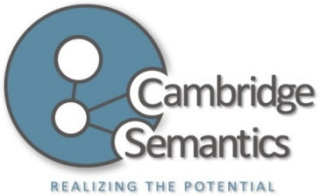
Semantic Technology for Information Management

Gilbane Conference

November 29, 2007



Discussion Agenda

- Case Study - A Global Pharma's R&D Information Challenge
- Enterprise Semantic Architecture and Strategies

Global Pharmaceutical R&D

- Data exploding faster than technology can cope –clinical and research knowledge trapped in millions of publications
- Disconnects between science, technology, business information
- Interdependence between private and public knowledge
- Disparate, isolated information models and limited insight into information and data, sources and structures
- Semantic and syntactic heterogeneity, ambiguity and uncertainty

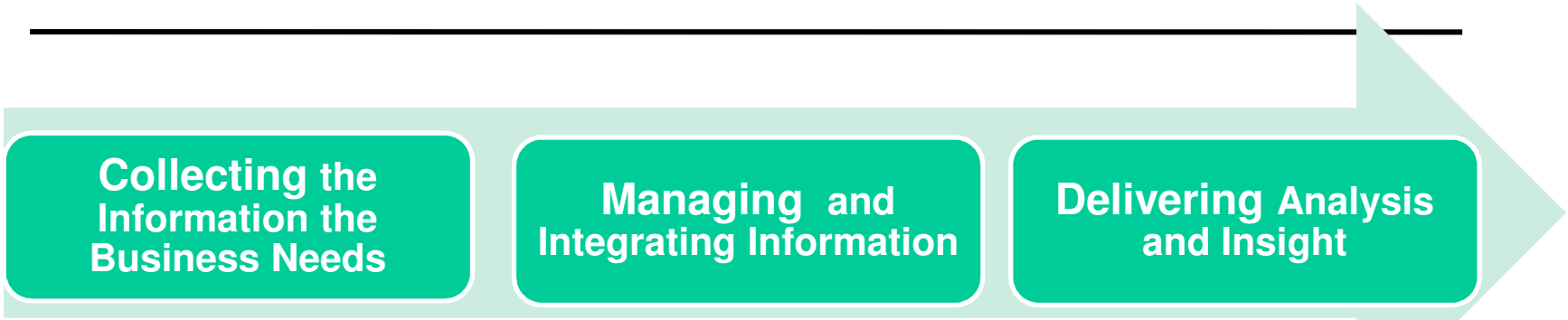
Unable to translate information into meaning and insight

Cambridge Semantics

Cross Integration

PA

The vision – Drive information value



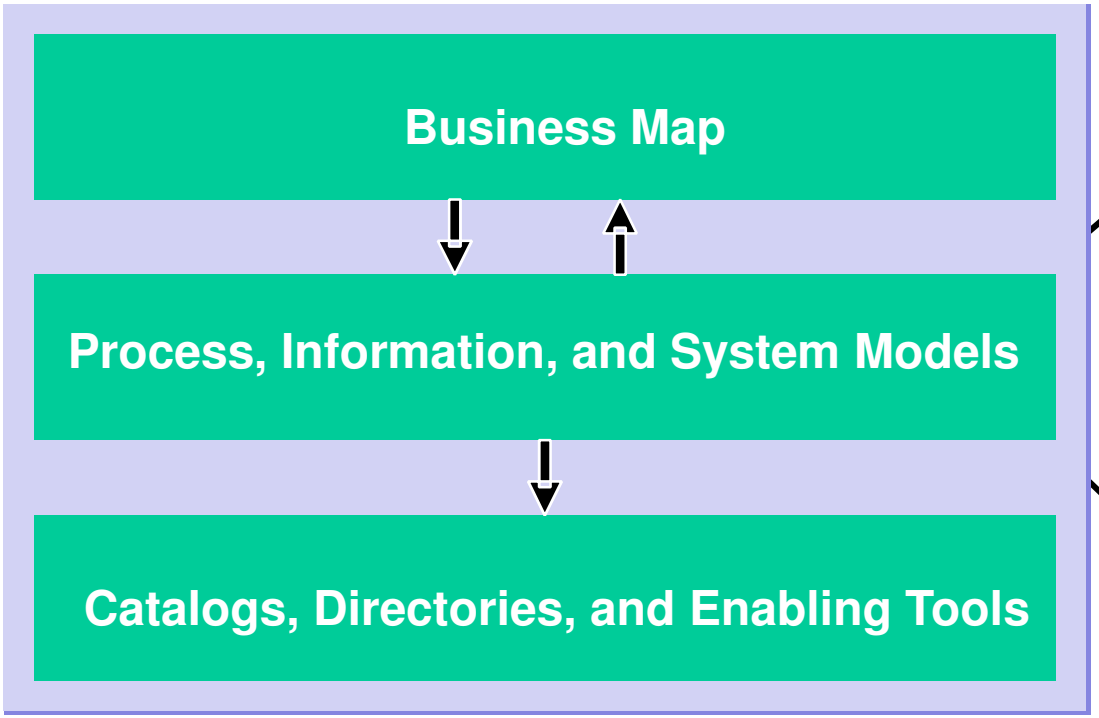
information into assets	*information as assets*	*information assets for value*
▪ Collection throughout the development continuum	▪ Authoritative sources of interconnected data	▪ Clear guidelines, security & privacy
▪ Standards, Formats for data retention and analytical re-use	▪ Trusted roles and processes for information governance	▪ Rapid access to information sources, lineage, rules
▪ Processes for managing relationships	▪ Consistent information architectures, standards	▪ Rapid application development
	▪ Effective data management, quality, compliance	▪ Information intelligence, earlier understanding
		▪ Better scientific analysis

