domains
   ⇒ e-commerce, health, security, media/social, BI/marketing, …

What XBRL can offer Semantic Web

- A centrally important domain of knowledge and tasks

- Practicality
  - Potentially a platform for wide acceptance & adoption of SW

- Firm connection to XML-Schema

- $\Rightarrow$ *Incentive for SW community to collaborate with XBRL*

Outline

1. Overview

2. Drill down:
   - ➢ Uses of Semantic Rules for XBRL
   - ➢ Knowledge Representation (KR) & Semantic Rules on Web
     (if there's time)

3. Wrap-up: Directions

Uses of Semantic Rules for XBRL

- Ontology mappings: contextual, reformulation
  - *Examples*:
    - Price with vs. without shipping, tax
    - Earnings last 4 qtrs vs.{last 3 qtrs + forecast next qtr}
    - Profit with vs. without depreciation
    - Historicals when statutory treatment changes
    - Implicit context: use a typical definition of revenue
  - Your vs. my pro-forma or analytic view
    - Between companies, governmental jurisdictions
  - Exception handling, special cases, one-time events
    - Footnotes – "where the real action is"
    - *Example:* Revenue includes sale of midtown NYC headquarters bldg

Uses of Semantic Rules for XBRL, cont.'d

- Policies, "rules", regulations, laws
    - Trust: confidentiality, access, authorization
    - Regulatory & legislative, compliance & governance
    - Other business and government operations

- Analytics & monitoring
    - Queries & (persistent) views
        - Supersumes SQL, SPARQL, XQuery
    - Contextualized (see last slide)

- Decisions & triggered actions – automated, monitor-based

Example: Exception in Ontology Translation (in SILK)

/* **Company BB reports operating earnings using R&D operating cost which includes price of a small company acquired for its intellectual property. Organization GG wants to view operating cost more conventionally which excludes that acquisition amount. We use rules to specify the contextual ontological mapping.** */

@{normallyBringOver} ?categ(GG)(?item) :- ?categ(BB)(?item);

@{acquisitionsAreNotOperating} neg ?categ(GG)(?item) :-
    acquisition(GG)(?item) and (?categ(GG) ## operating(GG));

overrides(acquisitionsAreNotOperating, normallyBringOver); /* exceptional */

acquisition(GG)(?item) :- price_of_acquired_R_and_D_companies(BB)(?item);

R_and_D_salaries(BB)(p1001);   p1001[amount -> $25,000,000];

R_and_D_overhead(BB)(p1002);   p1002[amount -> $15,000,000];

price_of_acquired_R_and_D_companies(BB)(p1003); p1003[amount -> $30,000,000];

R_and_D_operating_cost(BB)(p1003); /* BB counts the acquisition price item in this category */

R_and_D_operating_cost(GG) ## operating(GG);

Total(R_and_D_operating_cost)(BB)[amount -> $70,000,000]; /* rolled up by BB cf. BB's definitions */

Total(R_and_D_operating_cost)(GG)[amount -> ?x] :- … ; /* roll up the items for GG cf. GG's definitions */

As desired: |= **R_and_D_salaries(GG)(p1001); …**

neg R_and_D_operating_cost(GG)(p1003); /* GG doesn't count it */

Total(R_and_D_operating_cost)(GG)[amount -> $40,000,000];

Notation: @{…} encloses a rule label. ? prefixes a variable. :- means if. X ## Y means X is a subclass of Y. overrides(X,Y) means X is higher priority than Y.

Advantages of Standardized SW Rules

- Easier Integration: with rest of business policies and applications, business partners, mergers & acquisitions
- Familiarity, training
- Easier to understand and modify by humans
- Quality and Transparency of implementation in enforcement
  - Provable guarantees of behavior of implementation
  - Improved compliance and governance
- Reduced Vendor Lock-in
- Expressive power
  - Principled handling of conflict, negation, priorities

Outline

1. Overview

2. Drill down:
    - ➢ Uses of Semantic Rules for XBRL
    - ➢ Knowledge Representation (KR) & Semantic Rules on Web
      (if there's time)

3. Wrap-up: Directions

Background: What is Knowledge Representation (KR)?

- The *field* of KR studies and designs particular knowledge representation languages/systems (KR's).

- A KR includes:

  - A formal language for expressing <u>premises</u>.

