

The logo features the word "FORRESTER" in a white, serif font, centered within a dark green oval. The oval is set against a dark blue background with a pattern of thin, light blue, wavy lines that create a sense of motion or depth.

FORRESTER®

# The State of Canonical Modeling in IT in 2010

**Mike Gilpin**

VP / Research Director

Forrester Research

March 9, 2010

# Agenda

- How did we get here? A brief history to set context.
- The current state of canonical modeling.
- The future of canonical modeling.



**Canonical modeling  
throughout history...**

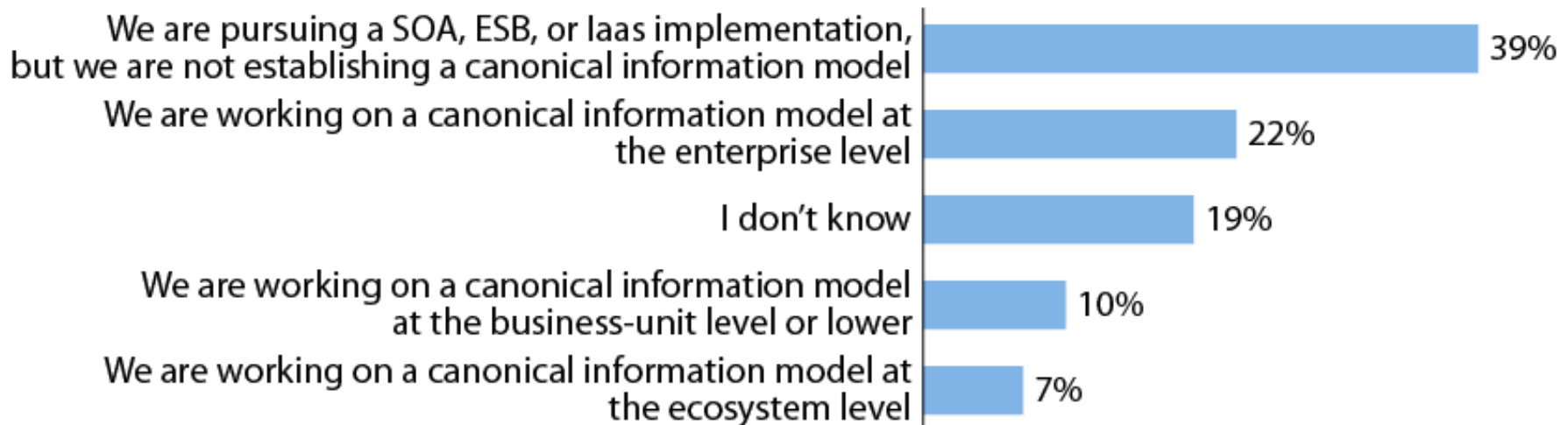
# Definition

- A canonical information model is a model of the semantics and structure of information that adheres to a set of rules agreed upon within a defined context for *communicating* among a set of applications or parties.

Canonical models are about ***data in motion***,  
not *data at rest*

# Canonical modeling has a long history

**“Is your organization working on establishing a canonical information model as part of your SOA, ESB, or Information-as-a-Service (IaaS) implementation?”**