Cambridge Semantics

Cross Integration

PA

Developing the Information Framework



Business Map provides high level representation.

Models describe the information and knowledge available at varying levels of detail, depending on need

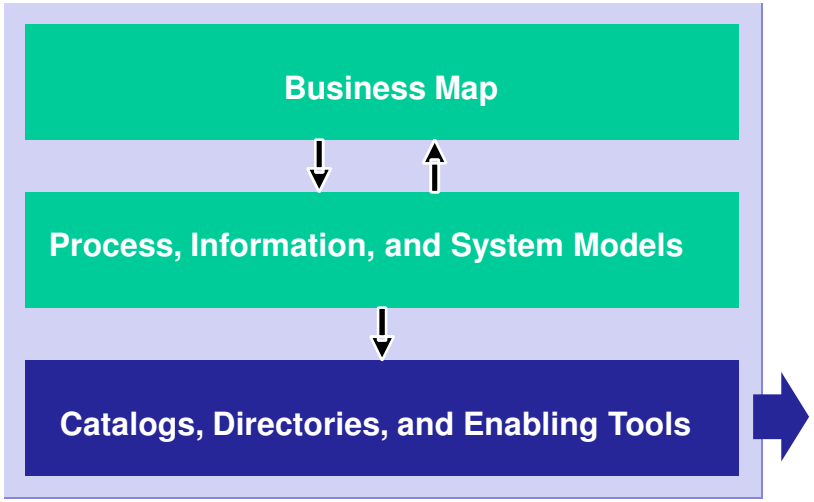
Catalogs, directories, and enabling tools support information and knowledge access and management

Cambridge Semantics

Cross Integration

PA

Seeking innovative tools, solutions



Can emerging Semantic technology help address key needs?:

- **Organization** of federated heterogeneous information?
- Quick and **intuitive access** to information?
- Data **extraction, analysis and exploration** methods?
- Text **mining** with identification of entities and the concepts they belong to?
- Effective **data integration, data exchange and interoperability** of applications?
- **Knowledge sharing, BI** and collaboration?

Cambridge Semantics

Cross Integration

PA

Enterprise Semantic Architecture and Strategies

"Semantic Technologies provide the ability to infer the meaning of data and determine its utility within a context"

Why use Semantics?

Contextual data integration

- Distributed semantic query finds information relevant to a particular user and task

Common information models

- Extensible, mergeable, shareable
- Unite shared concepts from disparate business units
- Discover relationships between information from different sources

Enterprise-wide information provenance, history and compliance

- "The right information at the right time“
- Track source of information as it’s assembled and acted upon
- Maintain an audit trail of changes to data
- Enables improved compliance

Why use Semantics (2)?

Policy-directed business rules

- Specific rules governing business processes, workflow, and data quality are derived from policies aligned with business goals. Rules can directly apply to information
- Policies can lead to machine-made decisions or assist knowledge workers' decisions
- Real-time, event-based behaviour

Business agility via data flexibility and reuse

- All business assets are declarative; nothing is locked inside of impenetrable program code
- Reuse data, metadata, policies, rules, models, workflows, actions, …
- Help reduce the enormous burden of maintenance

Why use Semantics (3) ?

