

Case Studies in Applying Semantics to Enterprise Systems

Dave McComb, Semantic Arts
February 2011

Semantic Arts

- ▶ Small consulting firm, specializing in helping large organizations apply semantic technology to their enterprise architectures

Semantic Arts' Clients



The logo for SallieMae, featuring the brand name in a bold, white, sans-serif font with a registered trademark symbol (®) to the upper right of the 'e'. The text is centered within a solid dark blue rectangular background.

SallieMae[®]



Sallie Mae

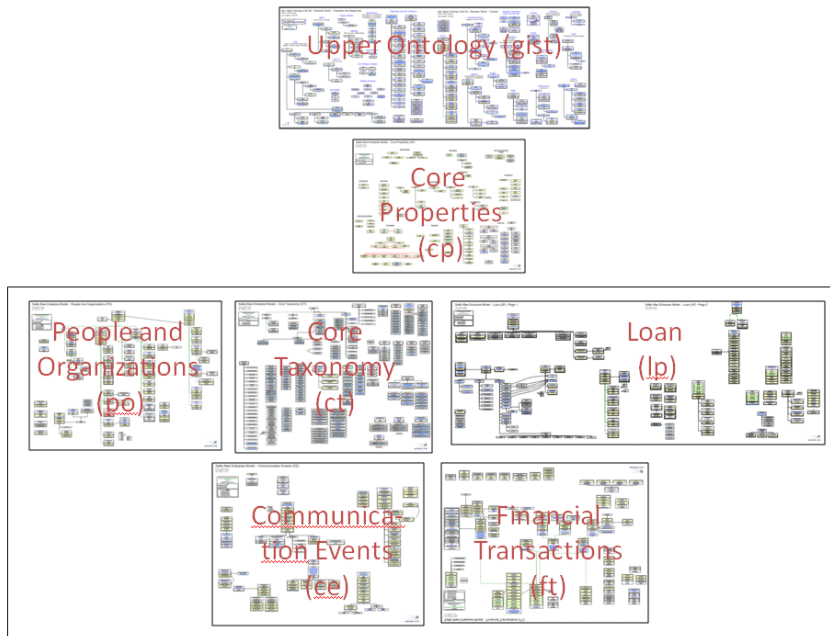
- ▶ Leading provider of student loans
- ▶ We built an Enterprise Ontology for them in early 2009.
- ▶ In late 2009 they had an opportunity to apply it...

Getting a handle on complexity

	tables	attributes
Class	582	10,230
LoanCons	133	15,295
Eagle I	356	13,538
Eagle II	464	12,502
	1,535	51,565

These are the number of distinctions being made in the current systems

Sallie Mae Enterprise Model – May 2009



Classes	574
Object Properties	250
Data Type Properties	38
Total T-Box Axioms	1470