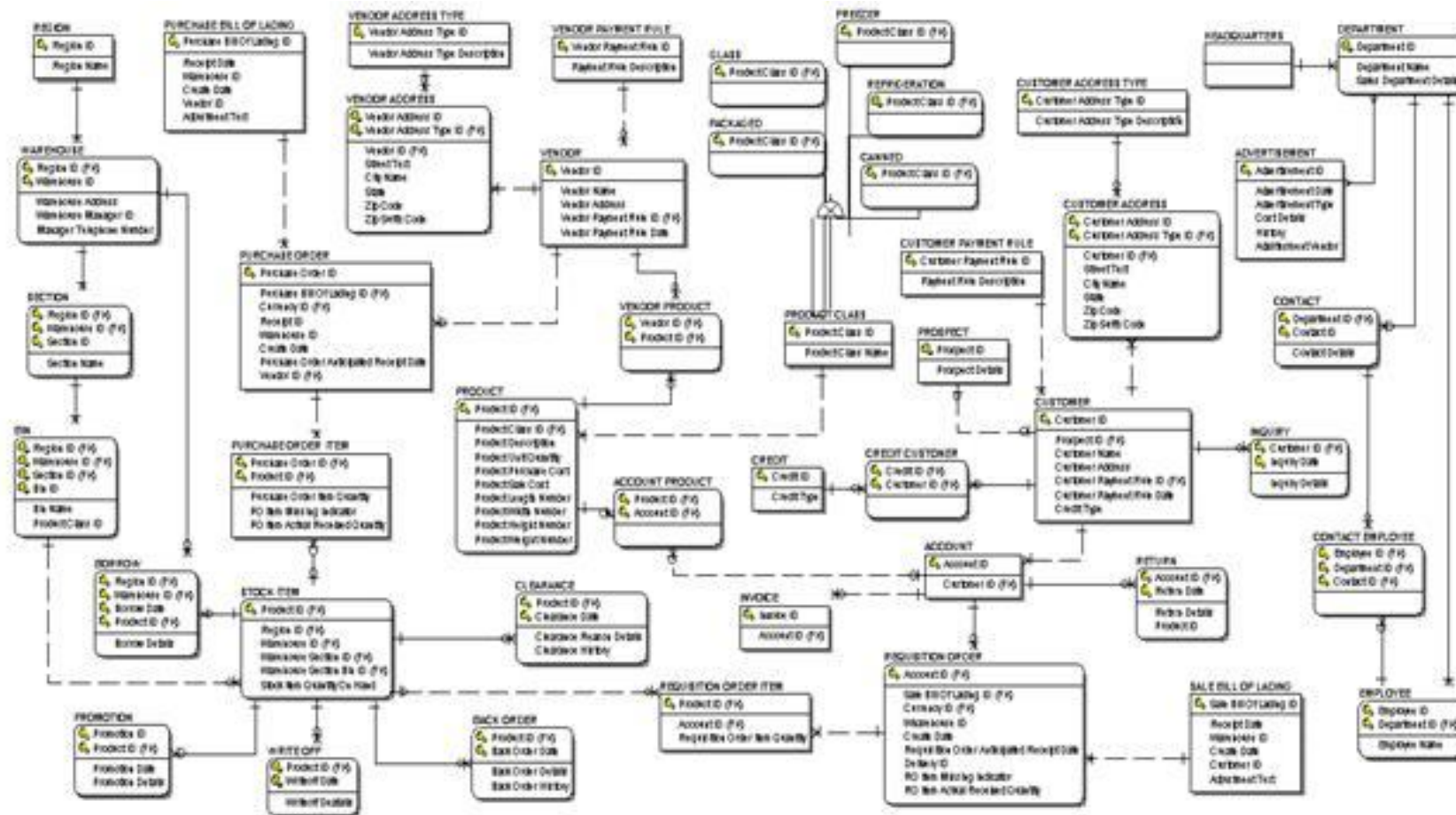
Base: 76 enterprise architecture professionals

