

Provisioning An SOA, or Now That I've Got The Infrastructure, Where Are The Services?

Chuck Mosher

cmosher@metamatrix.com

SOA for E-Government Conference

24 May 2006



Agenda

- Challenges of SOA adoption
 - Data services: a possible solution?
- Data services
 - Definition, Challenges
- Model-driven architecture for enabling data services
- Semantic matching technology
- Summary

Challenges of SOA Adoption (Bob Brown)

- Organizational
- Power & control
- (Anti)-incentives
- Ownership

- A common theme: Where is the ROI?
 - And more importantly, when?
 - Why spend the money (do I get the credit)?
 - I.e., what is the benefit in (my) taking the risk?

Implementing SOA – Where to Start?

- There is a lot of infrastructure with SOA: registries, message buses, directories, repositories, etc., but that is just the *skeleton* for an SOA.
- The meat is the services, and what will virtually every service require?

DATA

- Integrated, federated, semantically-rationalized, COI-targeted

Data Services – A Necessary 1st Step

- SOA is about integration and interoperability
 - Fundamental to this is *data*
- Processing & data must be abstracted
 - Or more silos result
- Indeed, multiple levels of abstraction
 - Support numerous use cases/COI's
 - Same data, different vocabularies
 - Support for change built in
- Use SOA principles to build SOA
 - For data, ROI is often immediate

Data Service

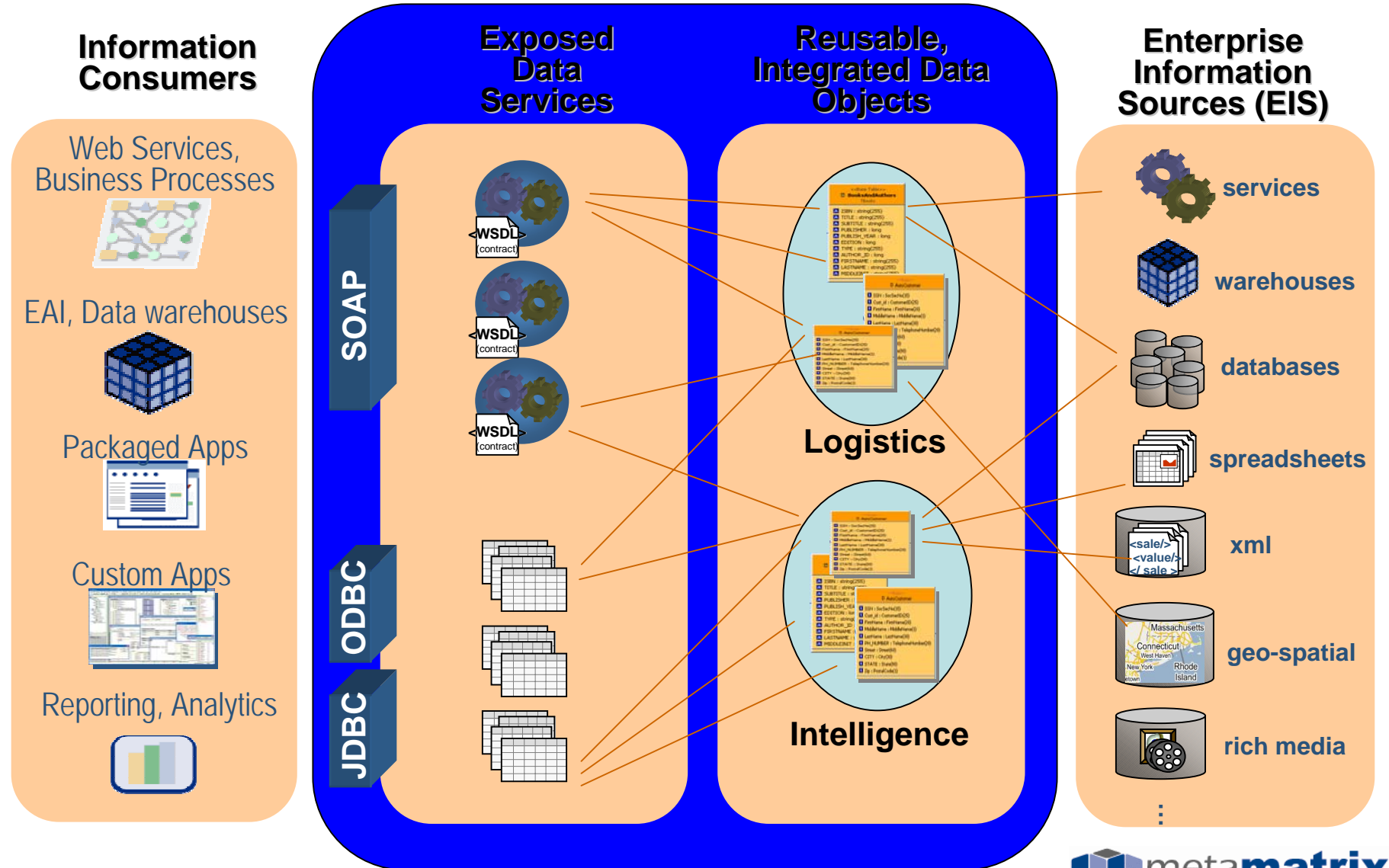
- A type of Web Service
- Does all of the work to transform any data in any format to a W3C compliant service
 - Implements all of the logic to effect the transformation
 - Provides access to data sources, regardless of source API, technology
- Does not implement application/business logic
 - Provides common data layer for business web services
- Decouples the data from the application while making the data discoverable and accessible

Data Service Challenges

- Design criteria
 - Top-down, bottom-up, inside-out
 - Query patterns (ad-hoc vs. reporting, granularity)
 - Leverage metadata for rapid data service creation
- Semantics
 - Arguably the most difficult problem
- Performance
- Security
- Metadata management
 - Leverage captured & created metadata