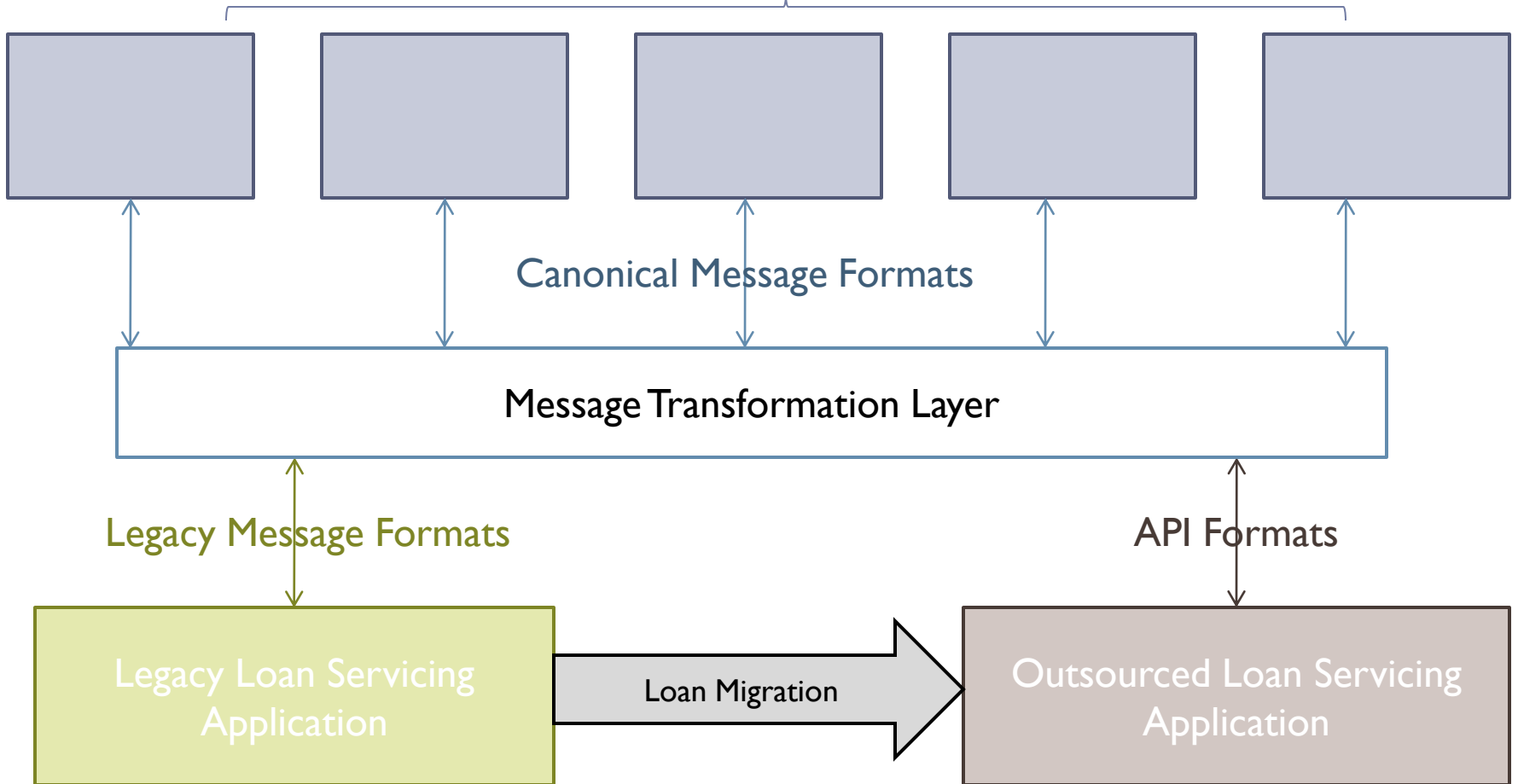
The original goals of the Sallie Mae Enterprise Semantic Model were to:

- ▶ Create **formal business definitions** of the principal concepts in use across the organization.
- ▶ Validate the model against existing data bases and interfaces, and start the process of **formally describing the existing data** using those enterprise definitions.
- ▶ Provide a basis for integrating structured and **unstructured data**.

Outsourcing Initiative

Customer-Facing Applications

Customer facing applications would be rewritten to use the canonical message formats.



Class Comparison

Sub Ontology	May 2009	March 2010
Loans	180	340
Communication	96	123
Social Beings	119	146
Finance	117	209
Core Properties	3	4
Core Taxonomy	99	284
Identifiers	21	56
Gist	130	129
GistComp		65
Message Model		134
CLASS specific (FinTran Codes)		130
All	610	1284

Mostly in the loan subject area as more detail on loan servicing events was added.