Source: October 2007 Enterprise Architecture Role Online Survey

## Some canonical modeling (Forrester) history

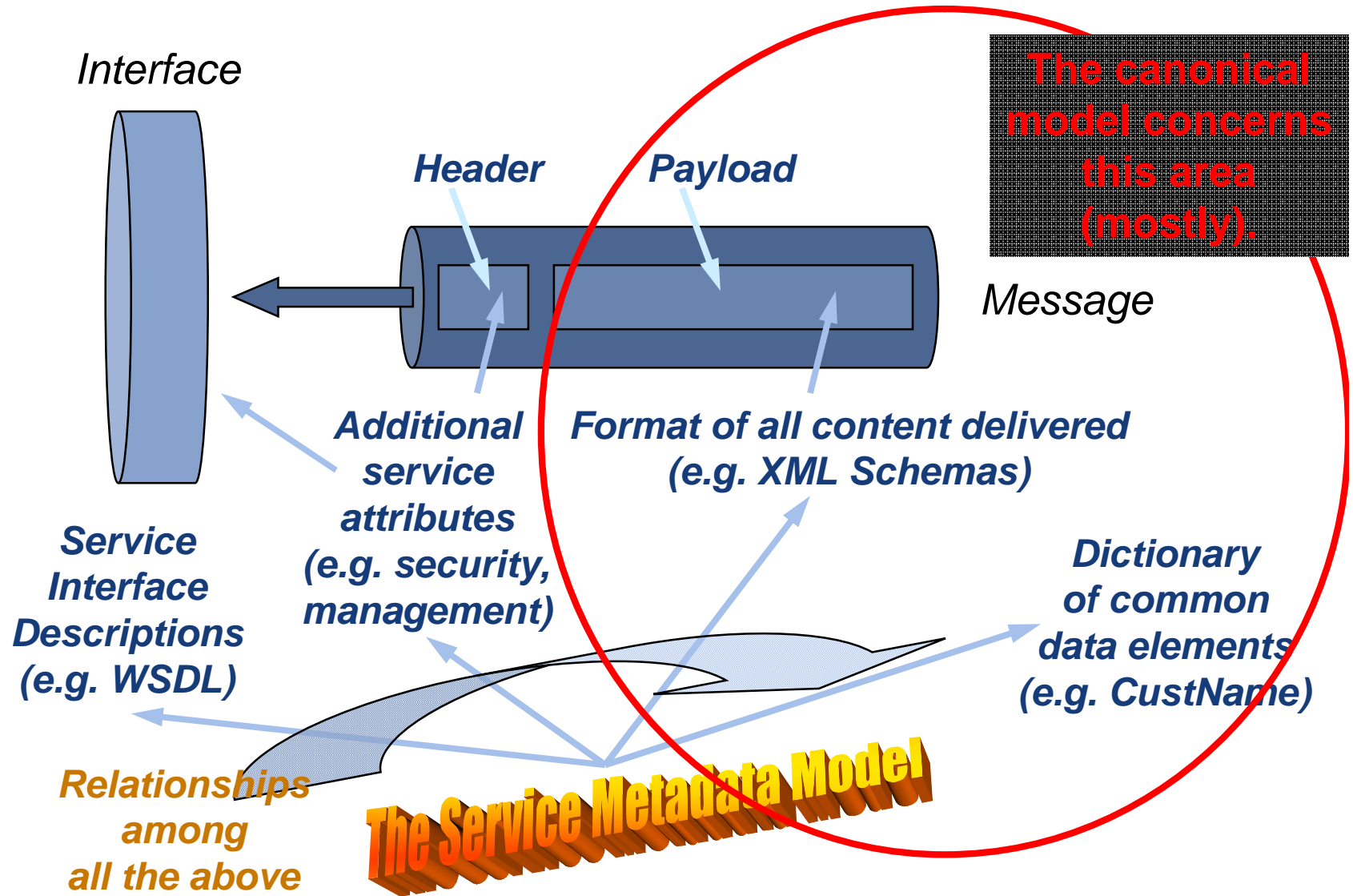
- UBS case study from late 90s – B2B trading flow
- Goldman question at SIA in 2000
- Large credit card operator project – 2006
- 2007 onwards **Data Services** have been an increasingly significant factor
- But “plain old” XML message interchange is still the main driver

# It's not the same as a “corporate data model”

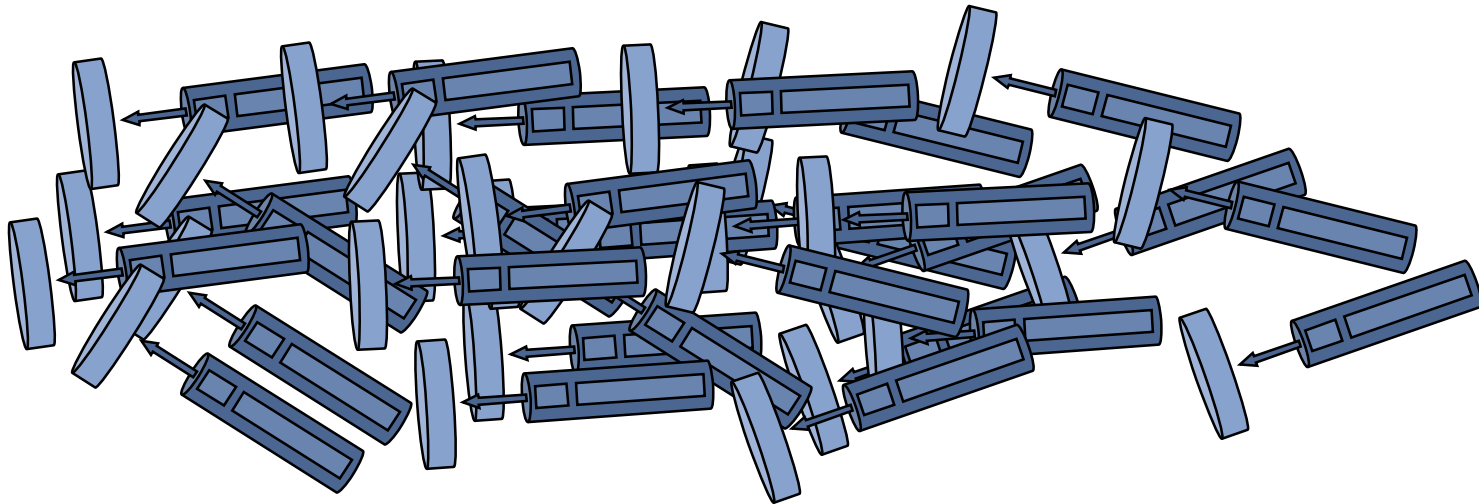


Shell Oil's corporate data model in the 80's had > 1000 entities!