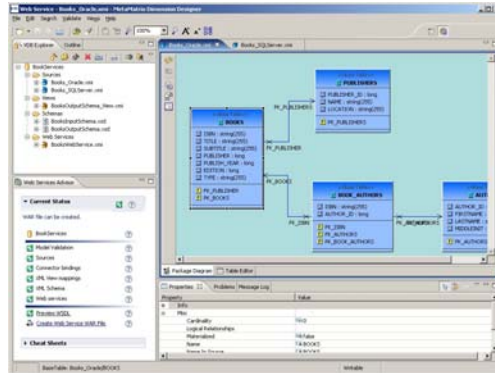
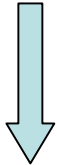
MetaMatrix Approach to Data Services

Designing data services



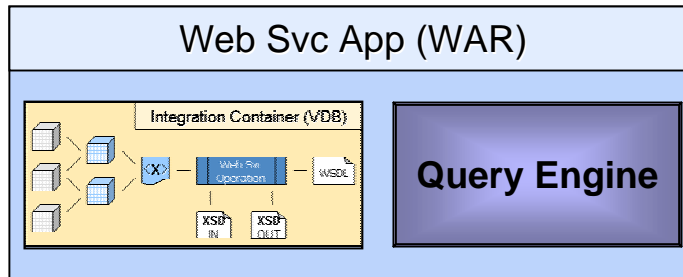
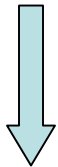
Rapid Web Service-Enablement

1 MODEL Web service



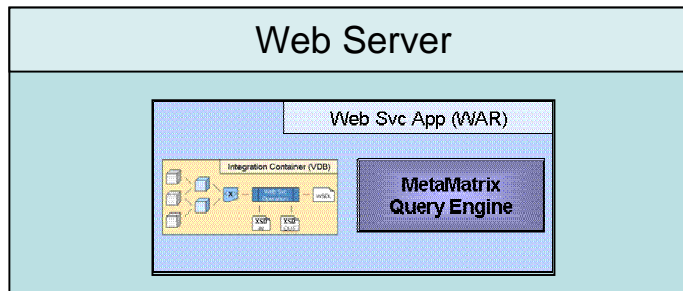
- Dimension Designer
- Model-based integration
- Expose multiple sources, integrated
- Expose business views of data
- Output an integration package (vdb)

2 PACKAGE Web service



- Standard WAR file
- Integration package is bundled
- Query engine is bundled
- Web service fully defined

3 DEPLOY Web service



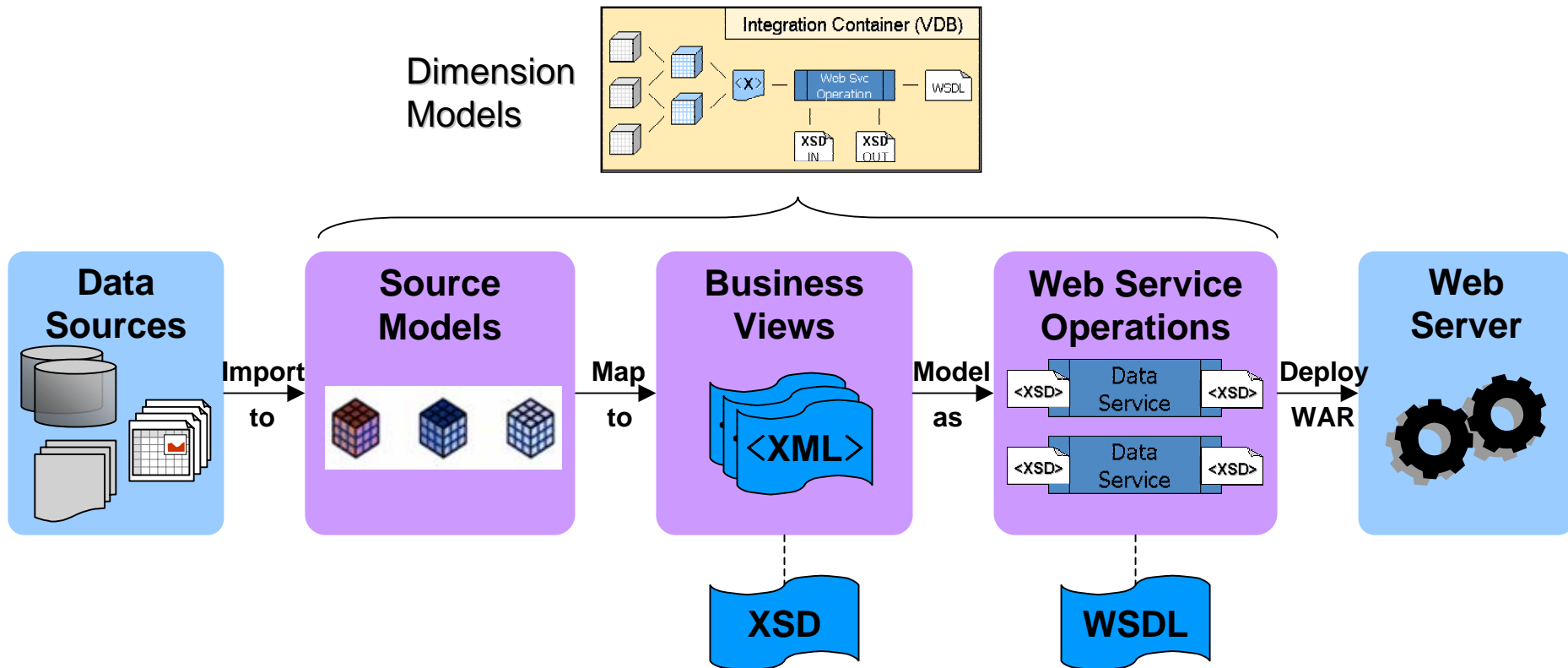
- Deploy WAR file to Web Server
- Web service fully executable
- Access via standard SOAP
- Data exposed as business views

