network inference



White House Conference

2004 SEMANTIC TECHNOLOGY SERIES

The Evolution of Integration Technology Towards an Enterprise Semantic Web

Updated: AUG-2004

© Network Inference 2004

AGENDA

- Grappling with system and data interoperability
  - Problem Recap: CODE == $$$$$
  - Solution Evolution: P2P, EAI, EII, WS, Semantic Grid
- Leveraging significant "new" technology
  - Background: What are data semantics anyway?
  - Standards: RDF & OWL
  - Technology: Mediation, Registry & Inference Engines
  - Architecture: Semantic Grid Services
- Do you need pain killers or antibiotics?
  - Case Study: Network-Centric Warfare STGP
  - Markets that need adaptive information
  - Why adopt now?
- Closing Comments

...more: Problem Statement

Same Old IT Problems, Same Old OPERATIONAL Barriers

Speed of Change Barrier:
The faster an organization wishes to change, the more expensive the IT impact of data interoperability.

The "H" Factor:
The single largest barrier to rapid IT change is humans themselves, humans must get out of the loop!

Today's Bottom Line:
More data = More code
More code = More $$$$

- Incompatible data meanings are the largest, most expensive, and time-consuming portion of IT visibility and IT interoperability projects:
  - Gartner...
  - Forrester...
  - NIST...
  - IDC...
  - CIO Mag...
  - GAO...
- The classic "n-squared" problem of interfaces is even more severe at the data layer:
  - Data-to-data interfaces outnumber "pipes"
  - Tightly-coupled is brittle, and requires code
- 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 Library's of Congress

Problem Statement

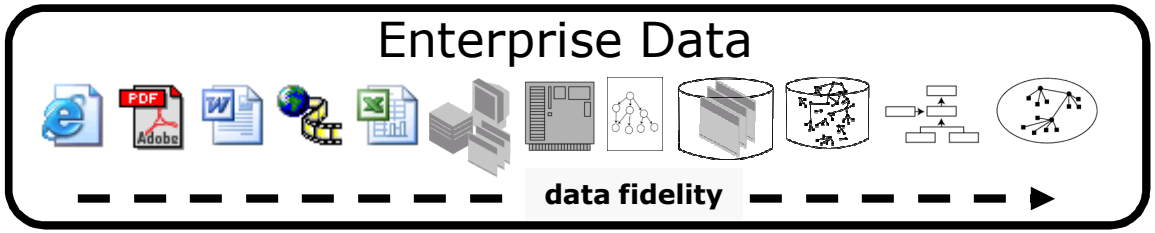
How Can ORDER be Wrought from CHAOS?

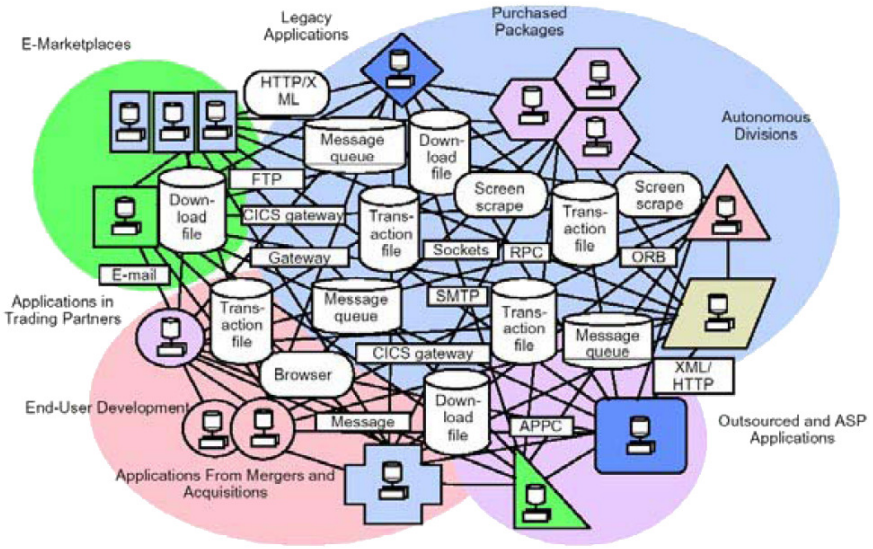
Enterprise data, created over decades, is chaotic, lacking meaning, and increasing at exponential rates.

Popular approaches to solving the problems of visibility and interoperability are incapable of coping with the complexity and scale of the root problems.

Leaving organizational decision-makers confused, and unable to see the forest for the trees.

Inference-driven systems solve these problems…

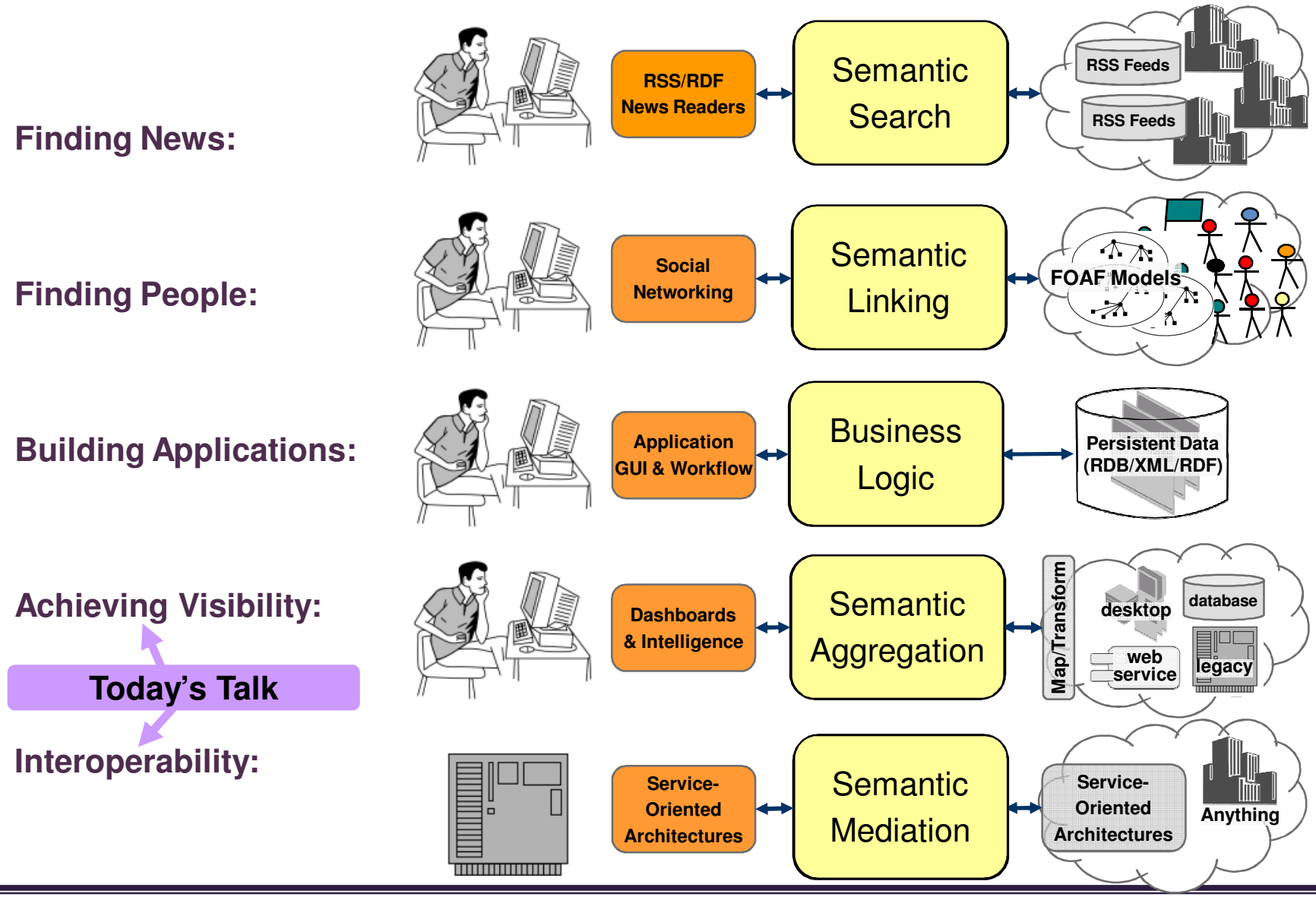






network inference

Semantic Technology Use Cases:
Some Early Glimpses of a Bright Future



network inference

De-Facto Solution

Ad Hoc, Point-to-Point Integration



Main Improvements:

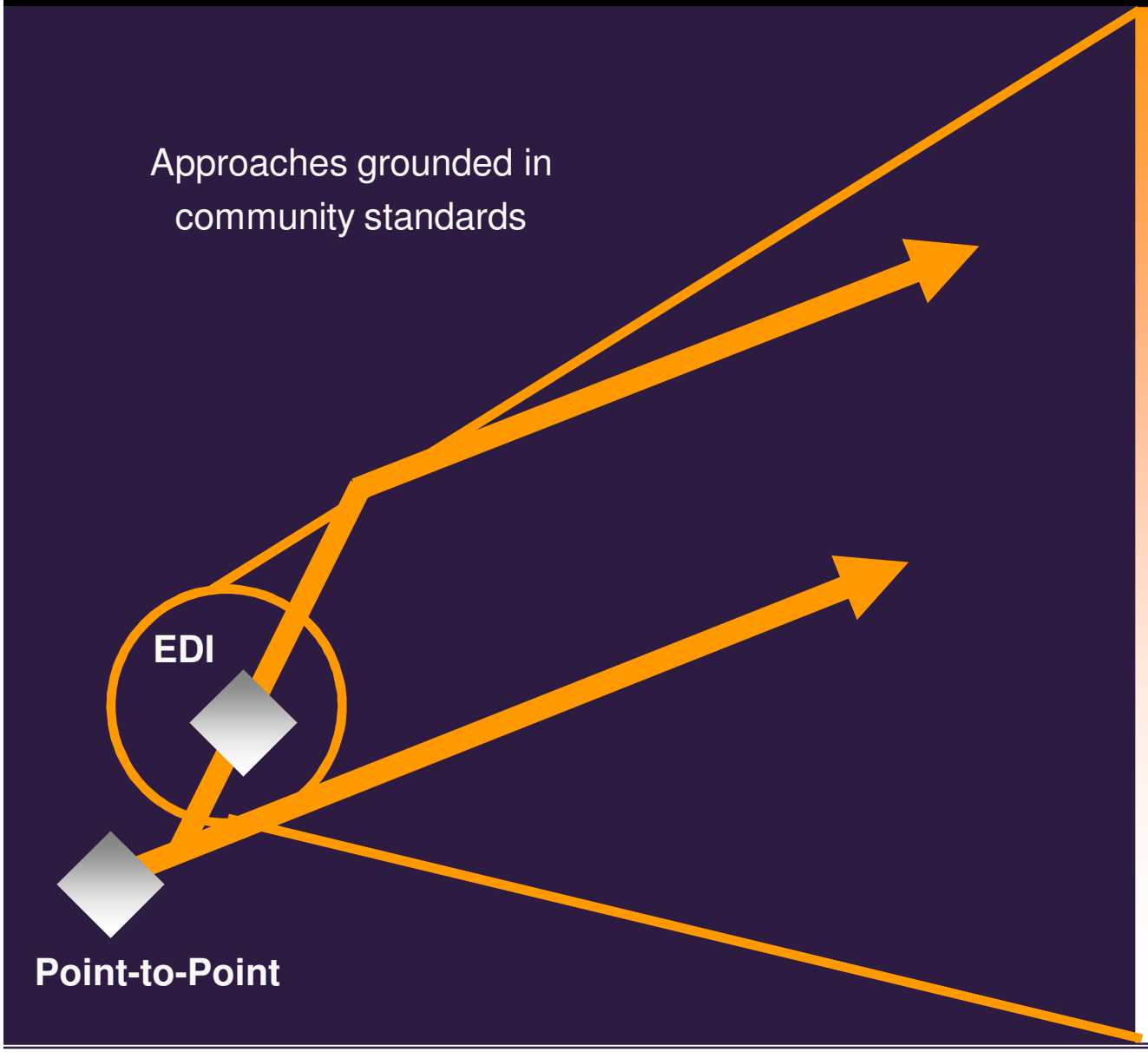
- Application connectivity is possible
- Data sharing need not be "swivel-chair"
- Some automation

Key Problems:

- No common protocols (at network)
- No common interfaces
- No tools for development
- Few tools for runtime workflow or process automation
- Too many interfaces to scale well
- **$$$$$**

Industry Standards and Information

Electronic Data Interchange



Main Improvements:

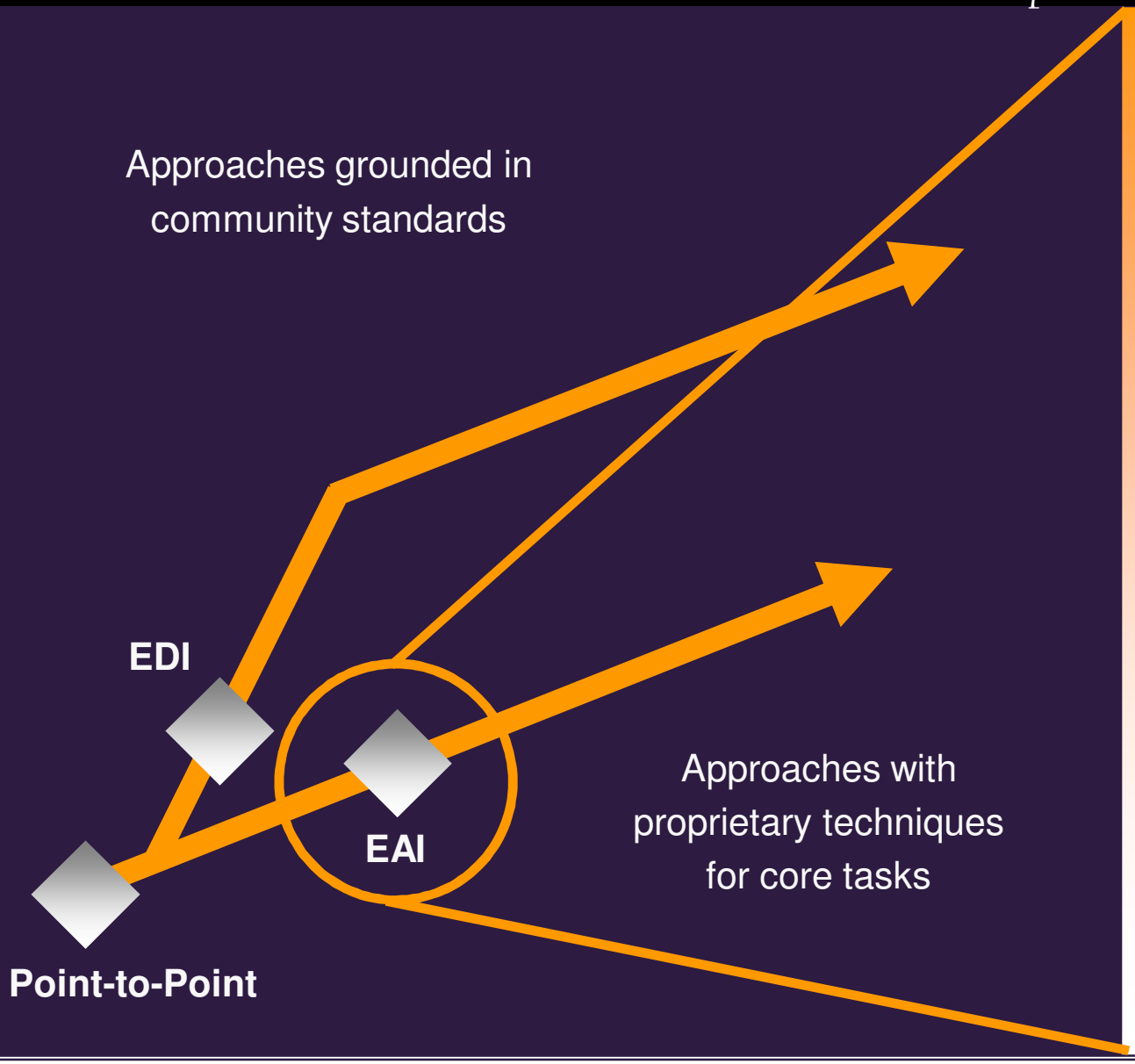
- Common vocabulary
- Standard protocols

Key Problems:

- Too much time to develop
- Too much time to change
- Fragmented vertical adoption
- Tight-coupling – inflexible
- No machine accessible semantics
- **$$$$$**

First COTS Solutions

Enterprise Application Integration (EAI)



Main Improvements:

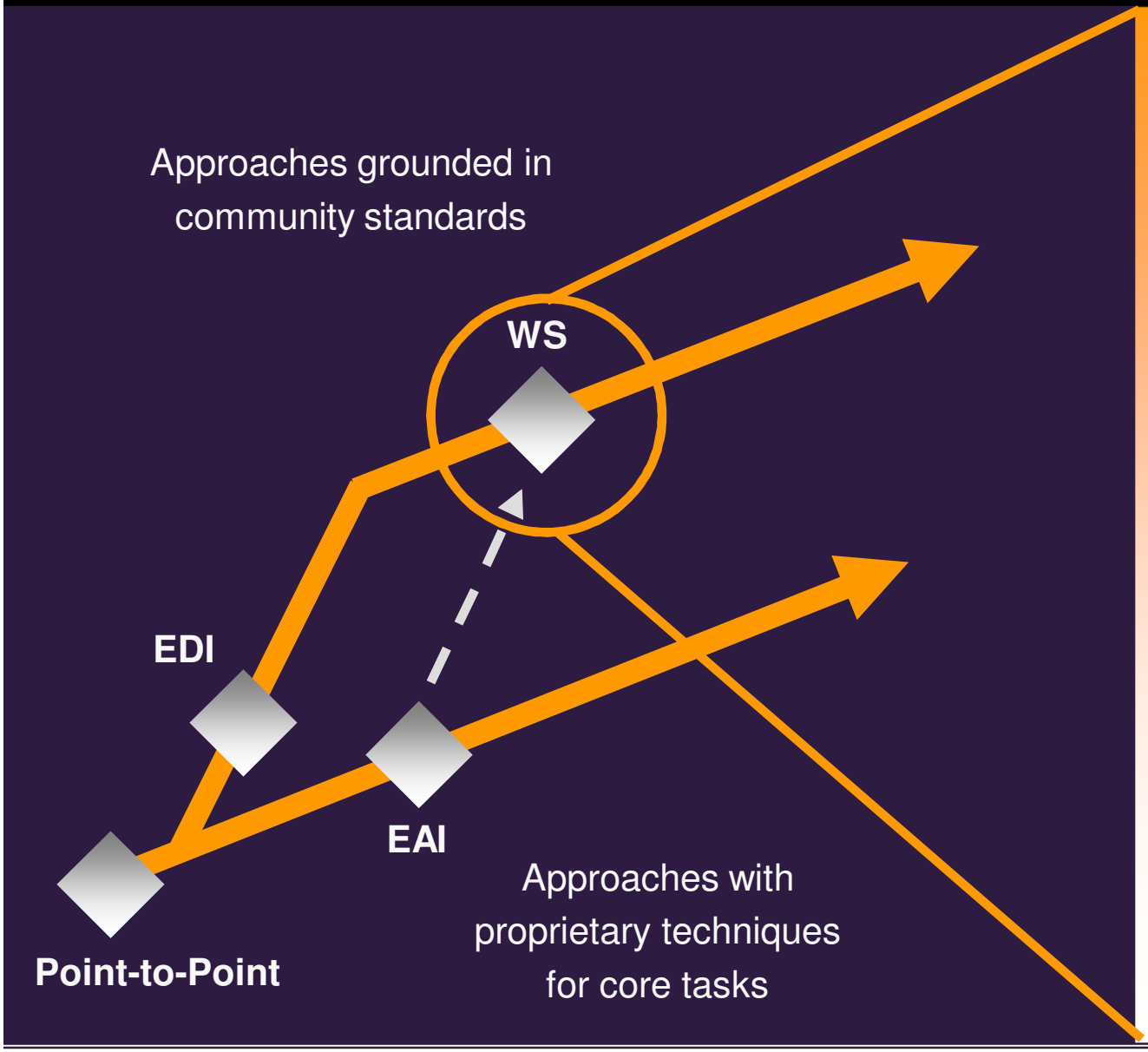
- Eliminate N^2 problem (for the plumbing)
- Extra "management" layers
- Support for different transaction models (pub/sub & req/rep)