Instance taxonomies were converted to classes

Many new classes were specific to the Message Model class

Total doubled

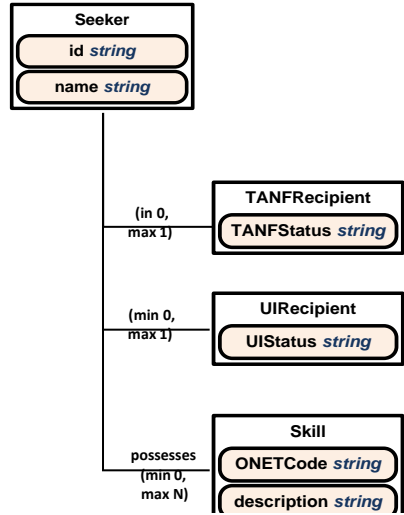
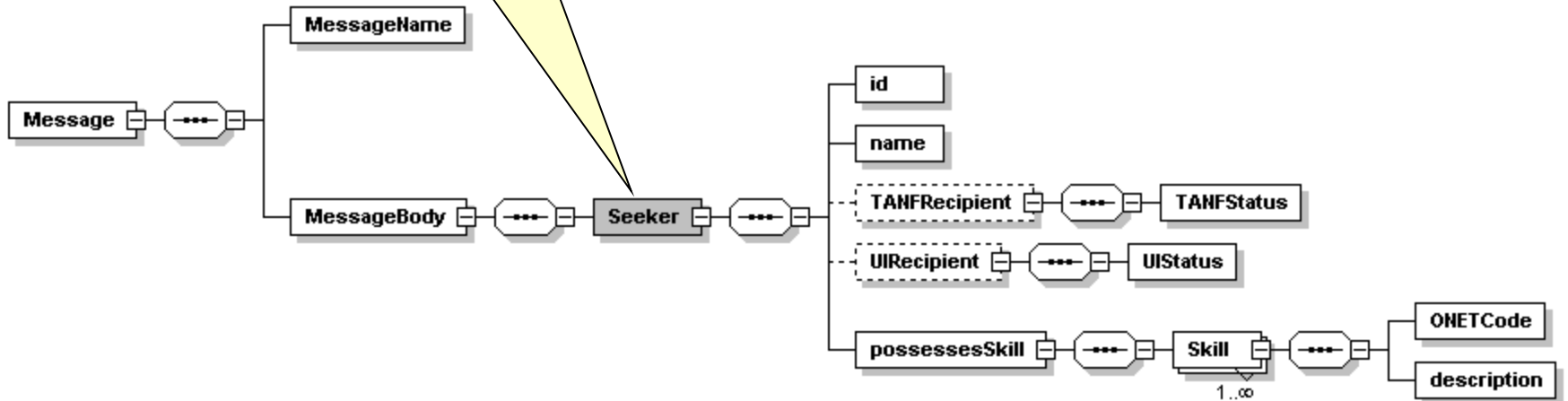
Properties (Object/Datatype)

SubOntology	May 2009	March 2010
Loans	43/1	61/0
Communication	31/0	32/10
Social Beings	46/8	49/7
Finance	35/0	31/1
Core Properties	148/32	186/15
Core Taxonomy	4/0	2/0
Identifiers	2/1	2/2
A lot of the net increase was in gist.	75/11	119/20
GistComp		42/0
Message Model		26/2
CLASS specific (FinTran		15/0
Surprisingly the total number of properties went up far less.	225/44	317/36