Modeling Scenarios

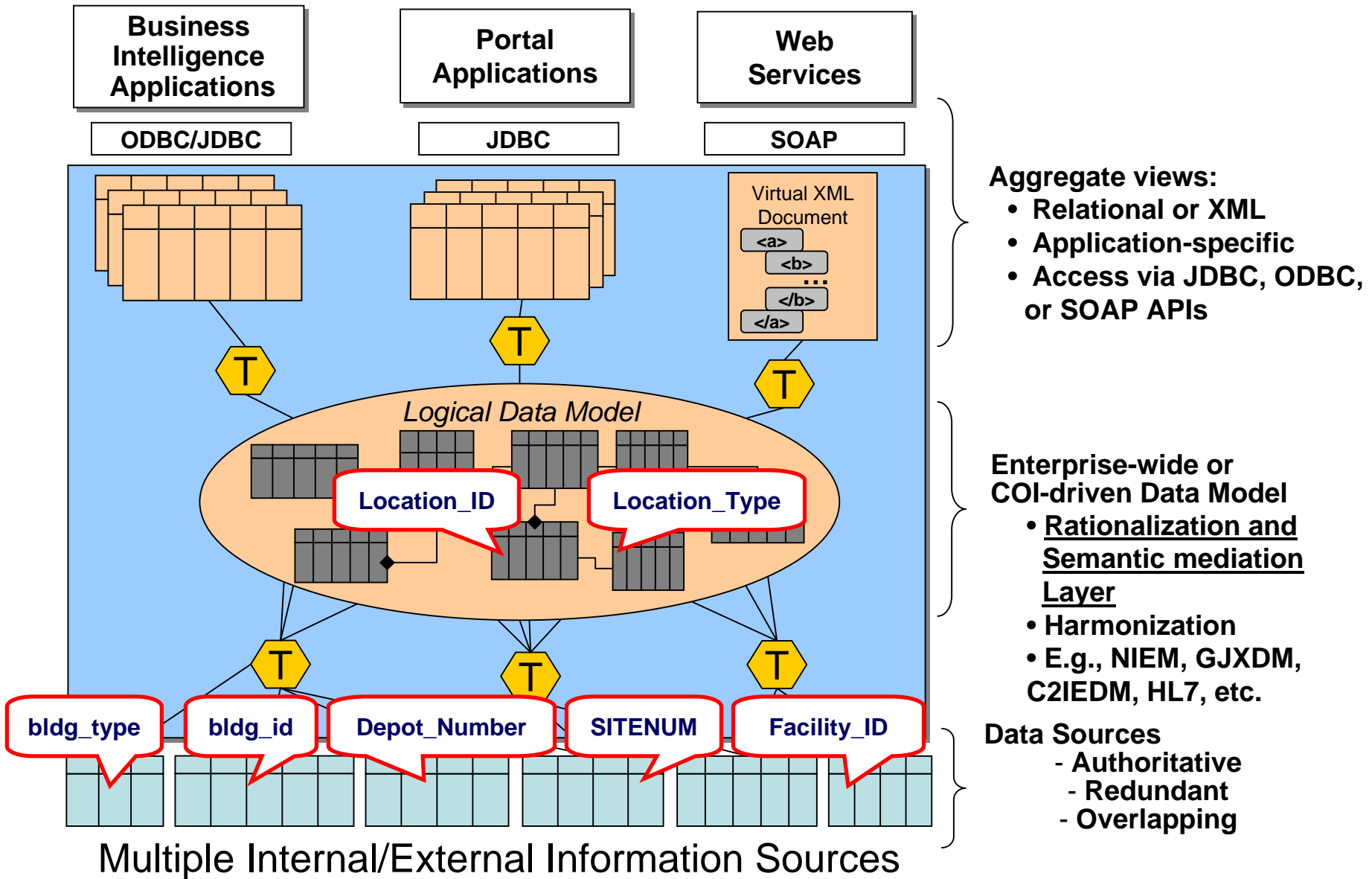
- Starting from any of these:
 - Source metadata, db schema (“bottom up”)
 - XSD-compliant XML View (“business view”)
 - WSDL definition of Web Service (“top down”)
- Expose any data as a **Web Service**
 - Relational tables and stored procedures
 - Delimited text files, incl. XML
 - Applications; e.g., ERP, CRM, mainframes, proprietary
- With any desired degree of ‘business semantics’
 - Simple “table-like” XML
 - Nested/recursive XML
 - Nested/recursive XML backed by an XSD

MetaMatrix Dimension Modeling

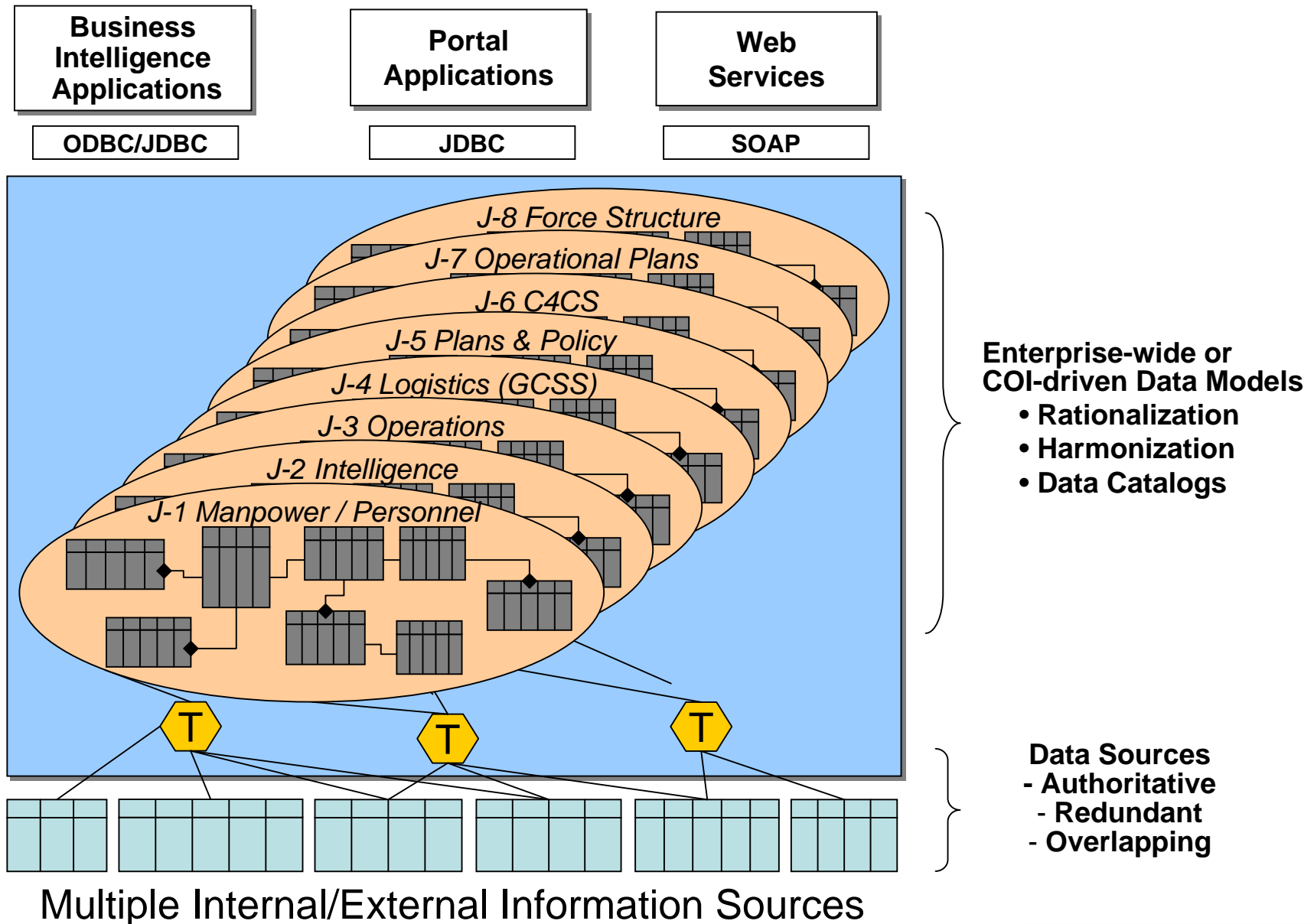
- Rapid design & deployment of **Web Services**
- Expose integrated data as XML-based business views
- Deployment of Web Services as standard Web apps
- Runtime execution optimized through use of MetaMatrix Query Engine