Key Problems:

- Vendor lock-in
- Still no common, standard protocols (at network)
- Proprietary data/info management
- N^2 problem at data layer
- Weak business process
- **$$$$**

Meanwhile, the Industry Builds "EAI Standards"

Web Services



Main Improvements:

- No vendor lock-in
- Standard APIs/Services
- Loose-coupled "pipes"
- Service "paradigm" (vs. object or RPC)

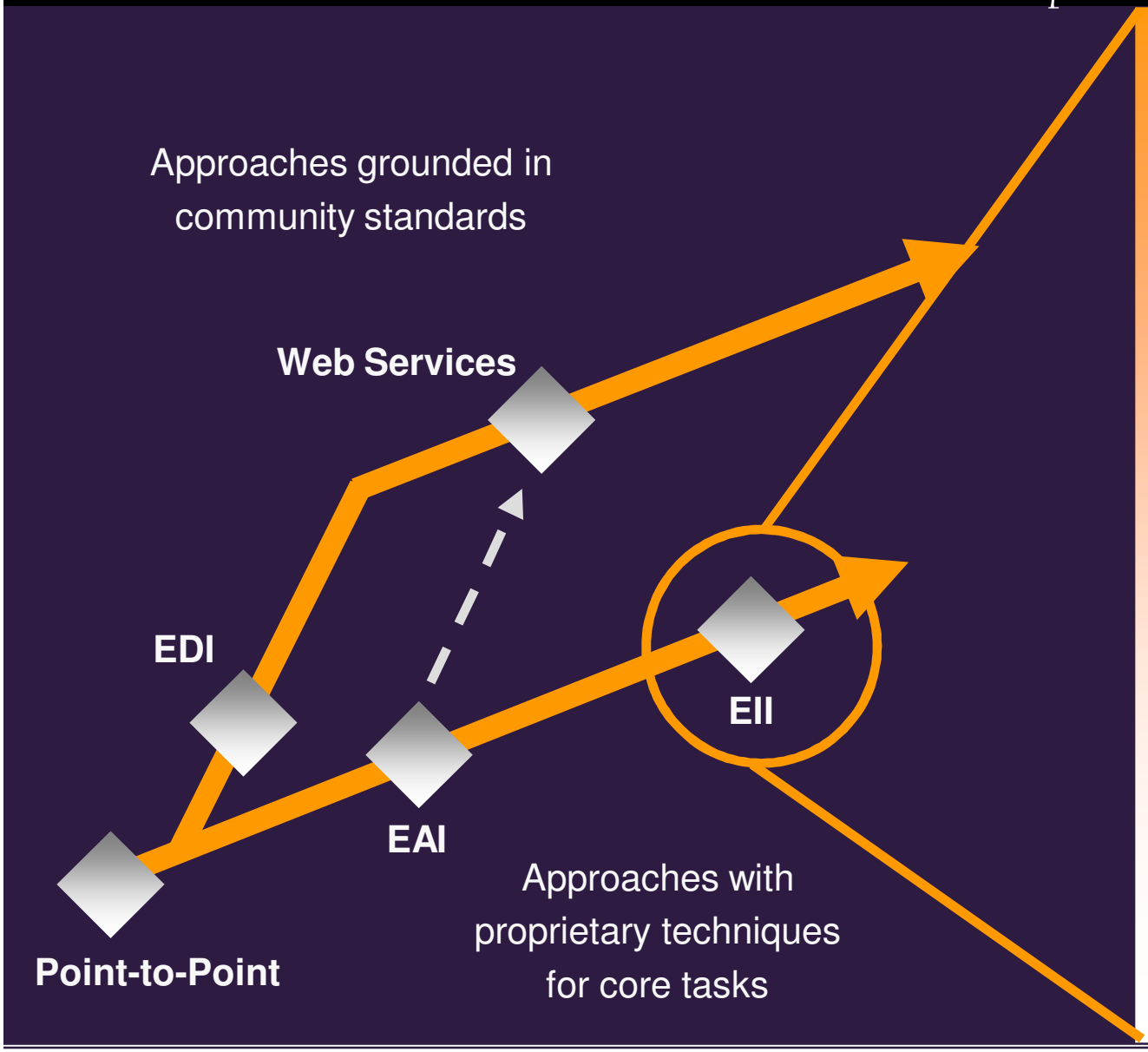
Key Problems:

- Difficult "management"
  - fewer tools
  - evolving vision
- N^2 problem at data layer
  - XSL/T & Java Xform
- Too much hype
  - not dynamic discovery
  - not "process aware"
- **$$$**

network inference

COTS Solutions (Finally) Focus on the Data

Enterprise Information Integration (EII)



Main Improvements:

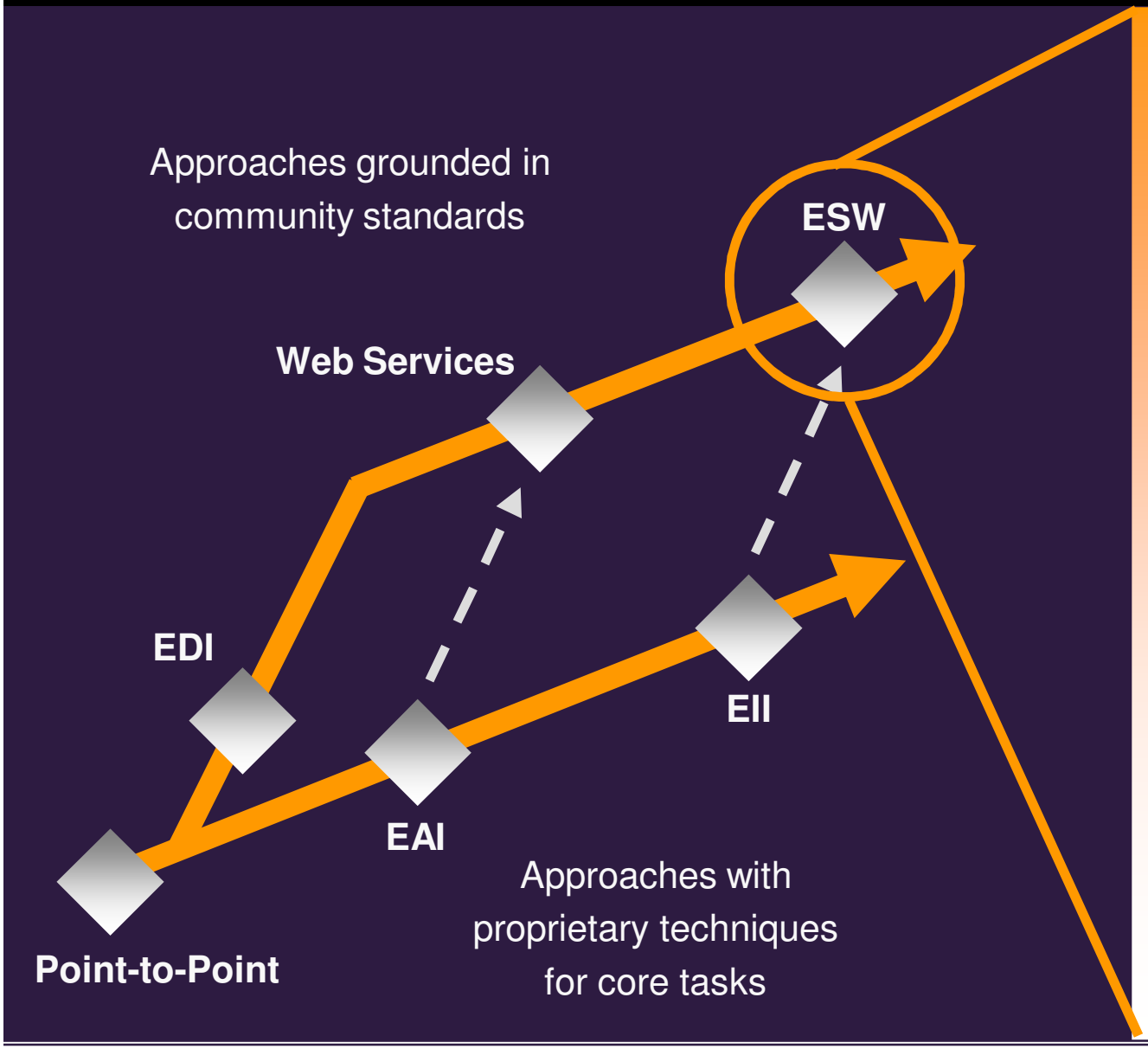
- Focus on core business value = information
- "Virtual" data views
- N^2 data problem solved
- Meta-data management
- Loose-coupled data

Key Problems:

- Vendor lock-in
- Proprietary metadata
- Fragmented approaches:
  - Relational
  - XML
  - Object
- Minimal automation
- **$$$**

Next, the Industry Standardizes Semantic Metadata

Enterprise Semantic Web (ESW)



Main Improvements:

- Standardized metadata
- New "smart" capabilities
- Much lower TCO
- Focus on core business value = information
- "Virtual" data views
- N^2 data problem solved
- Meta-data management
- Loose-coupled data
- Loose-coupled "pipes"
- Standard APIs/Services
- SOA Architecture

Key Problems:

- Emerging vendor support
- Use cases still evolving
- Up-front modeling costs
- **$$**

So, What do YOU Need to Know?