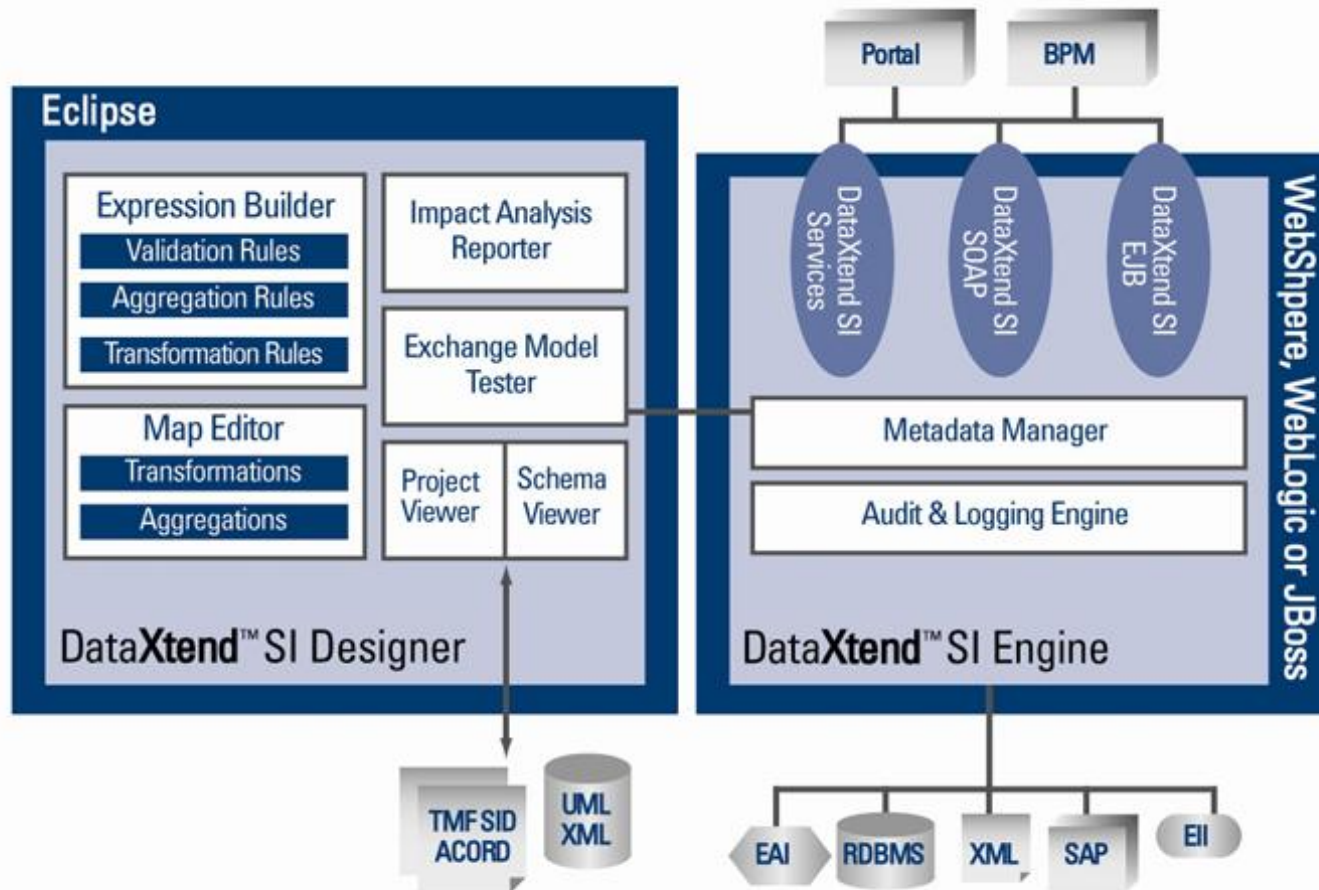
The Projection becomes the XSD Message Definition

The message body for the "BasicSeeker" message type is defined in XSD based on the projection.