# The metadata model of a service includes:



# When services proliferate, model chaos reigns

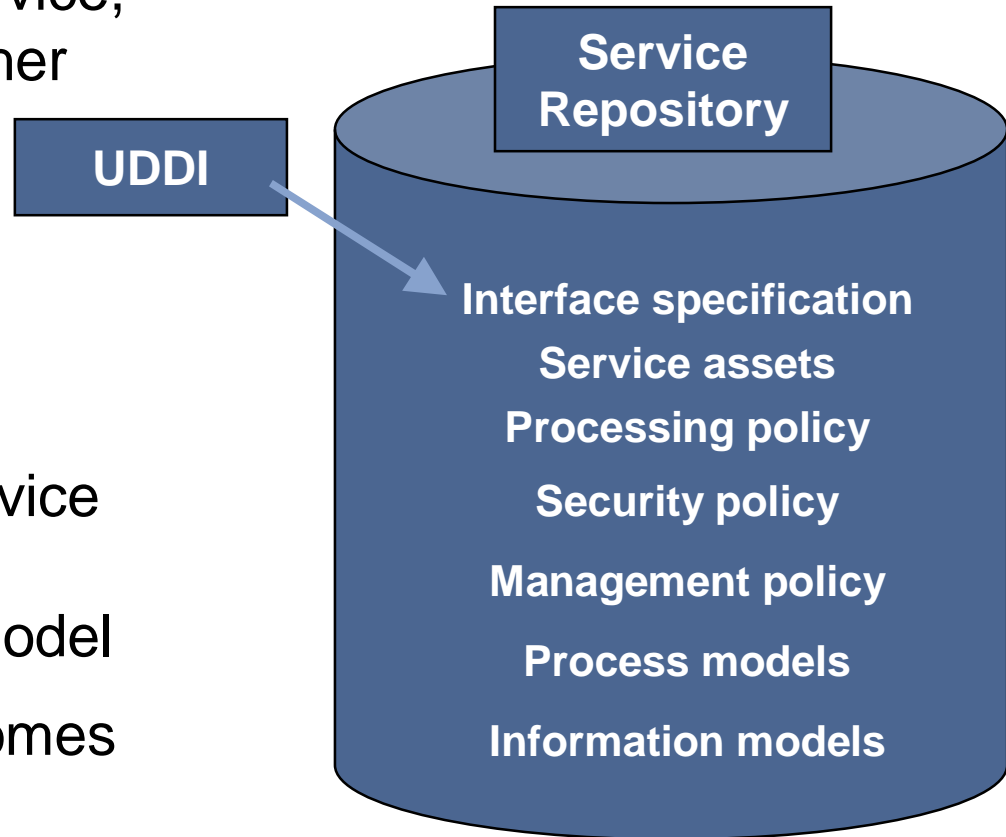



- It gets too hard to integrate new applications into processes
- Chaotic metadata also exacerbates data quality issues
- Bottom line: *A chaotic information model reduces the value of information*

*But organizations often lack the processes and technology to avoid information model and service metadata chaos*

# Your metadata needs a repository – perhaps *repositories*

- Manage all process, service, & event artifacts and other metadata as assets
- Extensible repository metamodel is essential
- UDDI is not enough
- The most advanced service repositories support a canonical information model
- A canonical model becomes the center for new development