  - A formal language for expressing <u>conclusions</u>.

  - A set of <u>entailment principles</u> that together, for any given set of premises, formally defines an associated set of sanctioned conclusions.

    - In "declarative" KR, these principles are independent of inferencing procedure/control-strategy, and thus constitute a <u>semantics</u>, e.g., a model theory

Background: Example KRs

1. Relational <u>databases</u>: relational algebra (SQL)
    - ≡ a subset of LP (below)
2. Mathematical <u>classical logic</u>: first-order logic (FOL)
    - Severe drawbacks for Web: brittleness, unscalability
3. Semantic rules, a.k.a. declarative Logic Programs (LP)
    - Basic case: Horn, e.g., W3C RIF Basic Logic Dialect (BLD)
    - State of art full-featured: Hyper LP, e.g., SILK, RuleML
4. Many others
    - Commercial rule and conceptual-modeling systems
    - Bayesian networks, inductive learning, fuzzy logic, etc.

LP is <u>the</u> Core KR in today's world … incl. Sem. Web

- **LP is the core logical KR of structured knowledge management today**
  - **Databases**
    - Relational / SQL
    - XML semi-structured / XQuery
    - RDF semi-structured / SPARQL (triple stores)
  - **Semantic Rule Standards**
    - RuleML standards design
    - Rule Interchange Format (RIF)**
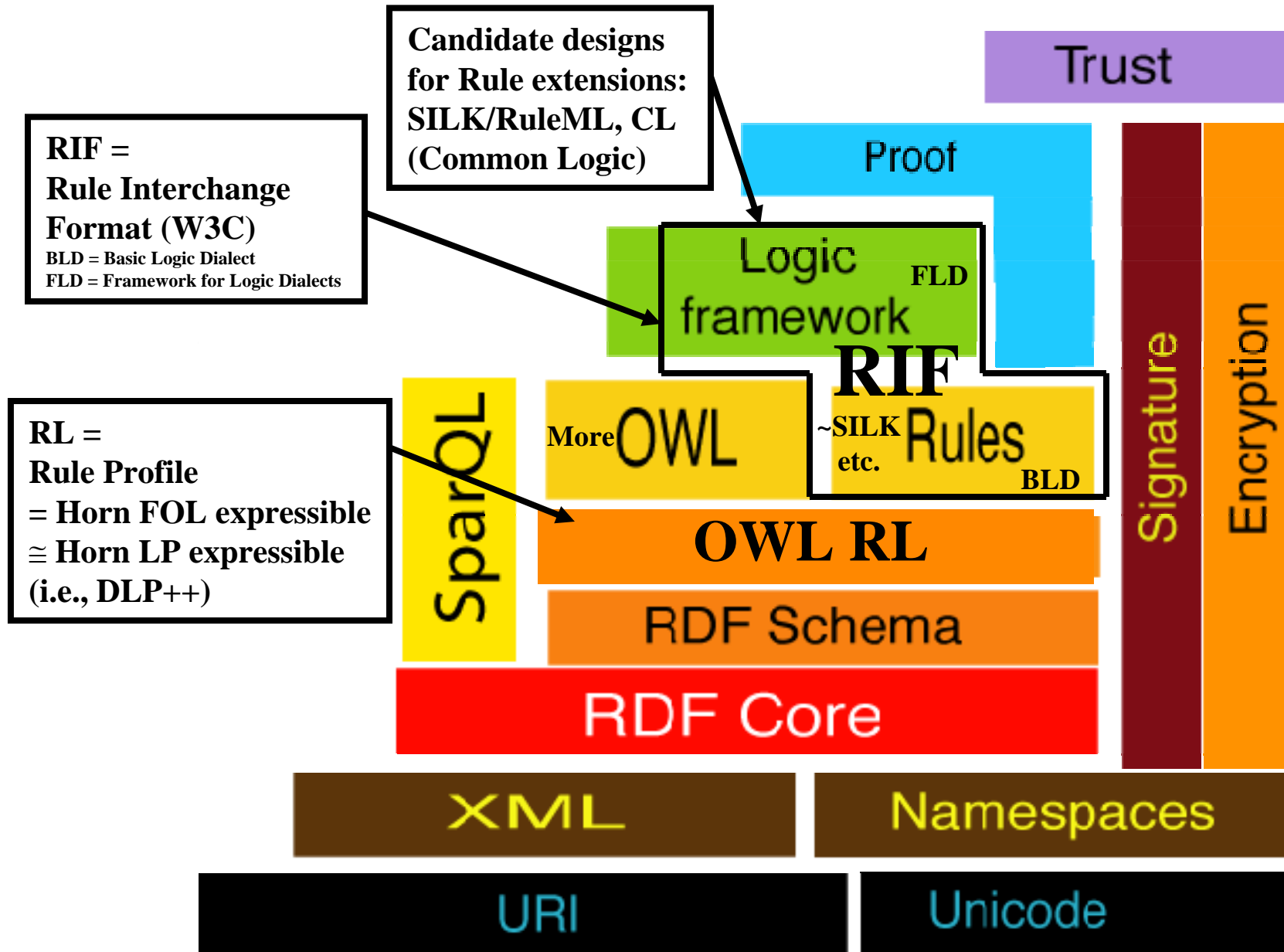 - **Semantic Ontologies**
    - Most commercial implementations of OWL are based on semantic rules: Description Logic Programs (DLP) + moderate extensions. E.g., Oracle.
    - OWL 2** standard includes the RL Profile, i.e., its Rules subset