XMLSPY XSD Design View

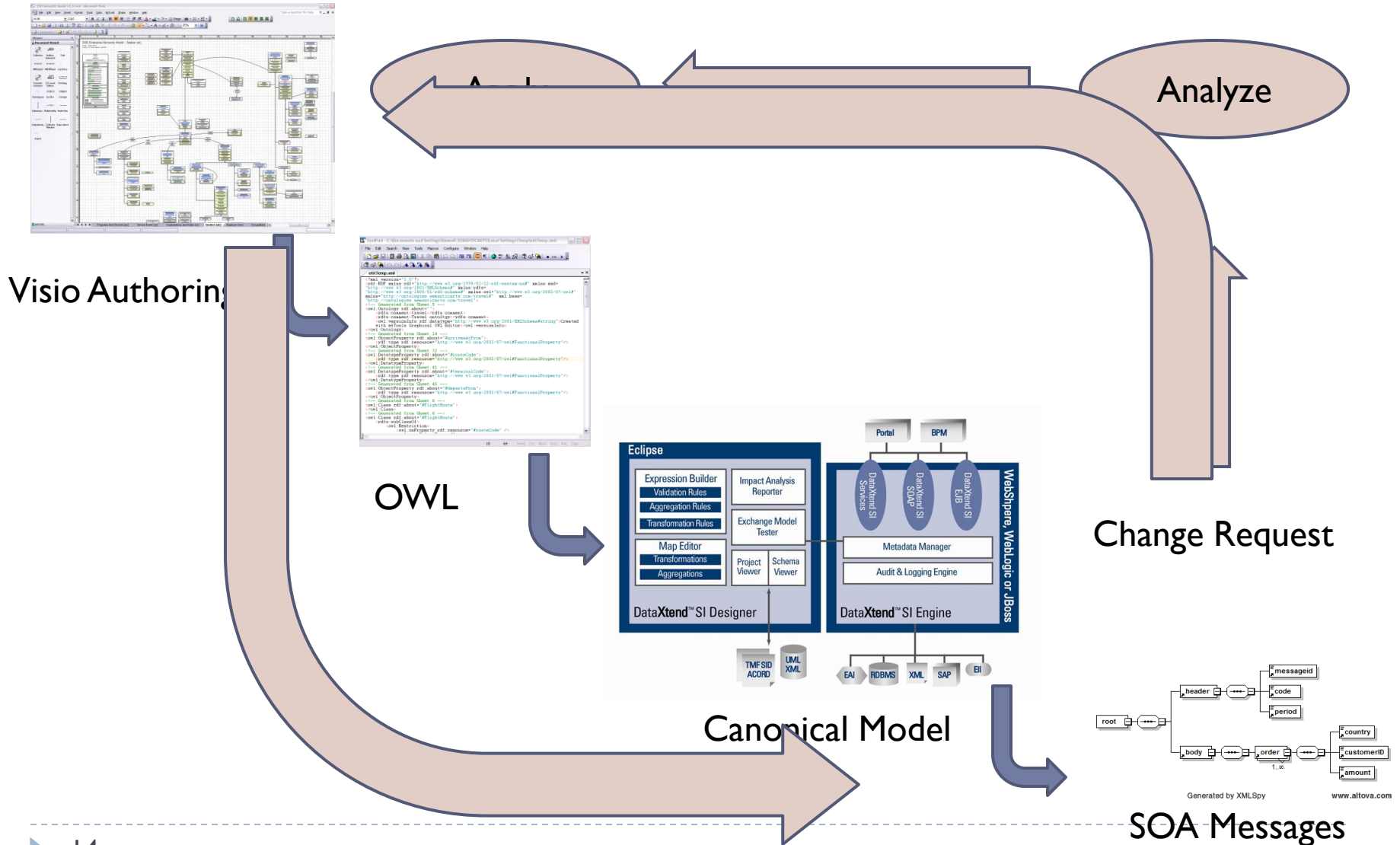


Progress/Data Extend (DXSI)



Toolset

Full loop about 1-3 hours



Net Result

- ▶ New outsourced servicing system was integrated into Sallie Mae's environment.
- ▶ One set of SOA messages handles both servicing systems.
- ▶ The rationalization of the messages was made possible by the enterprise ontology.
- ▶ Changes could be rapidly incorporated into the ontology and their impact reflected in messages within hours.