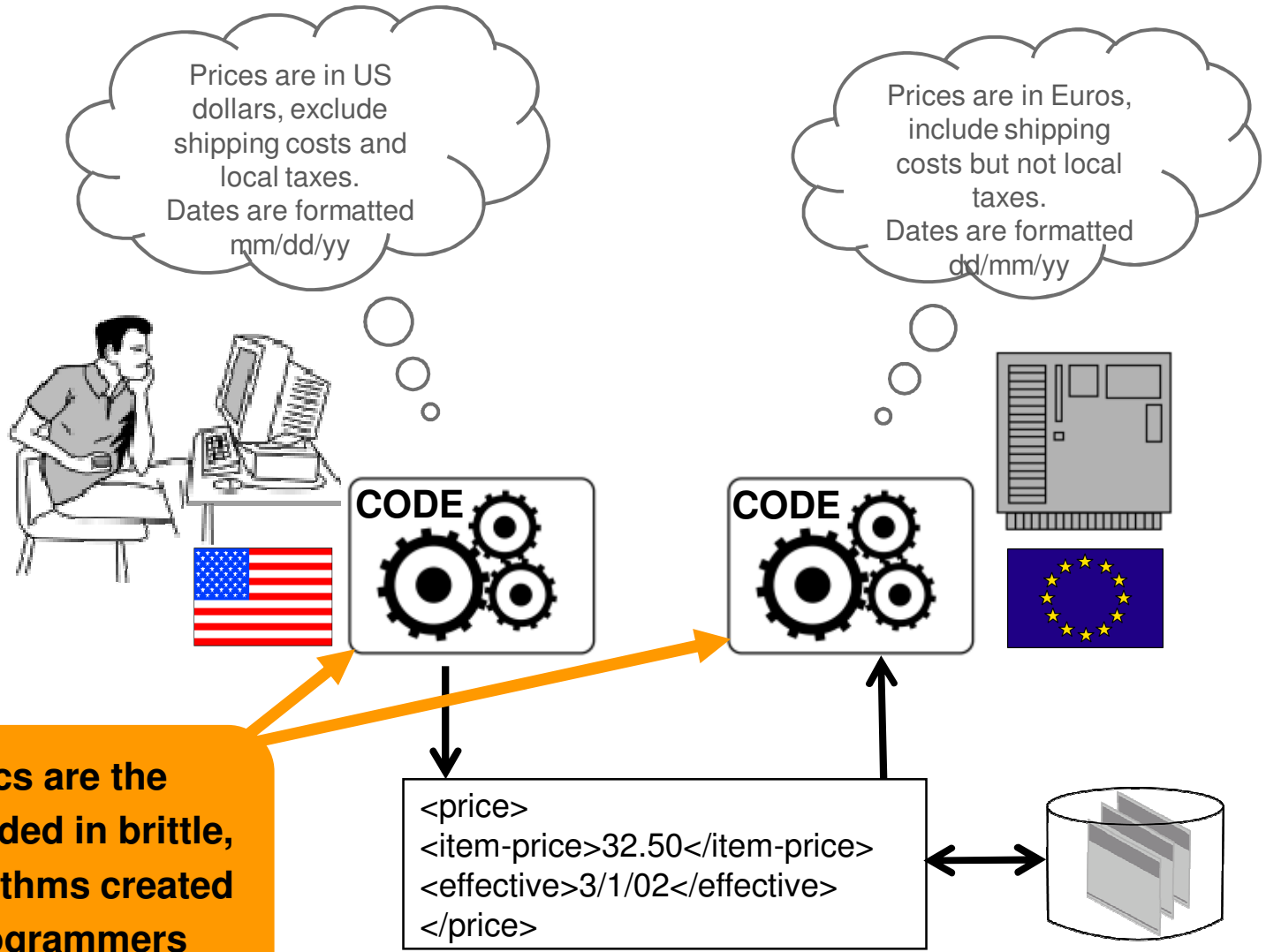
The Enterprise Semantic Web Has Arrived

"…(XML) is only the first step to ensuring that computers can communicate freely. XML is an alphabet for computers, and, as everyone who travels in Europe knows, knowing the alphabet doesn't mean you can speak Italian or French."

BusinessWeek
May, 2002

- A small bit of background…
  - Data semantics is how data is used
  - But today's popular data formats don't encode meaning – XML, OO, and relational data representations persist only simple structures.
- That newer, better data standards exist…
  - RDF – Resource Description Framework
  - OWL – Web Ontology Language
- Tools are available to help automate…
  - Inference Engines – deduce implicit knowledge
  - Mediation Engines – agree to disagree
  - Semantic Content Registry – DNS for semantics
- Architectures can evolve in realtime…
  - Semantic Grid Service Architecture

What are Data Semantics?



network inference

What are Semantic Conflicts?

Data Type	Different primitive or abstract types for same information
Labeling	Synonyms/antonyms have different text labels
Aggregation **Structure** **Cardinality**	Different conceptions about the relationships among concepts in similar data sets. Collections or constraints have been modeled differently for same information
Generalization	Different abstractions are used to model same domain
Value Representation	Different choices are made about what concepts are made explicit
Impedance Mismatch	Fundamentally different data representations are used
Naming	Synonyms/antonyms exist in same/similar concept instance values
Scaling and Unit	Different units of measures with incompatible scales
Confounding	Similar concepts with different definitions
Domain	Fundamental incompatibilities in underlying domains
Integrity	Disparity among the integrity constraints

What is Implicit Data, or "an Inference?"

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…

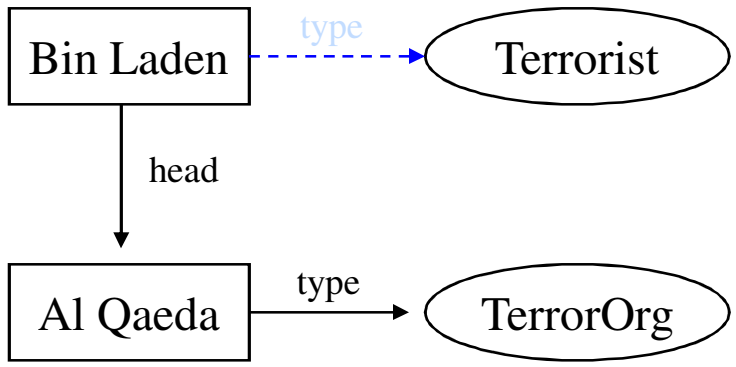
```
<owl:Property rdf:ID="head">
  <rdfs:subPropertyOf
        rdfs:resource="member" />
</owl:Property>

<owl:Class rdf:ID="Terrorist">
 <owl:sameClassAs>
   <owl:Restriction>
     <owl:onProperty
           rdf:resource="member" />
     <owl:someValuesFrom
           rdf:resource="TerroristOrg" />
   </owl:Restriction>
  </owl:sameClassAs>
</owl:Class>
```

The head of an organization is also a member of it

A member of a terror organization is a terrorist

Therefore, the head of a terror organization is a terrorist



network inference

Data Semantics Call to Action!

SEMANTICS WANT TO BE MODEL-DRIVEN, NOT CODED

Custom written software programs capture semantics in a tightly-coupled way, software models are a loosely-coupled alternative.

SEMANTICS ARE LOCAL, NOT GLOBAL

The meaning of anything, especially data, must be taken in context – and context is always local to a specific user or community.

SEMANTICS SCALE BEST WHEN FEDERATED

Centralized meanings limit adaptability and stifle adoption of new innovations, a flexible framework of meanings with multiple small centers is better.

SEMANTICS EVOLVE IN REALTIME

Static data models can't keep pace with business change; a software infrastructure that leverages dynamic and evolutionary data models is superior.

SEMANTICS ALWAYS OCCUR AT THE EDGES

Data meanings are most important when the edges of two "things" (human eyeballs to a monitor or one software layer to another) meet and exchange information.

SEMANTICS NEED NOT BE "STANDARDIZED"

Standard, shared vocabularies can artificially constrain the evolution of data communications because meaning will change over time and in new contexts.

So, What do YOU Need to Know?

The Enterprise Semantic Web Has Arrived

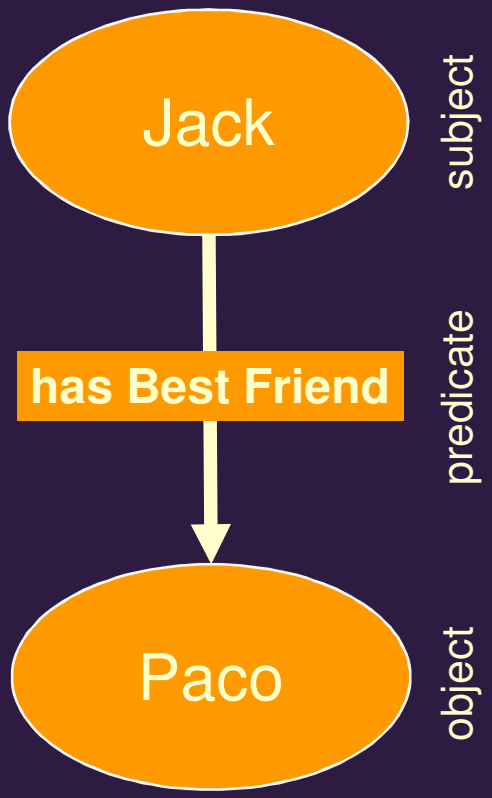
The standardization of OWL by the World Wide Web Consortium allows semantic Web technology to move out of the research and development community and into broad-based, commercial-grade platforms for building highly distributed, Web-enabled, cross-enterprise applications."

– DARPA, US DoD, February 2004

- A little bit of theory…
  - Data semantics is simply the meaning of data. But today's popular data formats don't encode meaning – XML, OO, and Relational data representations persist only simple structures.
- **That newer, better data standards are available today…**
  - **RDF – Resource Description Framework**
  - **OWL – Web Ontology Language**
- Tools are available to help automate…
  - Inference Engines – deduce implicit knowledge
  - Mediation Engines – agree to disagree
  - Semantic Content Registry – DNS for semantics
- Architectures can evolve in realtime…
  - Semantic Grid Service Architecture

What is RDF?