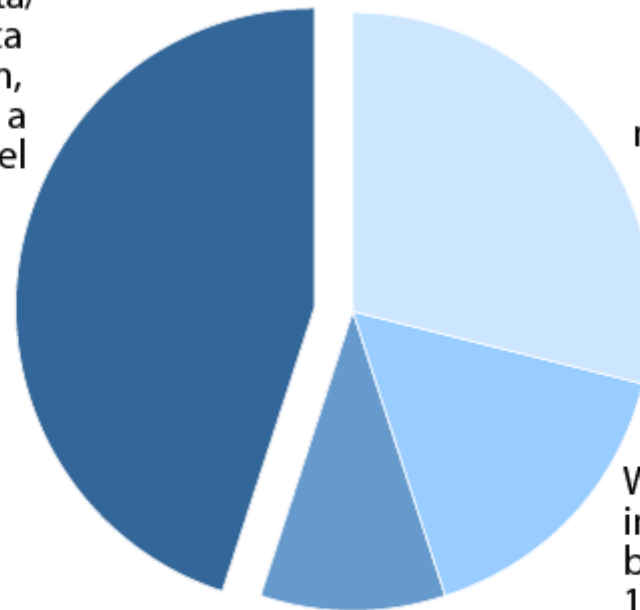


# Canonical modeling today

# Canonical modeling is increasing

**“Is your organization working on establishing a canonical information model as part of your SOA, data/information services, or data integration implementation?”**

We are pursuing an SOA, data/information services, or data integration implementation, but we are not establishing a canonical information model  
45%



We are working on a canonical information model at the enterprise level  
29%

We are working on a canonical information model at the business unit level or lower  
16%

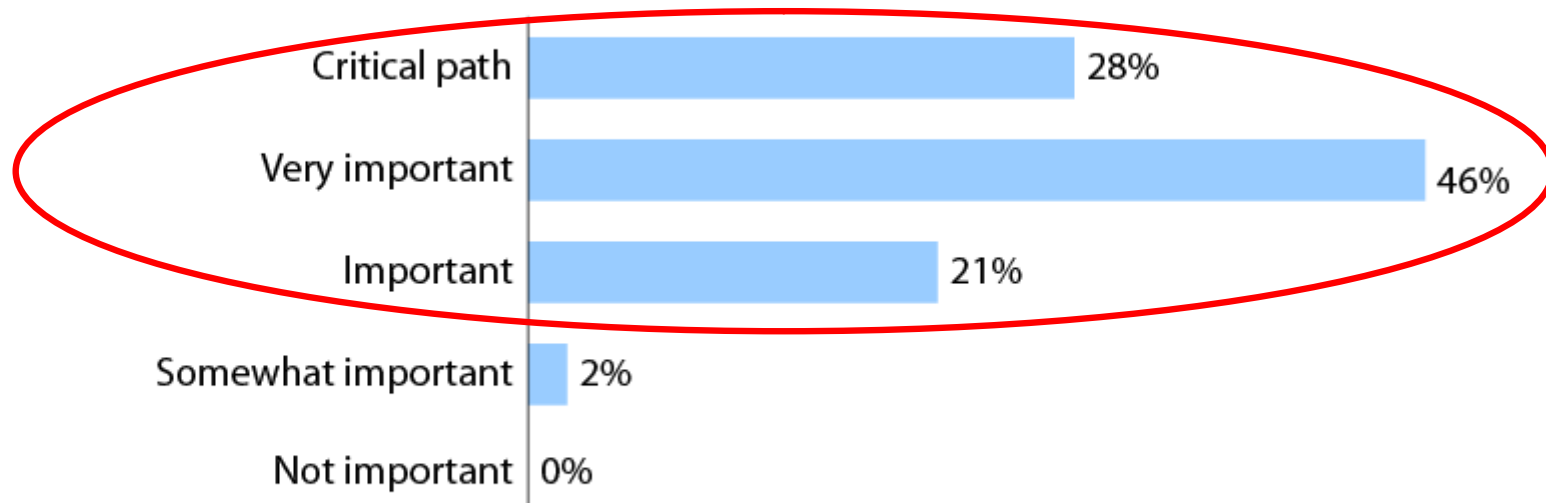
We are working on a canonical information model at the ecosystem level  
10%

Base: 107 EA professionals

Source: November 2009 Global Information Architecture Online Survey

# Data governance is crucial

**“How important do you/did you rate data governance in the success implementation and rollout of your information architecture?”**