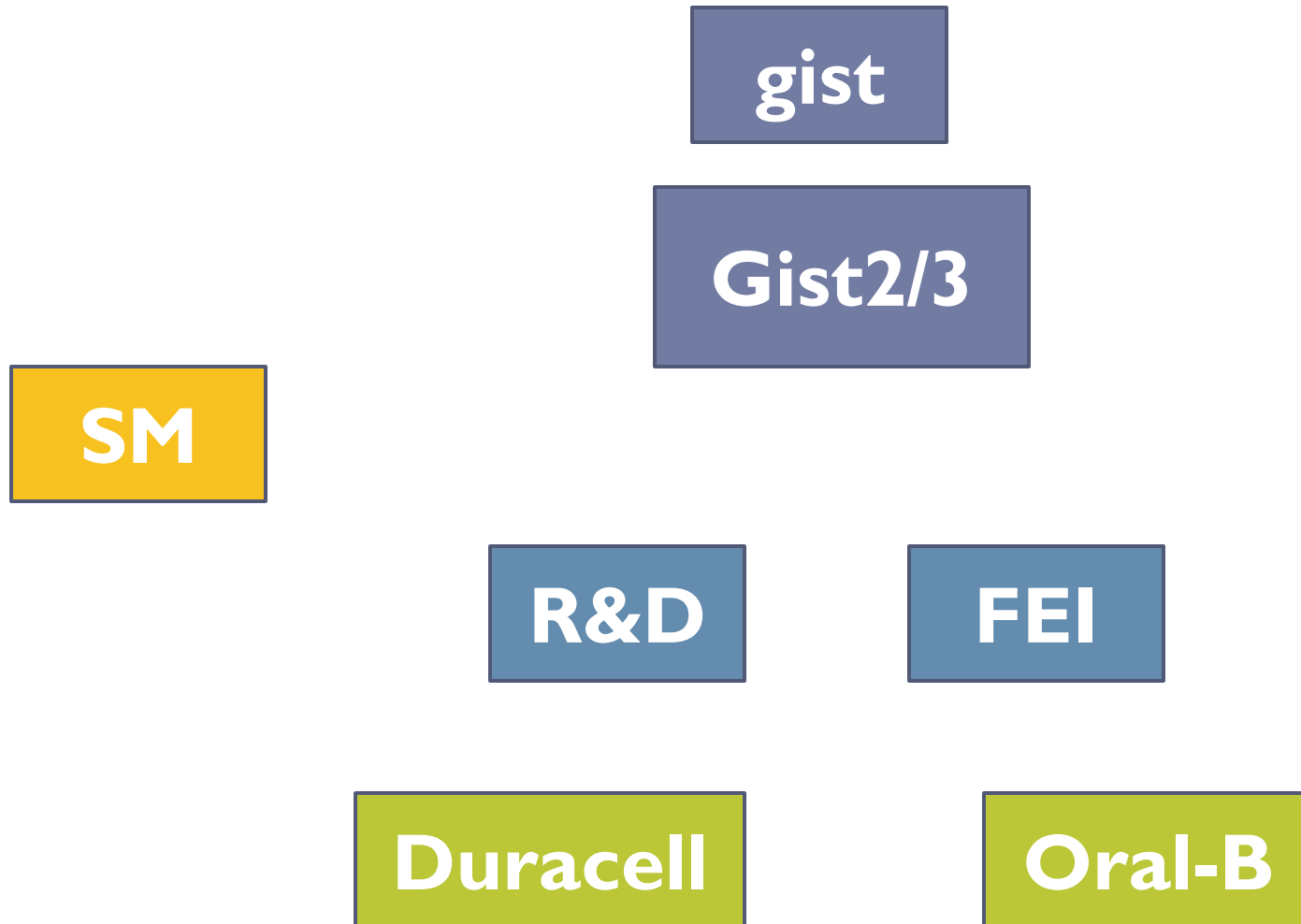
Procter & Gamble – Harvesting Knowledge from Researchers

- ▶ Large consumer products company
- ▶ Looking for ways to integrate research findings across disciplines
 - ▶ Over 10,000 researchers in nearly 100 disciplines
 - ▶ Each discipline has its own language
 - ▶ Traditional key word search not useful when searching across domains
- ▶ Problem compounded by departure of many key researchers (retirement, re-organization, etc.)

Work Performed

- ▶ We built an Enterprise Ontology for the R&D domain.
- ▶ In parallel with interviewing retiring researchers from two divisions: Duracell and Oral-B.

Structure of the model(s)



How the ontologies are layered

	Gist	P&G General R&D	With Dura & OralB
Classes	233	410 (added 177)	593 (added 183)
Object Properties	170	192 (added 22)	196 (added 4)
Data Properties	20	20	25 (added 5)

Upper Ontology Coverage

- ▶ Of the nearly 600 classes in the R&D ontology
- ▶ Only 2 were not derived from gist:
 - ▶ Brand
 - ▶ Invention
- ▶ Most R&D data is findable without needing to know the specialized dialect of each subdomain.