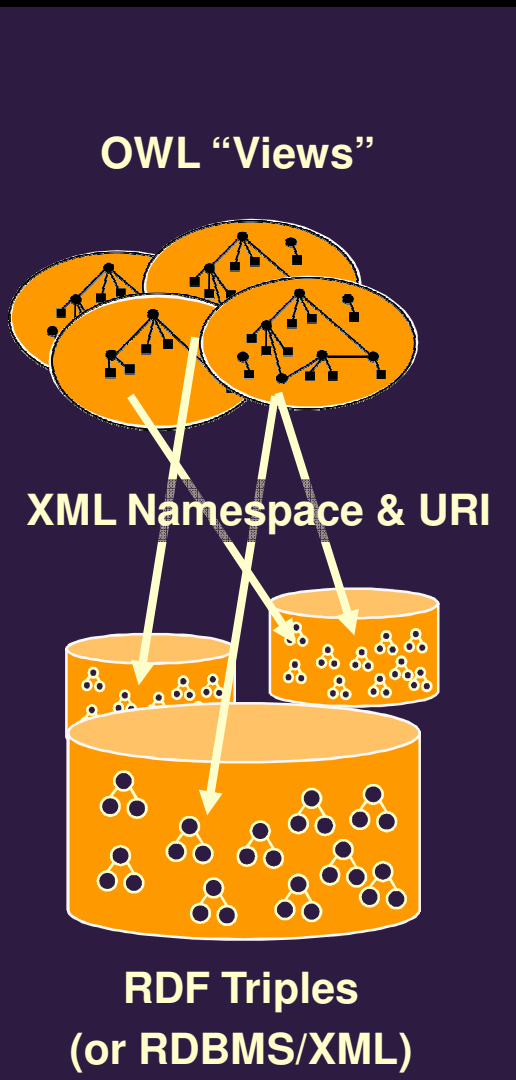
Resource Description Framework



RDF provides a great way of "leveling" an instance data set, while preserving the basic semantics required to reassemble the many parts in different "views" with OWL.

- Basic structure is a "triple"
  - [subject] => [predicate] => [object]
- RDF is implemented in XML, it inherits all syntax
  - Namespaces, for example
- Schemas need not be specified in advance
  - An RDF system need not require schema changes to cope with foreign vocabularies
- RDF data is fully expressible as RDBMS data
  - RDBMS data is also fully expressible as RDF

What is OWL?
Web Ontology Language



OWL is a born-and-bred ontology language for specifying domain knowledge, taxonomy, objects, classes, business rules and business logic in a model-based syntax.

- OWL provides loosely-coupled "views" of data
  - Federated knowledge-bases are easy to build and evolve
- OWL had a sound and complete semantic model
  - Like relational algebra, OWL(DL) is predictable
- OWL has machine-actionable semantics
  - Tools can do "things" with the models, data, metadata, rules, and logic without human assistance or code
- OWL is a highly expressive modeling language
  - Existing data (Relational, XML, OO) works with OWL

So, What do YOU Need to Know?

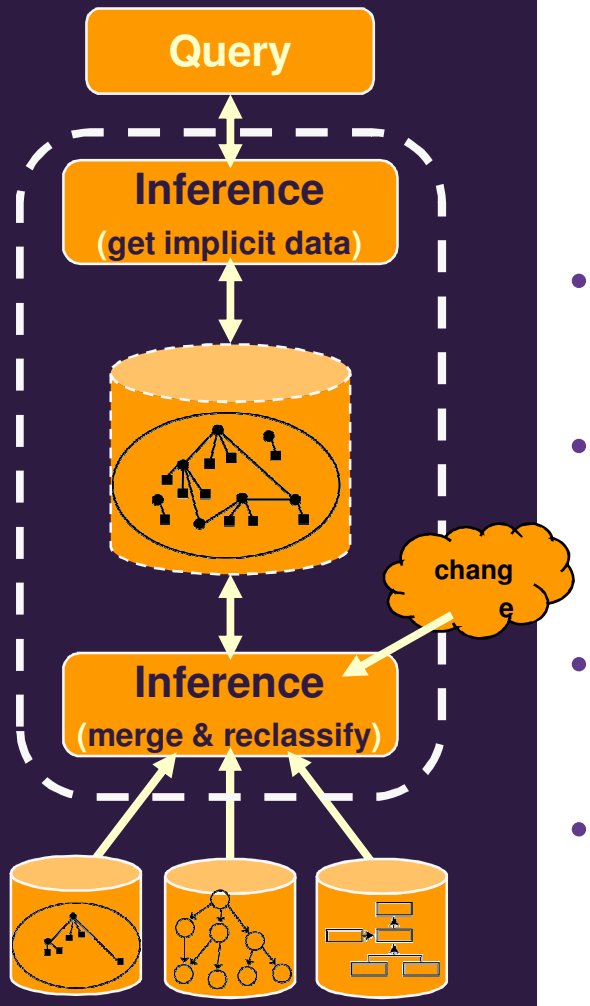
The Enterprise Semantic Web Has Arrived

"OWL has proven to be a crucial aspect of our strategy for adaptive information within our CareLink applications. Adopting OWL now [and inference-driven tools] means fewer manual changes to our data models in the future and greater flexibility in describing and extending knowledge and guidelines dynamically"

– Clinician Support Technologies, February 2004

- A little bit of theory…
  - Data semantics is simply the meaning of data. But today's popular data formats don't encode meaning – XML, OO, and Relational data representations persist only simple structures.
- That newer, better data standards are available today…
  - RDF – Resource Description Framework
  - OWL – Web Ontology Language
- **Tools are available to help automate data manipulation…**
  - Inference Engines – deduce implicit knowledge
  - Mediation Engines – agree to disagree
  - Semantic Content Registry – DNS for semantics
- Architectures can evolve in realtime…
  - Semantic Grid Service Architecture

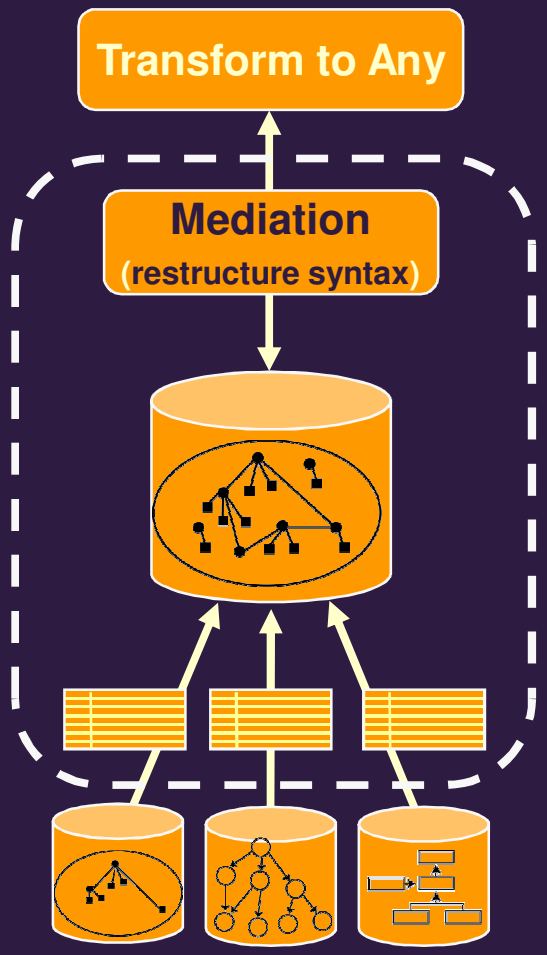
What is an Inference Engine?



Inference engines are services that can infer implied information from data, as well as adaptively reclassify knowledge, rules, and logic based on external conditions.

- Inference engines are services
  - They may be used in applications or middleware
- Inference engines are deductive/inductive "reasoners"
  - Native algorithms consume ontologies and can infer new facts or adaptively change how data is classified
- To an inference engine, data is just data
  - Data, metadata, rules, and logic are all equals
- Inference is most powerful on linked ontologies
  - Mapping and schema evolution is inherent, not coded
  - Federated data can "follow" concepts as they change and evolve along with schema/ontology changes

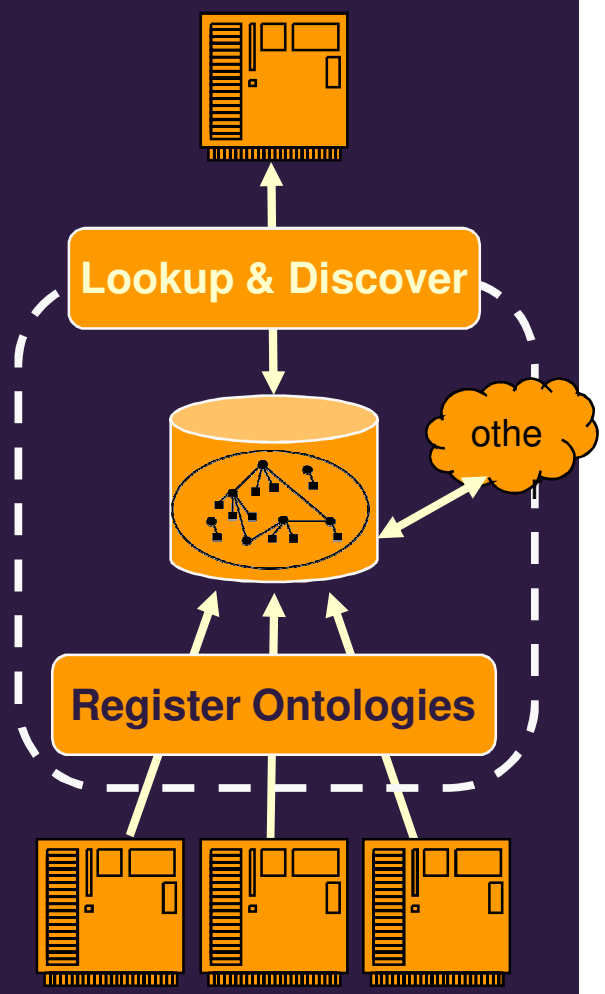
What is a Mediation Engine?



Mediation engines are tools that can dynamically output data in different formats and syntax that comply to disparate schema types. They enable automated data transformation without code.