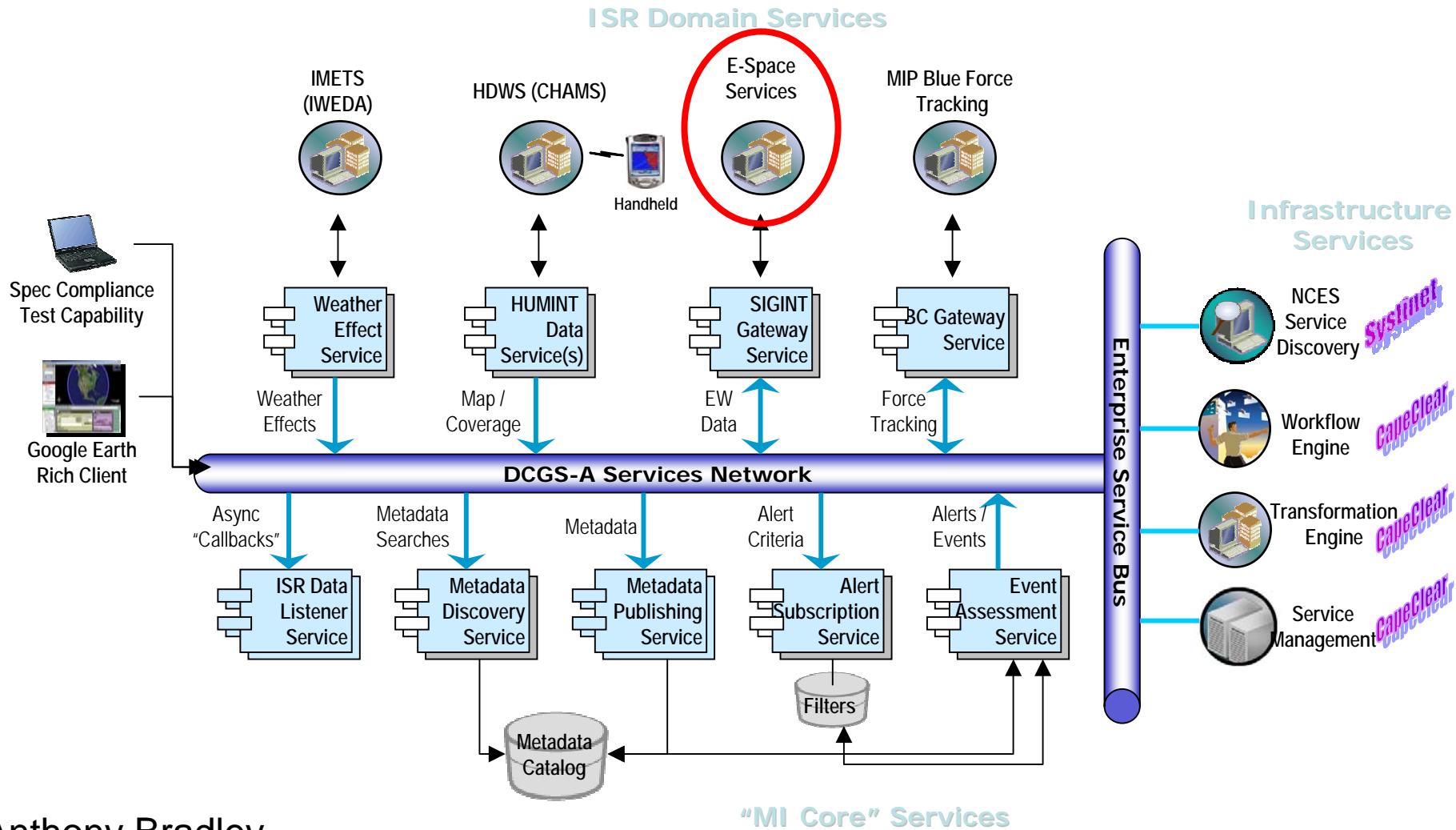
Using Models to Build a COI Data Dictionary