Results

- ▶ Semantic Wiki built based on ontology
- ▶ Two additional domains have been modeled (feminine care and baby care) and both reinforce the original abstractions
- ▶ Additional domains planned for this year





LexisNexis

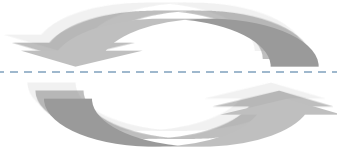
- ▶ Leading supplier of legal research
- ▶ Currently legal annotation is done by hand, an “editorial” process, or through scripts that hard automate the classification process.
- ▶ They recognize that they are running to the limits of this approach, at the same time that demand for more appropriate retrieval is climbing.

LexisNexis

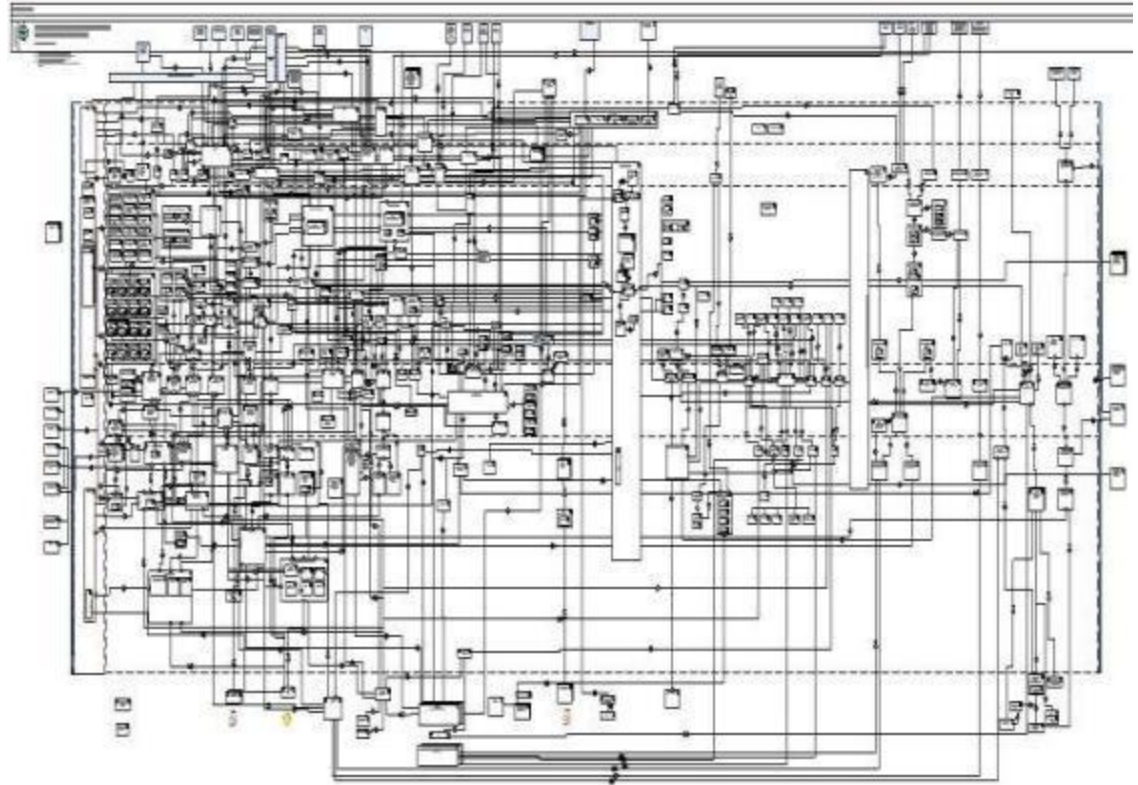
- ▶ They have launched a major initiative to convert their systems to be semantically based.
- ▶ Raw text will be processed to extract not only entities but relationships as well.
- ▶ This extracted information will be conformed to the new Enterprise Ontology.