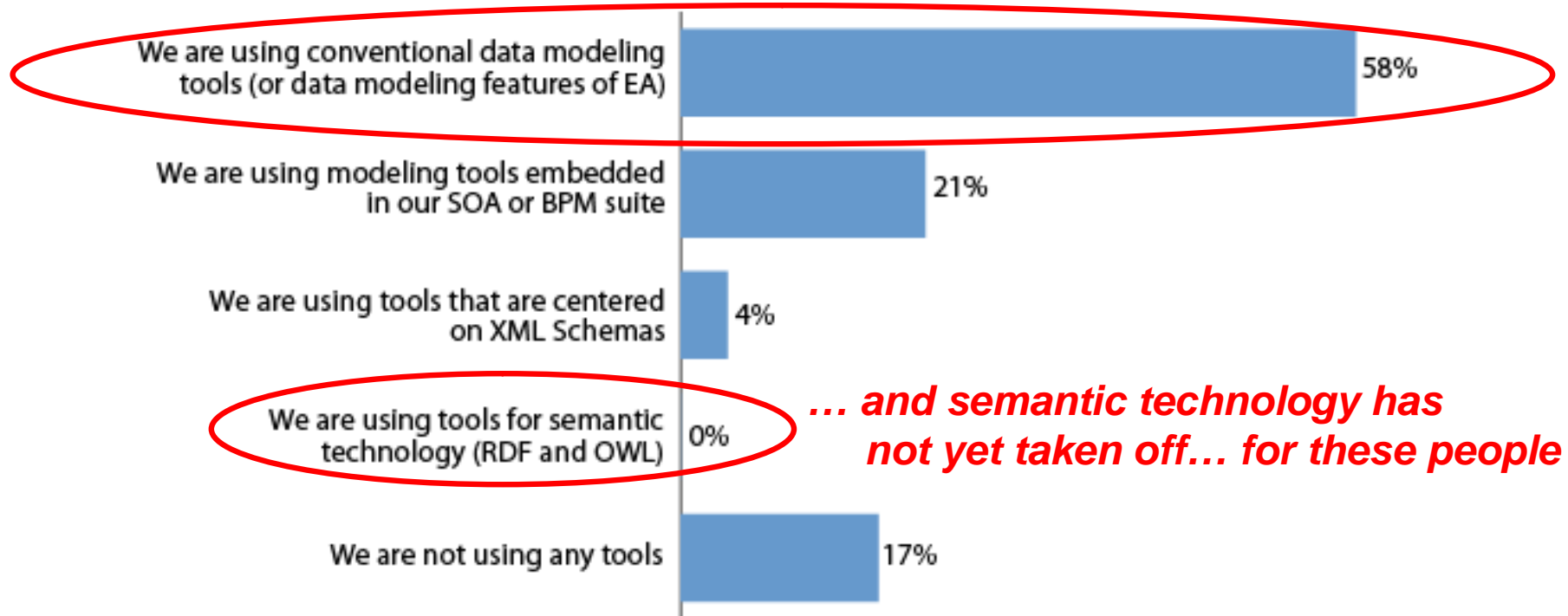


Base: 125 EA professionals  
(percentages do not total 100 because of rounding)

Source: November 2009 Global Information Architecture Online Survey

# Most architects seem to mainly be using older data modeling tools

"How are you developing or managing your canonical information model?"