Cost Savings	*Return on Assets*	*Return on Investment*
Doing same job faster, cheaper, or with fewer resources than it was done before	*Doing a better job than the one you did before, making other resources more productive and increasing their return of assets and attainment of mission*	*Changing some aspect of what the business does, resulting in growth, new value capture, mitigation of business risk, or other strategic advantage*
Impact of Semantic technologies		
• *20-80% less labor hours* • *20-90% less cycle-time* • *30-60% less inventory levels* • *20-75% less operating cost* • *25-80% less set-up and development time* • *20-85% less development cost*	• 50-500% quality gain • 2-50X productivity gain • 2-10X greater number or complexity of concurrent projects, product releases and units of work handled • 2-25X increased return on assets	• 2-30X revenue growth • 20-80% reduction in total cost of ownership • 3-12 month positive return on investment • 2-300X positive ROI over 3-years

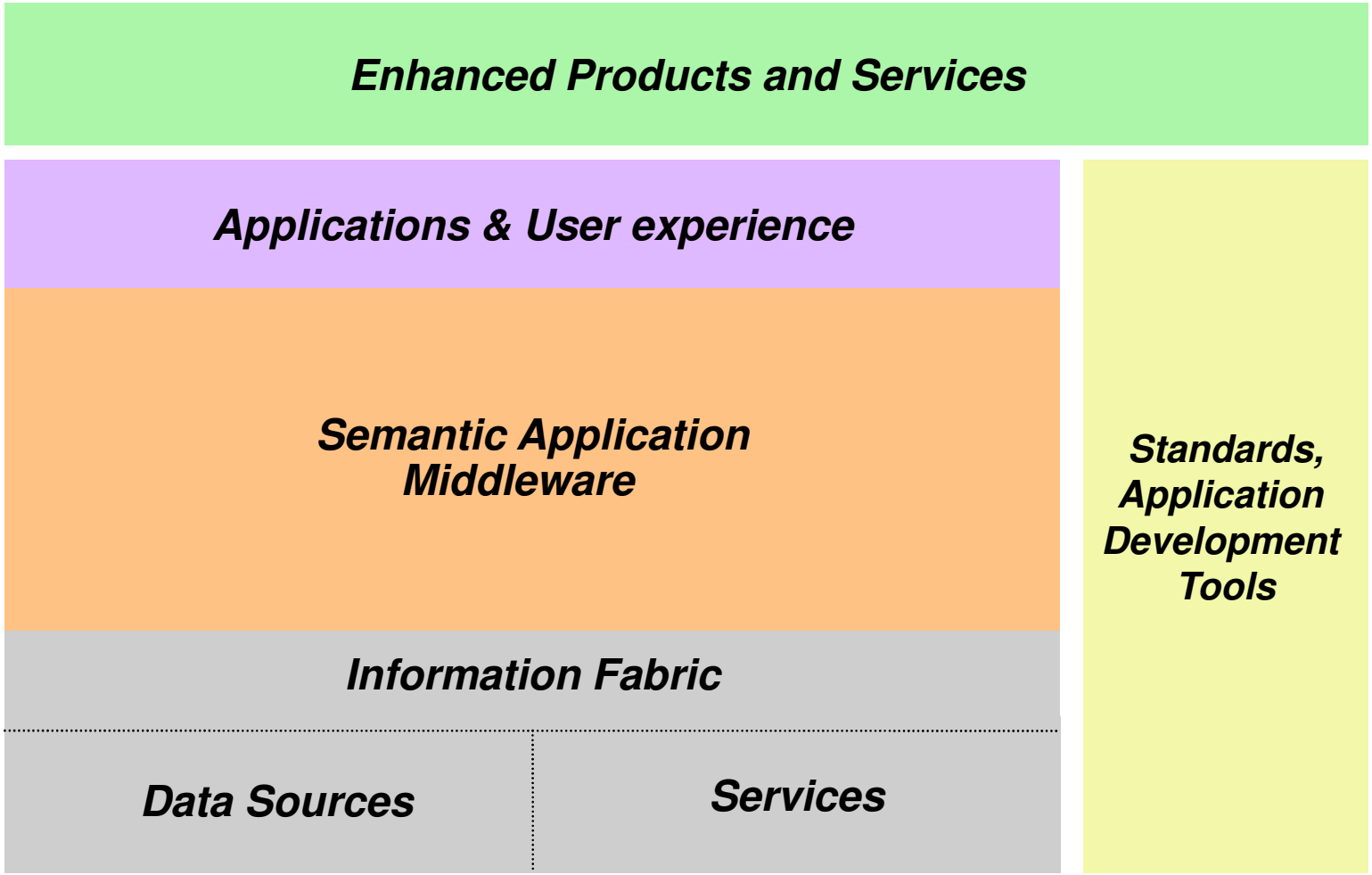
**** 10X Project; Mills Davis***

Cambridge Semantics

Cross Integration

PA

Required for a Global R&D Information Fabric



Cambridge Semantics

Cross Integration

PA

How to go about it

- Choose standards
- Model data and services
- Build an information fabric
- Select application-logic middleware
- Select user-interface middleware
- Configure, deploy, monitor, maintain, ...