Building an Enterprise Semantic Model



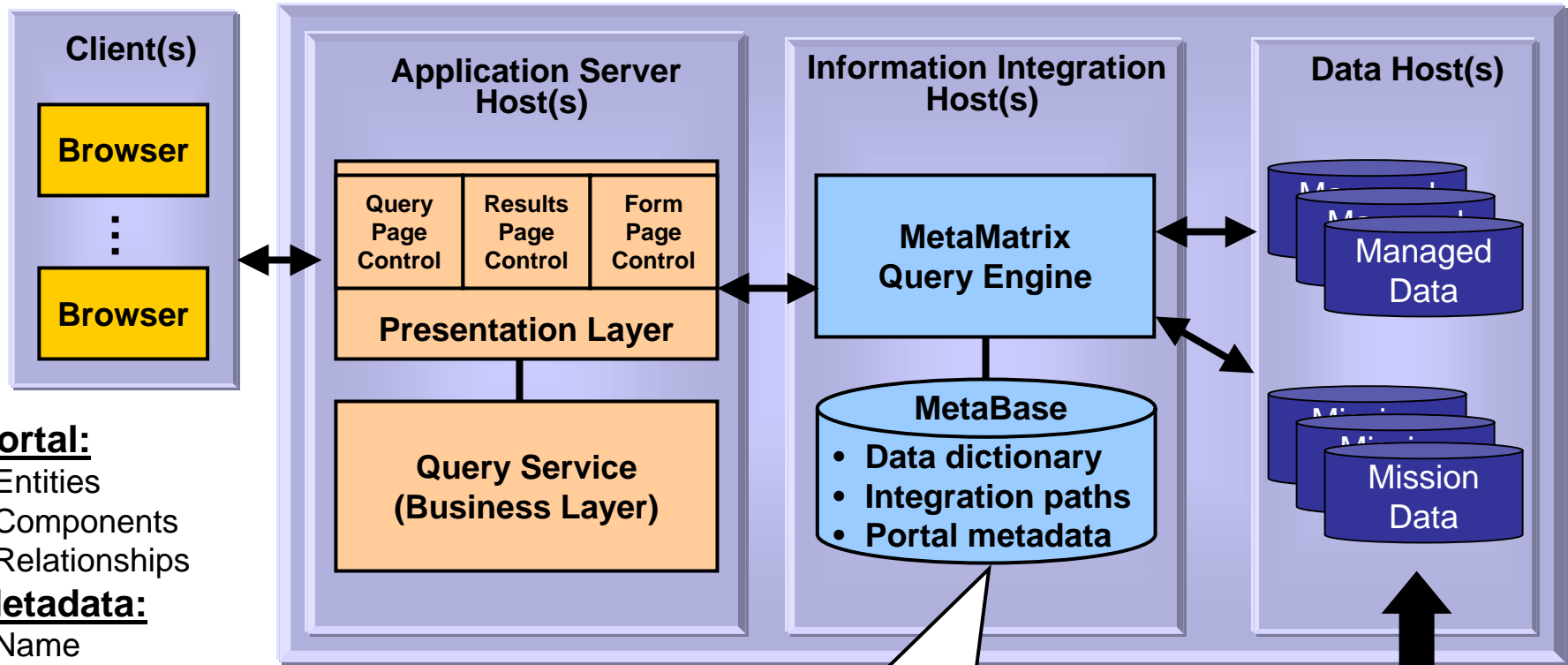
DCGS-A Reference Implementation Logical Architecture



Anthony Bradley

Booz Allen Hamilton, McLean, VA

Intelligence Agency: Information Portal

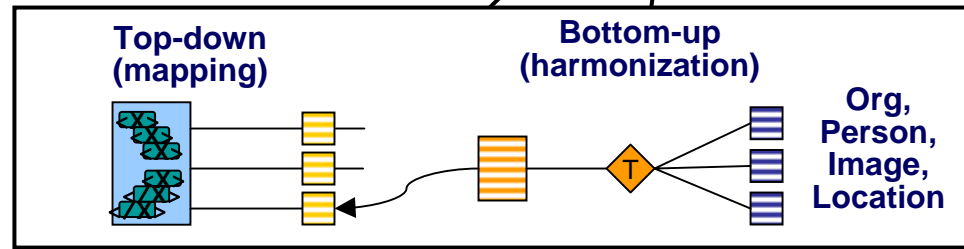


Portal:

- Entities
- Components
- Relationships

Metadata:

- Name
- InformationContext
- Usage
- Description
- Display Name
- Default Value
- Attribute Units
- Presentation Type
- Sort Order
- Visible



Biggest Challenge in Creating Data Services?

- Semantics!!!
- Structural differences are straightforward
- Differing definitions among data sources
- Differing vocabularies among COI's
- Established, emerging, and evolving data standards
 - C2IEDM, JC3IEDM, GJXDM, NIEM, GFM, many more
- Not addressed by ETL, EAI, SOA

A Previously Intractable Problem

- TWPDES has 1000+ core entities
- NIEM has 100,000+!
- Even a limited program with a dozen data sources could yield 10's of 1000's of potential mappings
- Humans cannot address this without help
- Indeed, it has stopped many data integration/reconciliation programs in their tracks.

The Matching Problem

- Given two symbols, calculate a measure of the relationship between them:

amount → quantity

Doesn't seem so hard...

The Matching Problem

- Given two symbols, calculate a measure of the relationship between them:

ftuqky → aqfkyeyr

This is what a computer “sees.”