*... and semantic technology has not yet taken off... for these people*

Base: 24 EA professionals

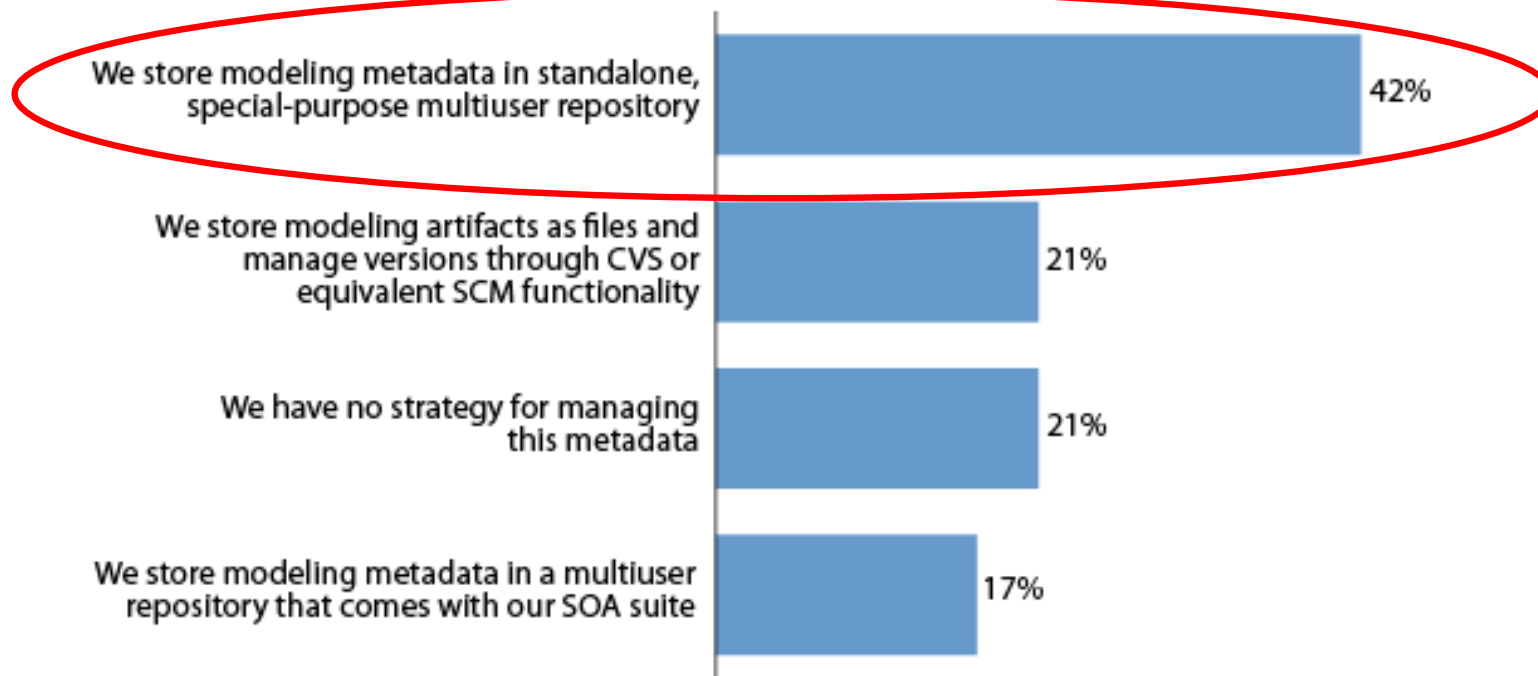
Source: November 2009 Global Information Architecture Survey

*but it's a small sample size...*

Source: Forrester Research, Inc.

# Most use standalone metadata repositories

**"How are you managing the metadata that arises from your canonical modeling efforts?"**



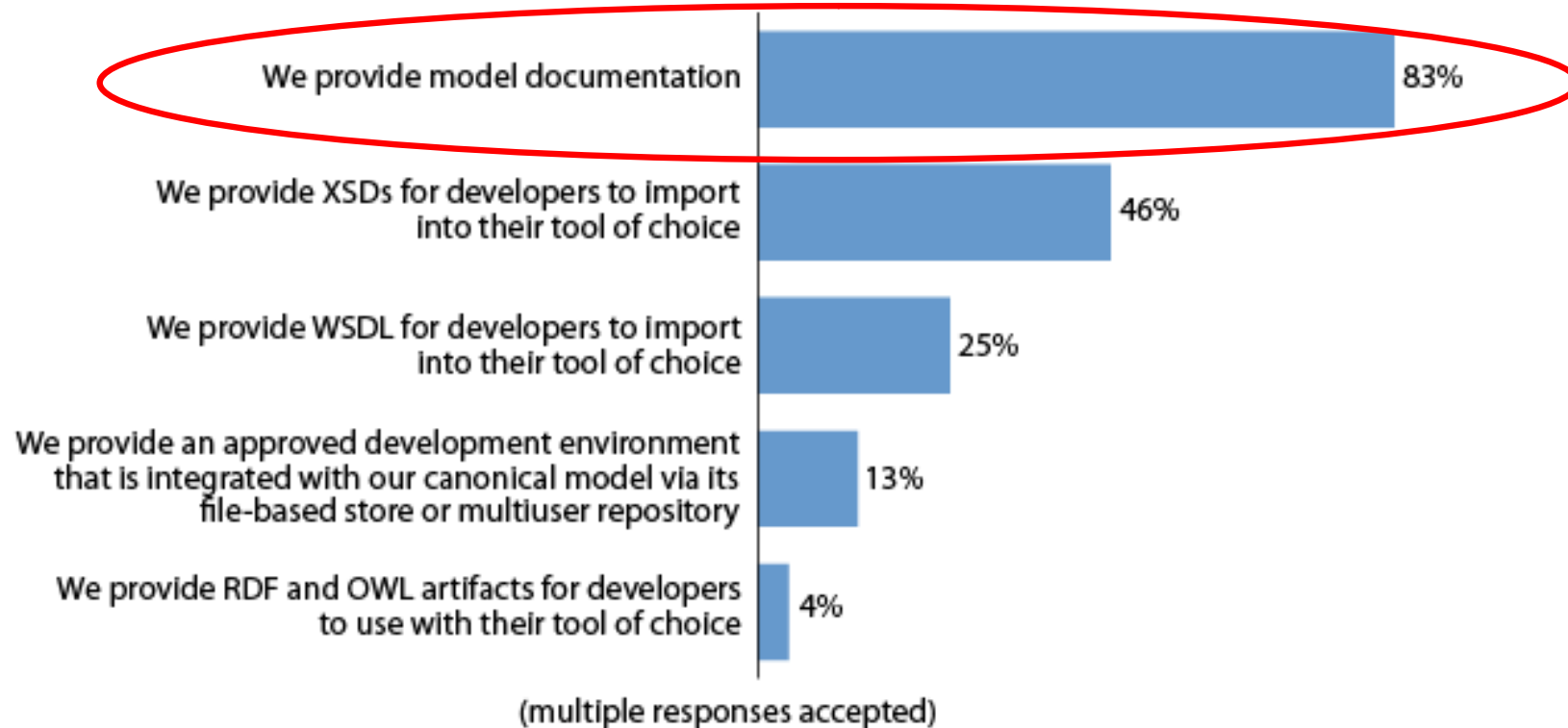
(percentages do not total 100 because of rounding)