- Mediation engines are services
  - They may be used in applications or middleware
- Mediation engines work with most structured data
  - Unstructured and semi-structured data must first be bound to a schema prior to creating the mediation maps
- Mediation engines let business agree to disagree
  - Use your own XML, relational or OO schemas
- Mediation engines enable interoperability of data
  - Automated transformation, any-to-any, is possible

What is a Semantic Content Registry?



Semantic content registries provide a federated registry for the semantics of schemas, ontologies, and applications – they will become the DNS for application semantics.

- Semantic content registries are services
  - They may be used in applications or middleware
- Semantic content registries are vocabulary managers
  - Registries are a neutral place to register and store data, metadata, and rules in the form of ontologies
- Semantic content registries use OWL and/or RDF
  - Better, standard, data and knowledge representation
- Semantic content registries enable discovery
  - With inference, discovery of new content and services will be as easy as issuing an ad-hoc query

So, What do YOU Need to Know?

The Enterprise Semantic Web Has Arrived

"Clearly, the time to forge a common framework based on semantic interoperability standards and e-business web services standards is now. "

– Patrick Gannon, CEO and President, OASIS "Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing & Enterprise Integration" – Book Forward

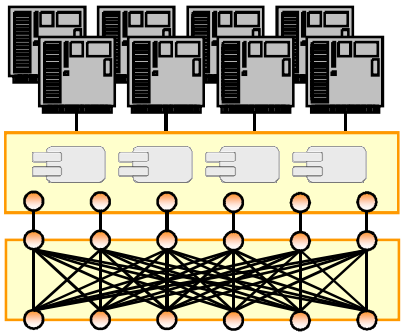
- A little bit of theory…
  - Data semantics is simply the meaning of data. But today's popular data formats don't encode meaning – XML, OO, and Relational data representations persist only simple structures.
- That newer, better data standards are available today…
  - RDF – Resource Description Framework
  - OWL – Web Ontology Language
- Tools are available to help automate data manipulation…
  - Inference Engines – deduce implicit knowledge
  - Mediation Engines – agree to disagree
  - Semantic Content Registry – DNS for semantics
- **Architectures can evolve in realtime…**
  - Semantic Grid Service Architecture

Evolution of Integration Architectures

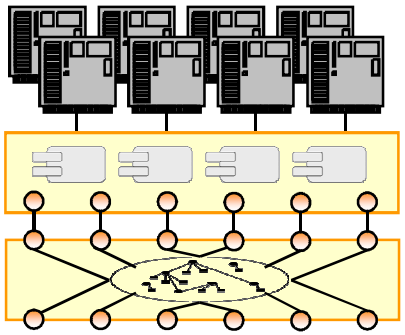
Steady Progression Towards Adaptive Capabilities

Fewer interfaces, looser-coupling and more adaptive configurations…

Web Services

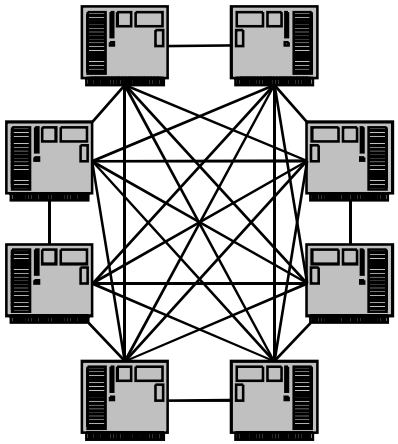


Semantic Grid

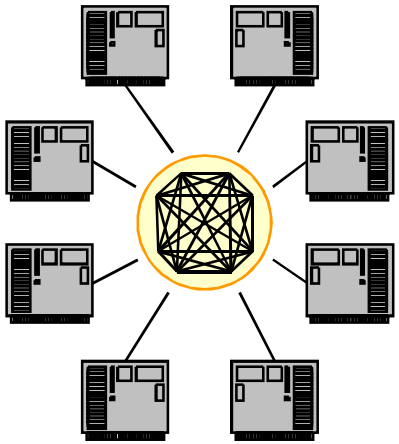


SOA

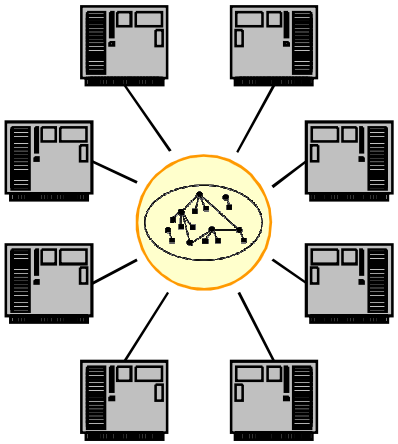
Point-to-Point => Hub and Spoke => Hub + EII



Point-to-Point



EAI



EII

network inference

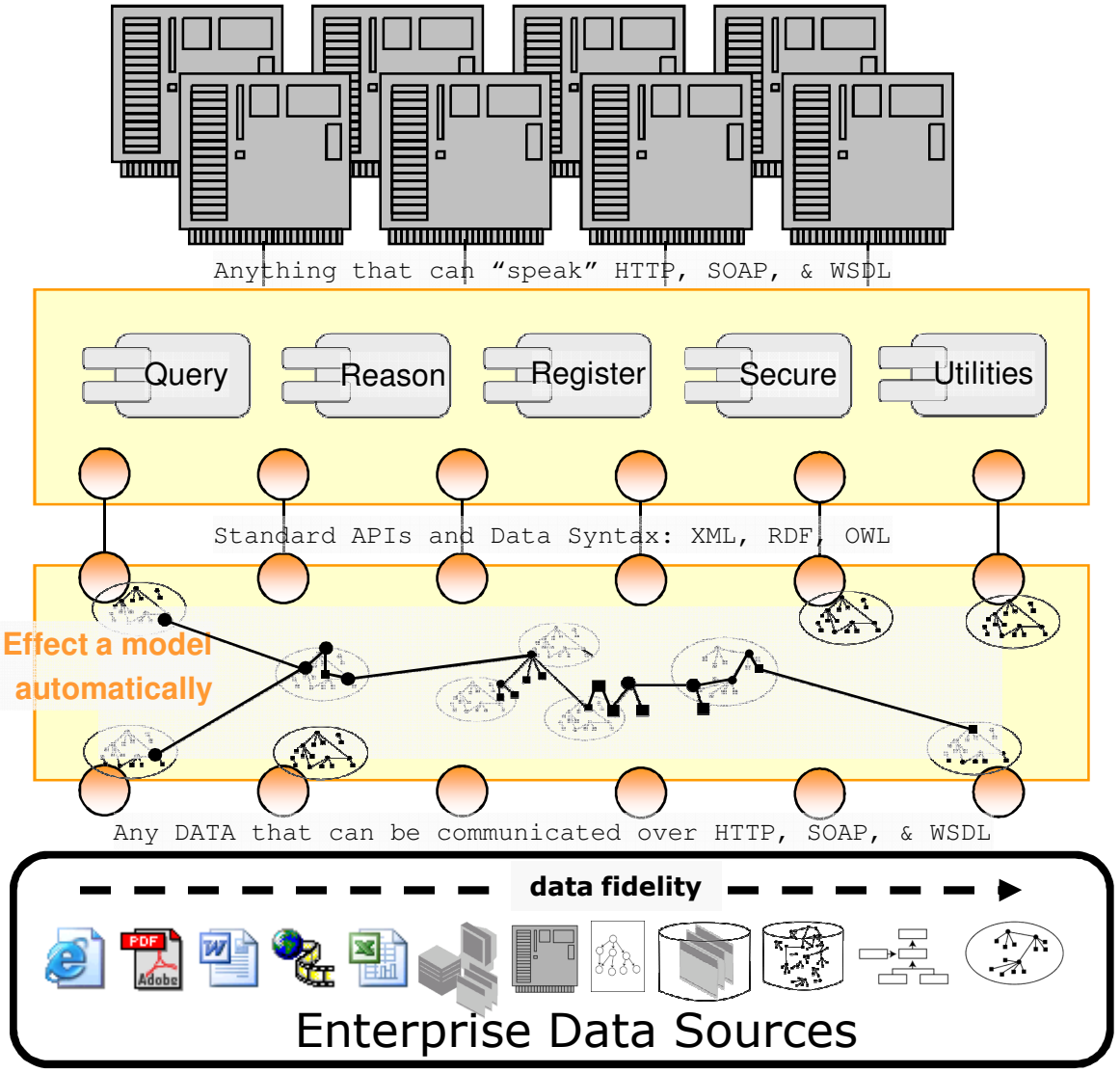
The Power of an Enterprise Semantic Web

Effect a Common Data Model (without nasty politics!)

Component driven Service-Oriented Architectures, using Web Services, are ripe for inference-driven systems.

Problems agreeing to vocabularies are political and cultural in nature.