The Matching Problem

- Even after extracting likely symbols, matching is a difficult problem.
- Symbols alone are not enough to generate good matches:
 - “ID” -> “SocialSecurityNumber” or “NY”
- The solution relies on *context*:
 - “NJ”, “MA”, “CA”, “**ID**”
 - “Ego”, “SuperEgo”, “**ID**”
- MatchIt provides that context

MatchIT 1.0 from Revelytix

- Integrated component of the MetaMatrix Semantic Data Services product suite
- Performs semi-automated semantic matching
- Based on ontology-driven semantic knowledge base
 - Word relationships, dictionaries, lexicons, thesauri
- Plug-in architecture
- Standards-compliant:
 - OWL
 - RDF
 - Inference engines
 - OSGI
 - Eclipse
 - JDBC

Matching Techniques

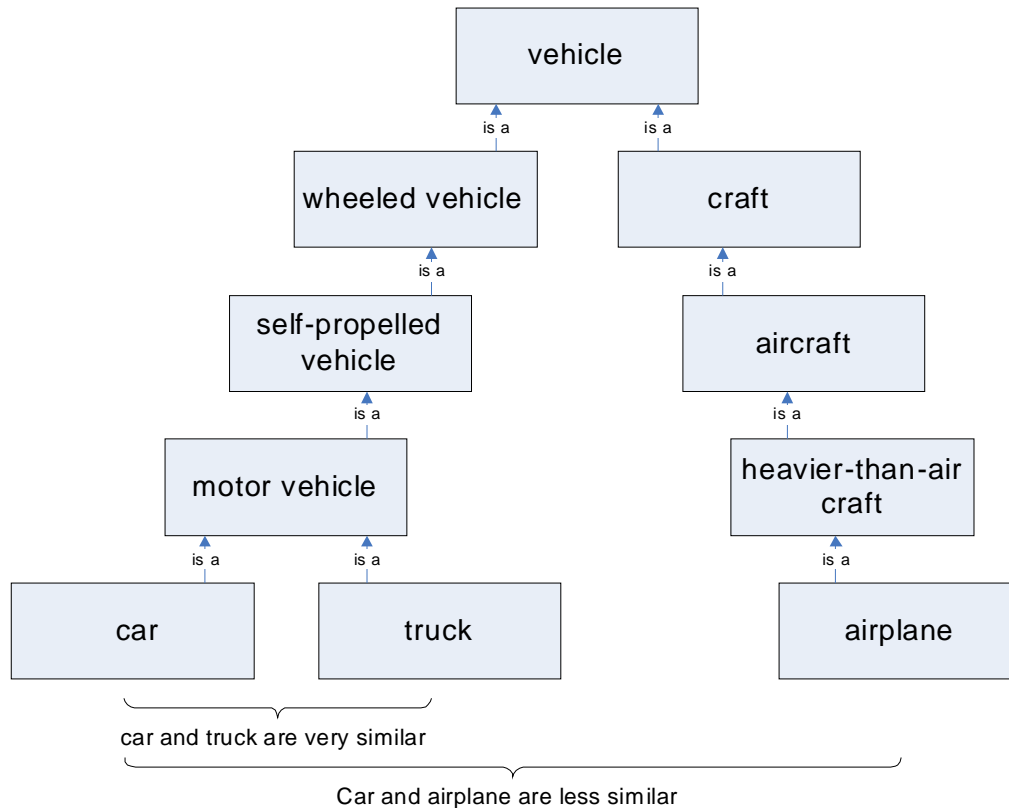
- MatchIT uses two types of matching techniques:
 - String Matching
 - Attempts to determine string similarity based on the *lexical distance* between them.
 - Semantic Matching
 - Attempts to determine string similarity based on the *ontological distance* between them within a semantic ontology.
- Generate Match Sets
- Can be run individually or in combinations
- Pluggable architecture allows for algorithmic extendibility

String Matching

- What is the lexical distance between two symbols?
 - “PUZZLE”, “PUZZ”
 - “ID”, “IDENTIFIER”
 - “STRONG”, “SONG”
 - “PKG”, “PACKAGE”

Semantic Matching

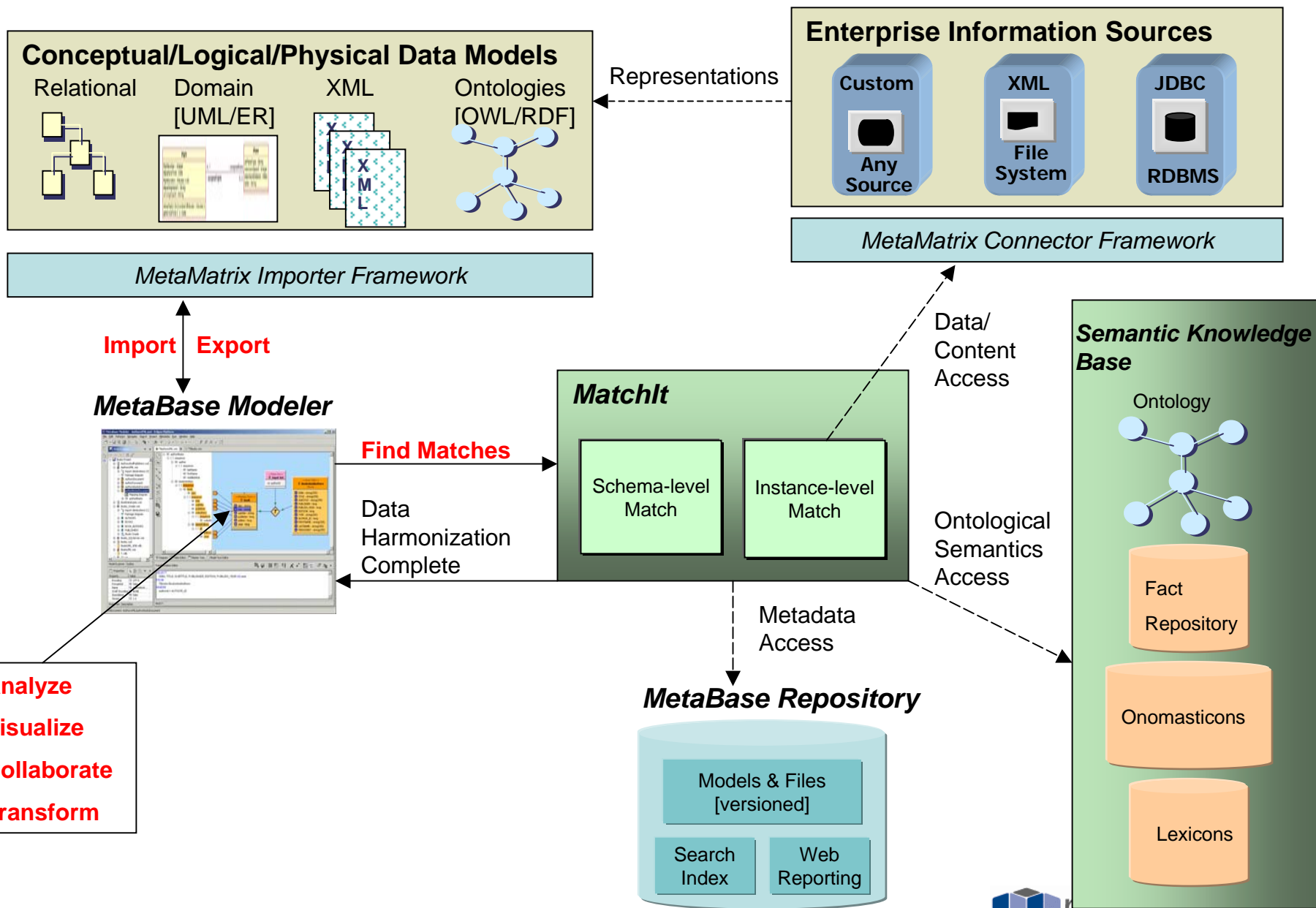
- How semantically similar are two concepts?