Base: 24 EA professionals  
Source: November 2009 Global Information Architecture Survey

Source: Forrester Research, Inc.

# Support for developers is often limited

“How do you make your canonical model available to developers (multiple items can be selected)?”



Base: 24 EA professionals  
Source: November 2009 Global Information Architecture Survey

Source: Forrester Research, Inc.

# Lessons from current case studies

- Focus on a *limited* number of process domains
- Industry or other domain/ecosystem standards are crucial to success – NIEM, UCore, FIX, SWIFT, FpML, SEPA, ISO 20022, , Deriv/SERV, Markit Wire, Acord, ...
- Multiple usage scenarios:
  - Domain standards driven – often multiple standards
  - Partner driven – dominant partner sets standard
  - Application driven – dominant application sets standard
  - Bottom-up – legacy data content (including existing XML Schemas) dominates requirements

# More lessons from current case studies

- Tools are immensely helpful, especially when they automate model discovery from myriad artifacts (especially in the bottom-up scenario)
- Your model is worthless if developers won't use it – so it must be easy for developers to consume & extend, and must map into their familiar vocabulary
- Complex environments need multiple domain models, rich support for versioning and schema migration
- “Open,” *lightweight* tooling is best – allows substitution of tools, repositories, etc. – rather than a closed stack
- Today, semantic (RDF/OWL) and XML approaches are mixed only by more advanced organizations

# Canonical modeling in the future



# Use this information architecture framework

Policies governing ownership and access	Governance	<b>Conceptual Business information conceptual entities</b>								
		Structured data				Content				
		Conceptual data model — major entities, attributes, relationships				Taxonomy				
	Principles	Standards	<b>Logical Information entity mapping to applications and repositories</b>							
			Data flows and systems of record				Schema			
			Logical data model — major entities, attributes, relationships							
	Standards	<b>Physical Physical data stores and repositories</b>								
		Operational and analytical data				Content				
		Customer	Product	Financial	Sales	Email	Docs	Images	Web	
		Storage								

# What will happen?

- Canonical models will become the model for access through shared data services “access layers”
- Real-time master data through data services will become the foremost approach to supply business processes with business insight
- Combining structured/unstructured data will become a more mainstream, everyday pursuit
- Integration via the Cloud will become commonplace, driving developers toward data services as the foremost way to access information – although this will necessitate much richer support for developers
- Federated repositories will become the rule rather than the exception – with some off-the-shelf integration
- Data integration leaders will support common models and tooling across data services, data interchange, ESBs, B2B, BI/warehousing, and messaging

# Thank you

Mike Gilpin

+1 301.469.4863

[mgilpin@forrester.com](mailto:mgilpin@forrester.com)

[www.forrester.com](http://www.forrester.com)

Twitter ID: mgilpin

# Selected Forrester research

- November 15, 2007, “Canonical Information Modeling Is Key To Many Information-As-A-Service And SOA Strategies”
- April 24, 2008, “How To Make MDM And SOA Better Together”
- January 13, 2010, “Topic Overview: Information Architecture”
- October 2, 2009, “What Semantic Technology Means To Application Development Professionals”