- **The Semantic Web today is mainly based on LP KR**
  - … and thus essentially equivalent to semantic rules
  - **You probably just didn't realize it!**

** W3C Last Call Working Draft

Updated: 10-2009 Semantic Web Standards "Stack"

Candidate designs for Rule extensions: SILK/RuleML, CL (Common Logic)

**RIF =
Rule Interchange Format (W3C)**
BLD = Basic Logic Dialect
FLD = Framework for Logic Dialects

**RL =
Rule Profile
= Horn FOL expressible
≅ Horn LP expressible
(i.e., DLP++)**

Trust

Proof

Logic framework   FLD

RIF

SparQL

More **OWL**   ~SILK etc.   **Rules**   BLD

OWL RL

RDF Schema

RDF Core

Signature   Encryption

XML   Namespaces

URI   Unicode

SILK & Hyper LP: Overview

- **A KR Language and KR System with reasoner, UI, interchange**
  - Syntax & semantics, forward & backward inferencing, API, translators

- **Goal: Expressiveness + Semantics + Scalability + Web**

- **Focus: Defaults and Processes, for question-answering in e-science & e-biz**

- **Largest rule research program in the US (that we're aware of)**

Outline

1.   Overview

2.   Drill down:
   - ➢   Uses of Semantic Rules for XBRL
   - ➢   Knowledge Representation (KR) & Semantic Rules on Web
     (if there's time)

3.   Wrap-up: Directions

What's Coming in Semantic Rules

- Semantic Rules will increase over the next several years
  - In adoption
    - Sem Tech sector
    - BRMS sector
  - In expressiveness
    - W3C RIF includes a framework for extensions
      - Framework for Logic Dialects (FLD)
    - Likely to include extension similar to SILK
      - Plan to propose in 2010 a new RIF dialect under FLD

KR Challenges Needing Applied Research

- Numerical reasoning more closely integrated with symbolic reasoning
  - Equalities and equations
  - Inequalities, "constraints"
  - Money
  - Time (and dates)

- KR context mappings (reformulations)

- *OWL and RDF are quite weak in the above areas*
  - *Semantic rule approaches, e.g., SILK, are much more capable*

Individual and Community To Do's

- Learn about Semantic Web and KR

- Think about how to use their strengths in your org/task

  – Plan ahead wrt design choices, based on requirements and what's coming soon

- Standardization of more expressiveness

  – esp. semantic rules

- Applied research on numerical and reformulation aspects of KR (see last slide)

Resources For More Info

- Generally:
  - Personal webpage http://www.mit/edu/~bgrosof/
  - SILK webpage http://silk.semwebcentral.org

- Specifically (available at the above sites):
  - Position paper from this Workshop (the one by me)
    - "Links and References", especially
  - Tutorial "Rules on the Web" (co-authored by me)
    - Half-day presentation at Intl. Semantic Web Conf. 2009 (upcoming on Oct. 26)

Thank You

Disclaimer: The preceding slides represent the views of the authors only.
All brands, logos and products are trademarks or registered trademarks of their respective companies.