Semantic Matching Objectives

- Find and rank the potential matches, but let the user review and decide for sure.
- I.e., eliminate 99+% of the things that don't match, and let the user review the <1%.
- Many times, a user can visually scan a small list of the top 1% and very quickly agree or disagree with the results.
- Favor false positives over false negatives.

Semantic Matching in MetaMatrix



Semantic Matching Demo

- Import two nontrivial vocabularies
 - ERwin model of large data warehouse
 - TWPDES XML schema
- Extract symbols
 - Schema-specific tokenization algorithms
- Assign semantics to each
 - Symbols are keys into dictionaries
- Perform semantic matching between them
- Analyze results

TWPDES XML Schema

MetaBase Modeler - TWPDES.xmi - MetaBase Modeler

File Edit Refactor Navigate Search Project Metadata Window Help

Model Explorer Outline

*TWPDES_Dictionary DW_Dictionary TWPDES.xmi EWSModelPhysical.xmi

Person

- PersonSurName [excluded]
- sequence
 - PersonAffiliationDetails
 - AgencySubjectInterest [excluded]
 - PersonBiometricDetails [excluded]
 - PersonCitizenshipDetails [excluded]
 - PersonCommentText [excluded]
 - PersonContactDetails [excluded]
 - PersonEducationDetails [excluded]
 - Employment [excluded]
 - PersonAssignedIDDetails [excluded]
 - PersonKinshipDetails [excluded]
 - PersonLocationDetails [excluded]
 - PersonMedicalDetails [excluded]
 - PersonMultimediaIDBinary [excluded]
 - PersonName
 - PersonPhysicalDetails
 - PersonCapability [excluded]
 - choice [excluded]

Mapping Classes:

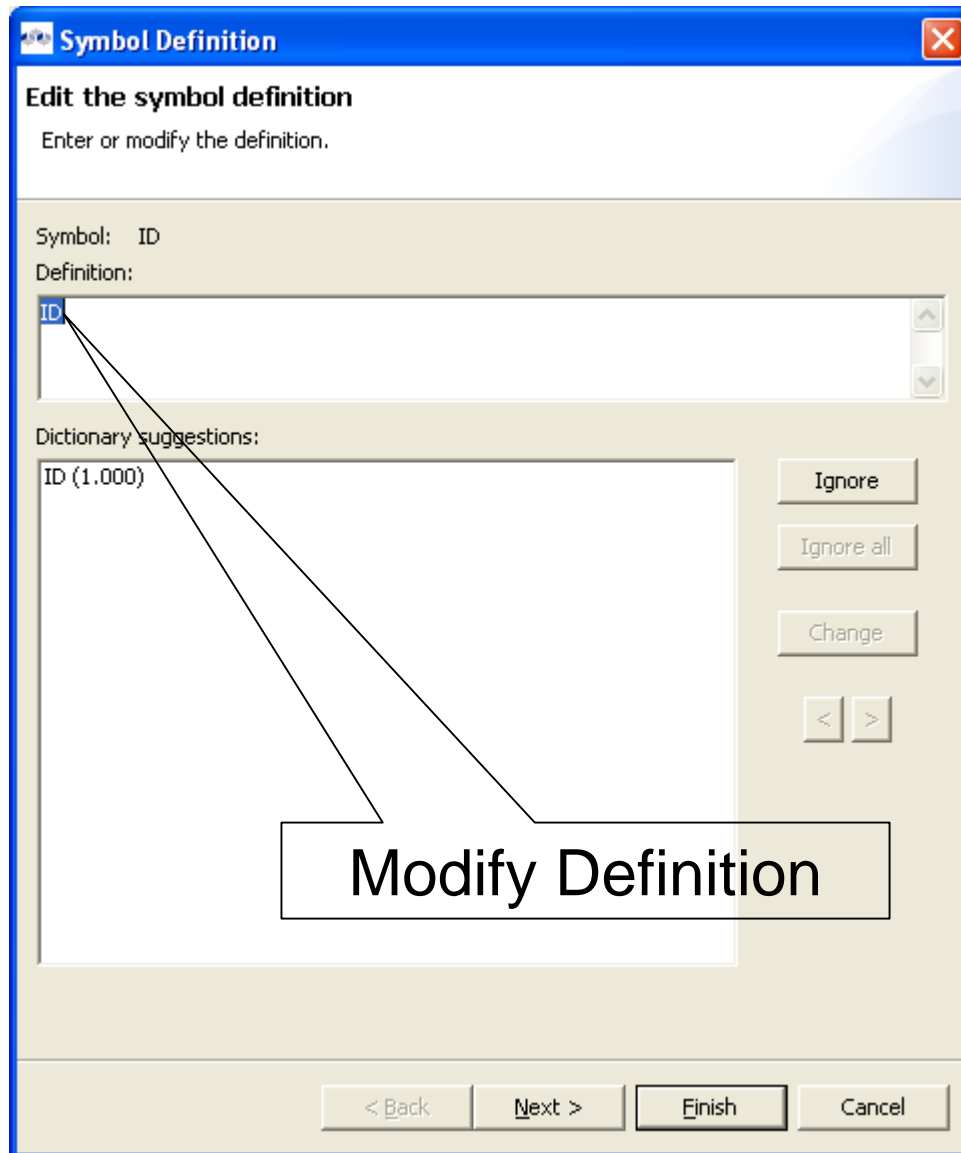
- Person
 - PersonSurName : string
 - PartyID : bigdecimal
- PersonAffiliationDetails
 - OrganizationName : string
 - OrganizationTypeText : string
 - AddressFullText : string
 - StreetFullText : string
 - AddressMailDeliveryUnitText : string
 - LocationStateName : string
 - ID : string
 - LocationCountryName : string
 - LocationCountryCode_fips10_4 : CountryCodeType
 - LocationCountryCode_iso3166Alpha3 : CountryAlpha3CodeType
 - LocaleRegionName : string
 - LatitudeDegreeValue : decimal
 - LongitudeDegreeValue : decimal
- AgencySubjectInterest
 - AgencyName : AgencyNameCodeType
 - SubjectReasonOrListText : string
- PersonBiometricDetails
 - ID : string
- PersonCitizenshipDetails
 - personCitizenshipStatusCode : PersonCitizenshipStatusCodeType
- PersonCommentText