Choose standards

- Standards avoid vendor lock-in, allow for incremental improvements, and encourage interoperability with legacy and future technology investments
- The W3C publishes mature semantic technology standards that can form the basis for a global R&D information fabric:
  - RDF: Flexible and extensible data modeling; unambiguous naming; a machine readable lingua franca for information
  - OWL: Sophisticated ontology language for modeling information relationships; drives consistency and compliance checking; infer new information from integration of old information
  - SPARQL: Distributed query for a semantic information fabric; tailored to easily develop data-integration queries; universal, semantic data access
  - http://www.w3.org/2001/sw/

Model data and services

- Most existing data and services already have a model:
  - Relational DDL, WSDL, UML, XML schema, enterprise data dictionaries
- OWL provides a standard way to capture the semantics of these models and to begin to interrelated them
- Semantic models tell people and software not only what the data looks like but also what the data is and how the data fits together
- Examples of semantic modeling tools:
  - Protégé - http://protege.stanford.edu/ - open source with many plugins
  - Knoodl - http://knoodl.com - Revelytix Inc. - wiki-based community ontology development
  - TopBraid Composer - http://topbraidcomposer.com/ - TopQuadrant – commercial Eclipse plugin
  - SWOOP - http://code.google.com/p/swoop/ - open source

Build an information fabric

- An enterprise information fabric is stitched together by finding commonalities and relationships between disparate services and data silos
- Data and services are all made accessible in Semantic standards forms
- New models (ontologies) can selectively and incrementally build out the information fabric
- Distributed, federated query (using SPARQL) can access the concrete enterprise data that makes up the information fabric using the ontologies as guides

Build an information fabric (continued)

- Examples Information fabric data-integration tools:
  - Wave® Semantic Data Services Layer for BEA Aqualogic SOA playform - http://modusoperandi.com - Modus Operandi – user-driven queries and semantic search
  - Metatomix MetaStudio database mappers - http://www.metatomix.com/software/metastudio.html - virtual integration layer
  - Semantic Discovery Systems - http://www.insilicodiscovery.com - In Silico Discovery – SPARQL endpoint atop a federated query engine
  - Virtuoso Universal Server - http://www.openlinksw.com - OpenLink Software – virtual database engine that supports SPARQL
  - D2RQ - http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2rq/ - open-source platform for mapping relational databases to RDF/OWL
  - Glitter - http://openanzo.org/ – Cambridge Semantics - open-source pluggable SPARQL implementation

Cambridge Semantics

Cross Integration

PA

Select application-logic middleware

- Business logic must have access to the semantic information fabric
- Semantic application middleware must support:
  - Development of business logic using traditional IT skills (e.g. Java)
  - Security – authentication and access control
  - Rules and policy
  - Data access and update
- Examples of Semantic application middleware:
  - Jena - HP Labs – open-source lightweight middleware http://www.hpl.hp.com/semweb/tools.htm
  - Metatomix Semantic Middleware - http://www.metatomix.com/
  - Open Anzo – Cambridge Semantics – open source http://www.openanzo.org/
  - Sesame – Aduna – open source http://www.openrdf.org/
  - TopBraid Services - Top Quadrant http://www.topquadrant.com/

Select user-interface middleware

- Web 2.0 trends encourage users to expect more from what's on their computer screens
- User experience must be tailored to specific jobs and must adapt as the needs of the job adapts
- Semantic applications have new requirements for working with the changing nature of information, policy-driven workflow, and newly inferred data
- Take advantage of contextual information
- Semantic user-interface middleware approaches include:
  - TopBraid Live - http://www.topquadrant.com/topbraidlive.html - TopQuadrant – based on Adobe Flex
  - OpenLink AJAX Toolkit - http://oat.openlinksw.com/ - OpenLink Software – JavaScript widgets and libraries
  - Metatomix Semantic Middleware - http://metatomix.com - based on Adobe Flex and AJAX

Configure, deploy, monitor, maintain, …

- Easy to forget where much of the real cost of enterprise R&D technology goes
- Important to evaluate semantic technologies in terms of tasks that go beyond application development:
  - How does the technology scale as new data sources are added? ...as the amount of data increases? ...as the amount of users increases?
  - Can I integrate other products
  - How are applications deployed? Can changes to the application be rolled out without disrupting productivity?
  - Can my existing IT skills be used to maintain and extend semantic applications?
  - How can I monitor the performance and usage of these applications to allocate hardware, predict and fix bottlenecks, etc.?

Conclusion

- Semantics provide both tactical and strategic benefits for implementing a global R&D information fabric
- Semantic technologies are incremental, mature, and standard
- A good semantic application architecture stacks applications atop middleware atop an information fabric atop existing data and services
- Commercial and open-source approaches exist for the various layers
- Semantic technology offerings must be evaluated with an eye towards enterprise deployment and maintenance

Cambridge Semantics

Cross Integration

PA

Thank you

- John Stone jcstone@verizon.net
  - (617) 686 7305
- Sean Martin sean@cambridgesemantics.com
  - (617) 794-0255

Cambridge Semantics

Cross Integration

PA