Current Situation

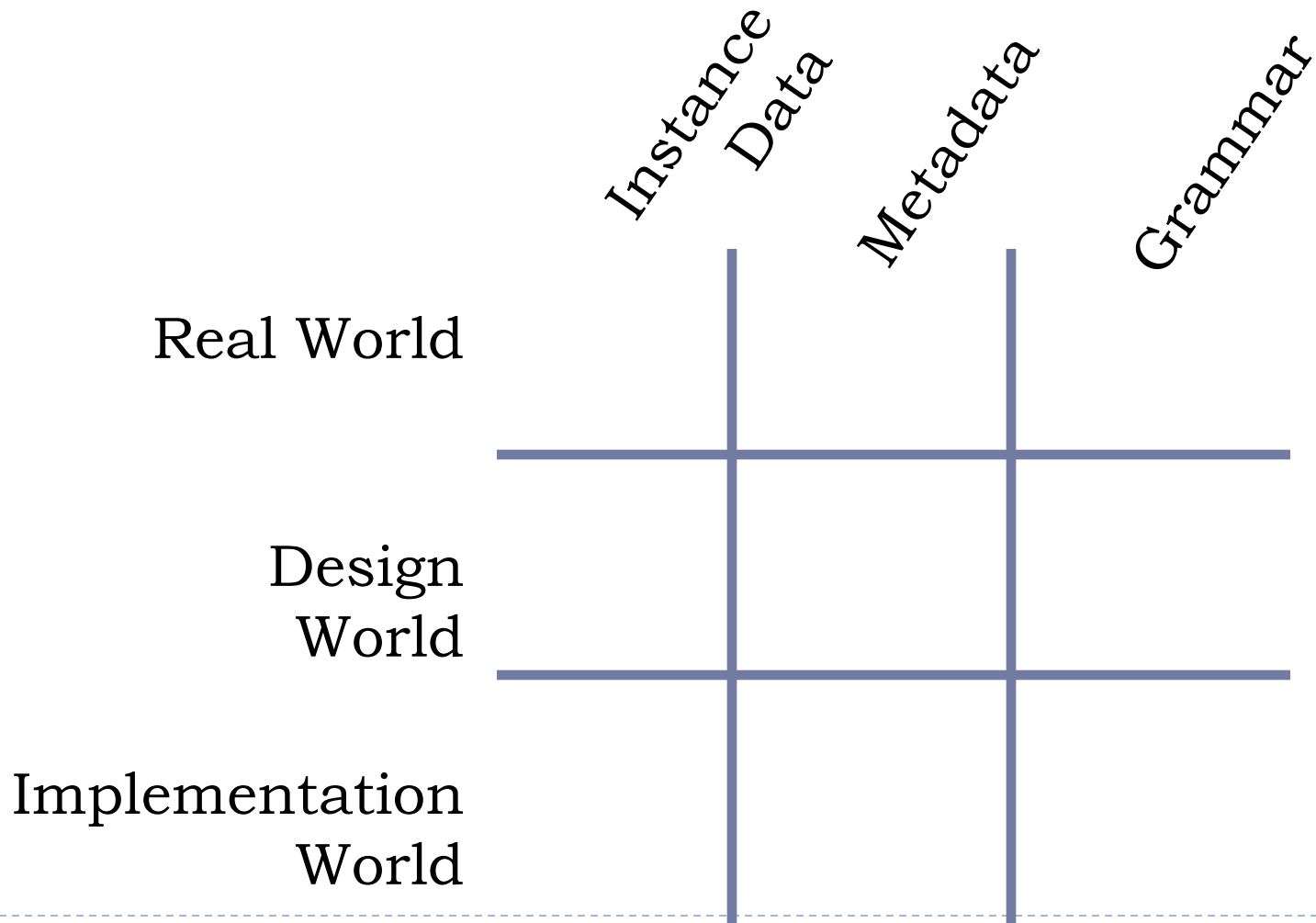
Content Complexity



Content Systems Complexity



Nine types of models (or schemas)



Results (still early)

- ▶ Big win will be “deep modeling” of their content (what a law or a court decision means, beyond how is it structured).

Summary

- ▶ Three different case studies of portions of Enterprise Architectures being rebuilt based on Enterprise Ontologies
- ▶ Each was built from a common upper ontology (gist)