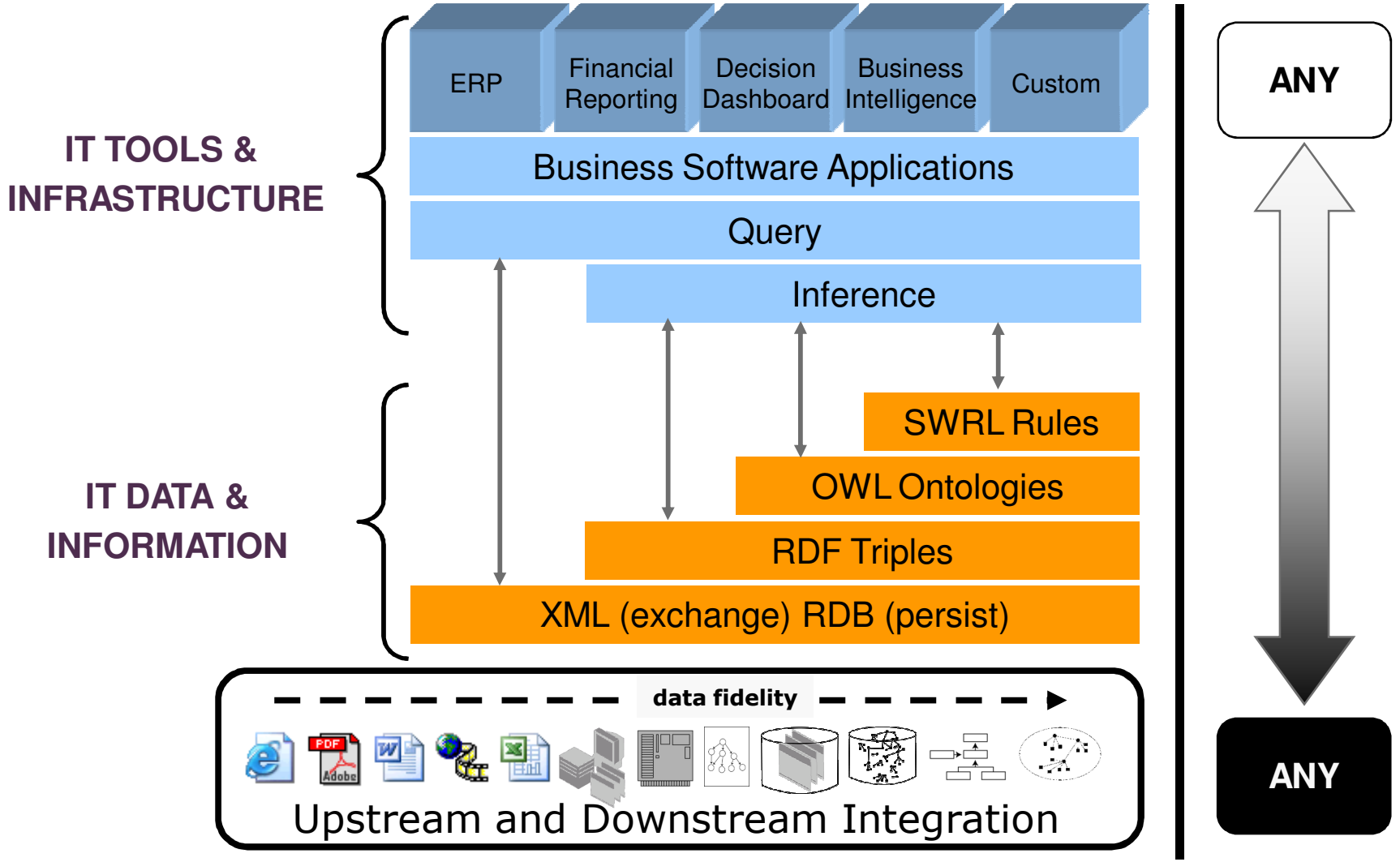
Reasoning systems (using RDF and OWL) provide a way around the problem by allowing each environment to have local view – while the overall system can have many common views.



network inference

Enterprise Architectures

Best Practice: Layer Your Architectures!

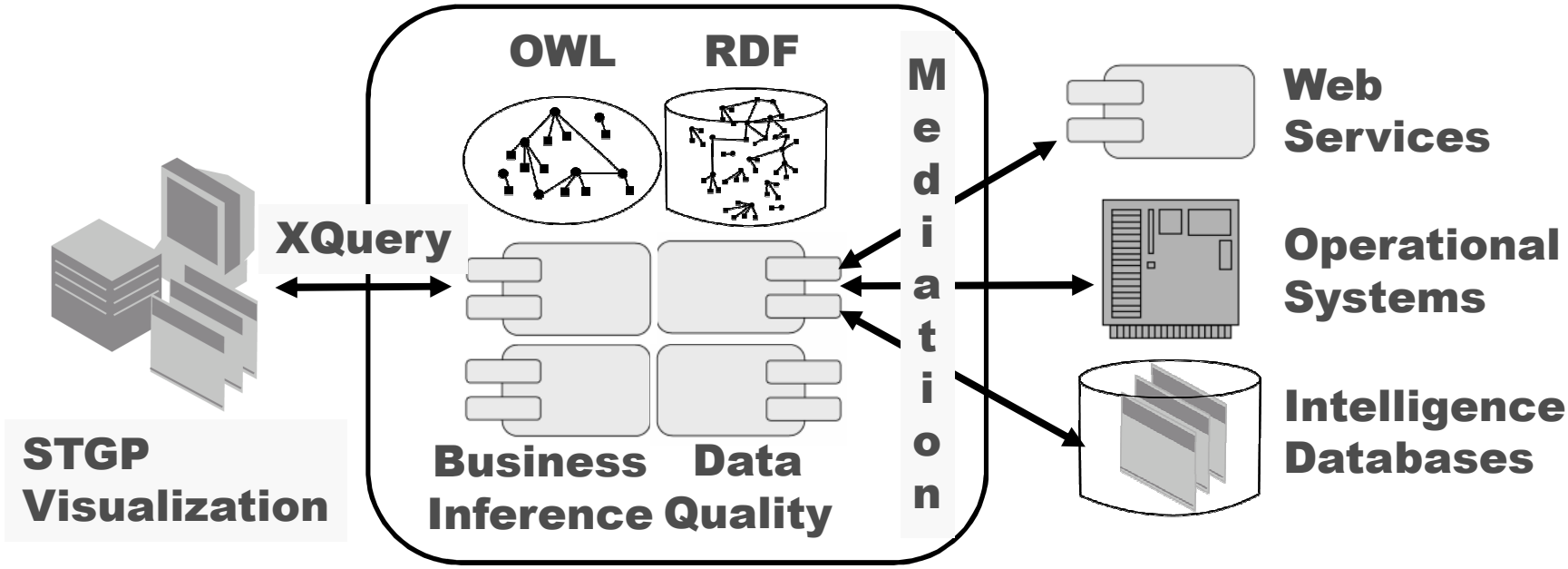


network inference

Enterprise Semantic Web In Action

Network-Centric Warfare, Shared Tactical Ground Picture

- Business Problem: Inflexible IT systems prohibit robust visibility to changing battle space IT systems
- Solution: Simple, extendable interpretation and access polices on top of mediated data – enabling actionable intelligence based on the changing rules of engagement



network inference

Why the Enterprise Semantic Web Adds Value

Lower Costs, Improved Capabilities, Increased Adaptability

Business decisions must reflect value in all areas of a balanced scorecard – that's what makes semantic technology compelling – value to ALL aspects with a dynamic computing environment.

Executive Stakeholder Value Axis	Value Proposition
Financial Performance	Dramatic TCO reduction for IT services
Customer Satisfaction	Faster IT response to business change
Business Process Improvement	More adaptable operational platforms
Organizational Benefit (Line of Business)	Reduced LoB costs/Greater value from IT
Knowledge Management	Loosely-coupled core IT "data ownership"
Human Capital Performance Management	More focus on core business
Measurement and Analysis	Greater visibility into real time data

Industries In Need

"Imperfect interoperability [of digital data] imposes at least $1 billion per year on the members of the U.S. automotive supply chain. By far the greatest component of these costs is the resources devoted to repairing or reentering data files that are not usable for downstream applications."

– RTI / NIST Report, March 1999 ["Interoperability Cost Analysis of the U.S. Automotive Supply Chain"]

- 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."

  - Defense Industry
    - Network-Centric Enterprise Services (NCES)
    - Department of Homeland Security (DHS)
  - Manufacturing
    - Automotive Supply Chain
    - Electronics & CPG Supply Chain
  - Life Sciences
    - Process Automation / Risk Analysis

Why Adopt Now?

"A little semantics goes a long way."

– James Hendler
Enterprise Architect, 2004

- Standards Leading the Momentum
  - W3C – core data and web technologies are semantically enabled
  - OASIS – ebXML is moving rapidly towards an OWL/RDF vision
  - PLCS – manufacturing lifecycle ontologies soon to be OWL ready
  - HL7 – healthcare and medical ontologies already OWL and/or DL
- Vendors Moving Rapidly With Support
  - Network Inference – 20 major customers installed w/OWL platforms
  - Sun Microsystems – Swordfish RDF data tools nearing production
  - IBM – commitment to RDF interfaces for DB2ii middleware
  - HP – market-leading open source RDF toolkit – Jena
  - Adobe – embeds RDF metadata in every binary file
- Your Friends/Competition are Adopting
  - NATO members adopting OWL net-centric warfare technologies, Fortune 500 electronics companies using OWL to drive financials, Fortune 500 medical companies using OWL/RDF for hospital maintenance systems, plus many, many more…

network inference



Network Inference, Inc.

QUESTIONS?

© Network Inference 2004

About Network Inference
Who We Are

Provide Adaptive Information Solutions

18+	**New Customers in 6 months**
3	**Years Developing Core Technology**
Solid VCs	**Nokia Ventures, Palomar Ventures**
5	**New customers in Q1**
4+	**Tier-A Business Partners**
3	**Sponsored Industry Standards**
USA	**California Headquarters**
25	**Employees; Proven Exec. Team**

network inference

network inference



Jeff Pollock

Vice President, Technology

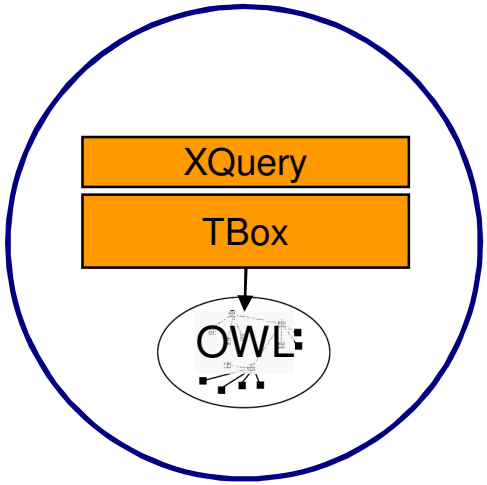
jeff.pollock@networkinference.com

© Network Inference 2004

THREE MAIN USES OF OWL INFERENCE

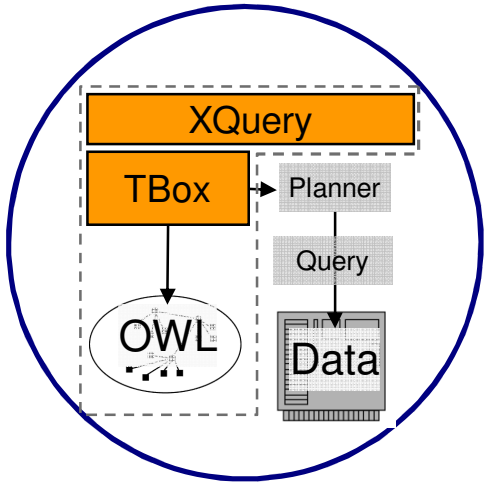
CONCEPT ONLY

- Reason on classes
- "TBox" = class only
- No instances/individuals
- Valuable for building logics & rules that can be used to drive application behavior



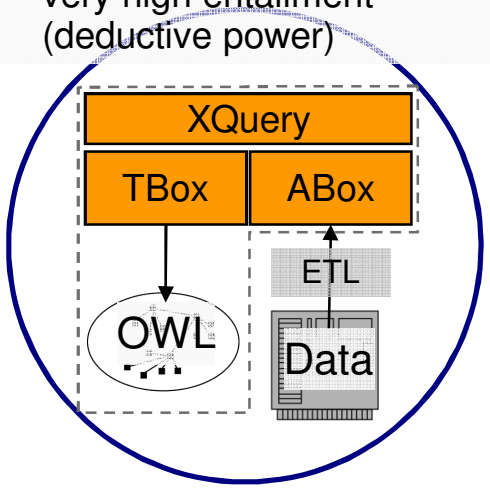
CONCEPT-TO-INSTANCE

- Reason on concepts first
- Then "get" data that belongs to concepts
- Instances come from native source (**RDF, RDBMS, WSDL, XMLDB**)
- Valuable for accessing native data at high speed and volume



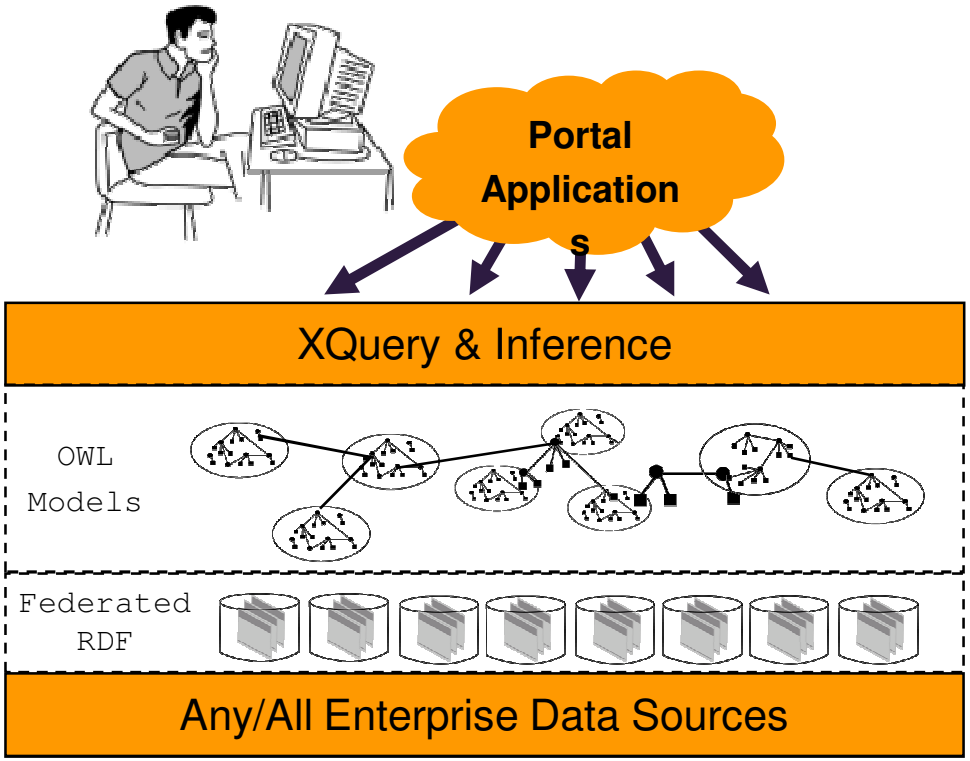
INSTANCE REASONING

- Concepts and instances
- "TBox" and "ABox"
- Instances may be in active memory, or in persisted memory (DB)
- Valuable for reasoning on real data values
- OWL instance reasoning is very high entailment (deductive power)



Enterprise Semantic Web Interface Layers

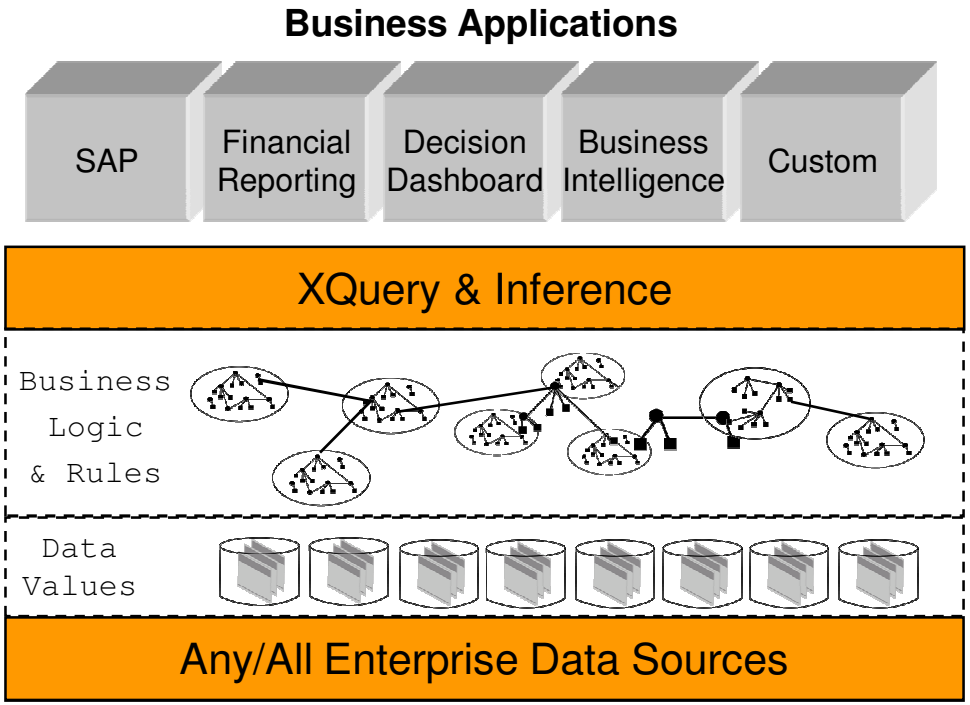
Apps talk to services, services reason about data, data originates everywhere



network inference

Enterprise Semantic Web Interface Layers

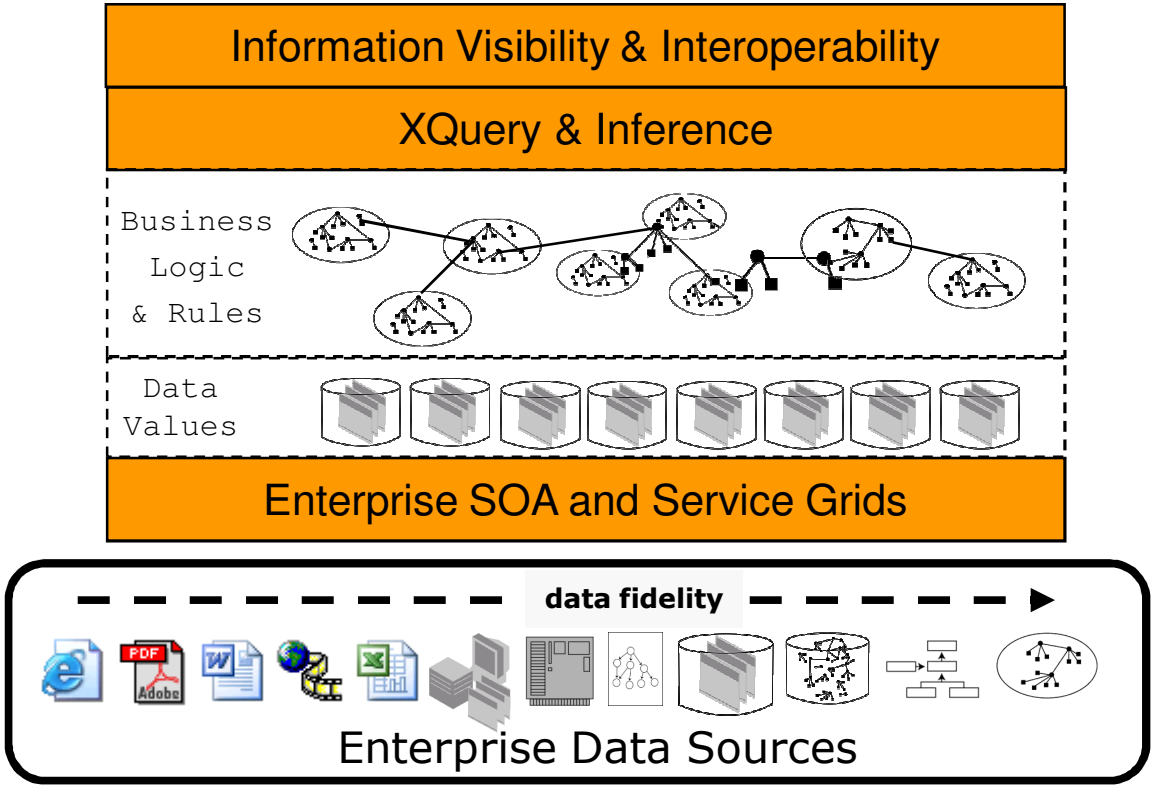
Apps talk to services, services reason about data, data originates everywhere



network inference

Enterprise Semantic Web Interface Layers

Apps talk to services, services reason about data, data originates everywhere



network inference

Enterprise Semantic Web Interface Layers

Apps talk to services, services reason about data, data originates everywhere