Mapping Class: TWPDES/Mapping Class Set Container/Mapping Classes/PersonAffiliationDetails

Writable

Mapping Classes for each XML frag in hierarchy

Editing the Dictionary



Editing the Semantics

Symbol Definition [X]

Edit the semantics
Assign senses to the keywords in the definition.

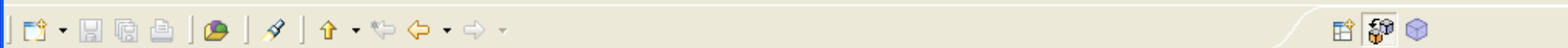
Symbol: ID
Definition: ID
Keyword: id

Available senses:
(id) noun 1: a state in the Rocky Mountai
(id) noun 3: (psychoanalysis) primitive ins

Assigned senses:
(id) noun 2: a card or badge used to iden

Control Senses

< Back Next > Finish Cancel



MatchIT Project Explorer

- FBI
 - DW_Dictionary
 - TWPDES_Dictionary
 - DataWarehouse
 - TWPDES
 - schema
 - SemanticMatches
 - StringMatches

SemanticMatches

Match Elements

Target element

Element: TWPDES.TerroristWatchlist.PersonDocument..PersonPhysicalFeature.PhysicalFeatureCategoryText

Type: Column

Similar source elements

Showing 11 of 11

<input checked="" type="checkbox"/>	Candidate	Type	Source	Similarity
<input checked="" type="checkbox"/>	Physical_Description_Securi...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PHYSIC...	0.50
<input checked="" type="checkbox"/>	Physical_Description_Value...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PHYSIC...	0.50
<input checked="" type="checkbox"/>	Party_Property_Security_T...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PARTY_...	0.50
<input checked="" type="checkbox"/>	Property_Color_Security_T...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPER...	0.50
<input checked="" type="checkbox"/>	Property_Document_Securi...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPER...	0.50
<input checked="" type="checkbox"/>	Property_Quantity_Unit_Text	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPERTY	0.50
<input checked="" type="checkbox"/>	Property_Security_Text_	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPERTY	0.50
<input checked="" type="checkbox"/>	Property_Source_Identifier...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPERTY	0.50
<input checked="" type="checkbox"/>	Property_Type_Security_T...	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.PROPER...	0.50
<input checked="" type="checkbox"/>	Case_File_Identifier_Text	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.CASE	0.48
<input checked="" type="checkbox"/>	Case_Security_Text_	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.CASE	0.48

Outline Properties

- PersonEyewearDescript
- PersonEyewearDescript
- PersonHairAppearance
- PersonHairAppearance
- PersonHairColorCode
- PersonHairFacialText_P
- PersonHairFacialText_P
- PersonHeightDescriptio
- PersonHeightDescriptio
- PersonHeightMeasure_I
- PersonHeightMeasure_I
- PersonPhysicalDisability
- PersonRaceText
- PersonSexCode
- PersonSkinToneCode
- PersonWeightDescriptic
- PersonWeightDescriptic
- PersonWeightMeasure_
- PersonWeightMeasure_
- PersonPhysicalFeature
 - PhysicalFeatureCatego
 - PhysicalFeatureDescript
 - PhysicalFeatureCategoryTe
 - PhysicalFeatureDescription
- sequence
- sequence_choice

Target Model

Match Results

MatchIT Project Explorer

- FBI
 - DW_Dictionary
 - TWPDES_Dictionary
 - DataWarehouse
 - TWPDES
 - schema
 - SemanticMatches
 - StringMatches

SemanticMatches

Match Elements

Target element

Element:

Type: Column

Similar source elements

Showing 1 of 1

<input type="checkbox"/>	Candidate	Type	Source	Similarity
<input checked="" type="checkbox"/>	Gender_Code	Column	DataWarehouse.EWSModelPhysical.Catalog.Schema.GENDER	0.67

Examine Details

Outline

- PersonEyewearDescript
- PersonEyewearDescript
- PersonHairAppearance
- PersonHairAppearance
- PersonHairColorCode
- PersonHairFacialText_P
- PersonHairFacialText_P
- PersonHeightDescriptio
- PersonHeightDescriptio
- PersonHeightMeasure_J
- PersonHeightMeasure_J
- PersonPhysicalDisability
- PersonRaceText
- PersonSexCode**
- PersonSkinToneCode
- PersonWeightDescriptic
- PersonWeightDescriptic
- PersonWeightMeasure_
- PersonWeightMeasure_
- PersonPhysicalFeature
 - PhysicalFeatureCategori
 - PhysicalFeatureDescript
- PhysicalFeatureCategoryTe
- PhysicalFeatureDescription
- sequence
- sequence_choice

Match Details

Match Details

Similarity Measure Details

Candidate: Gender_Code
Source: DataWarehouse.EWModelPhysical.Catalog.Schema.GENDER
Similarity: 0.67
Components of the similarity measure:

- 0.67 (Semantic similarity)

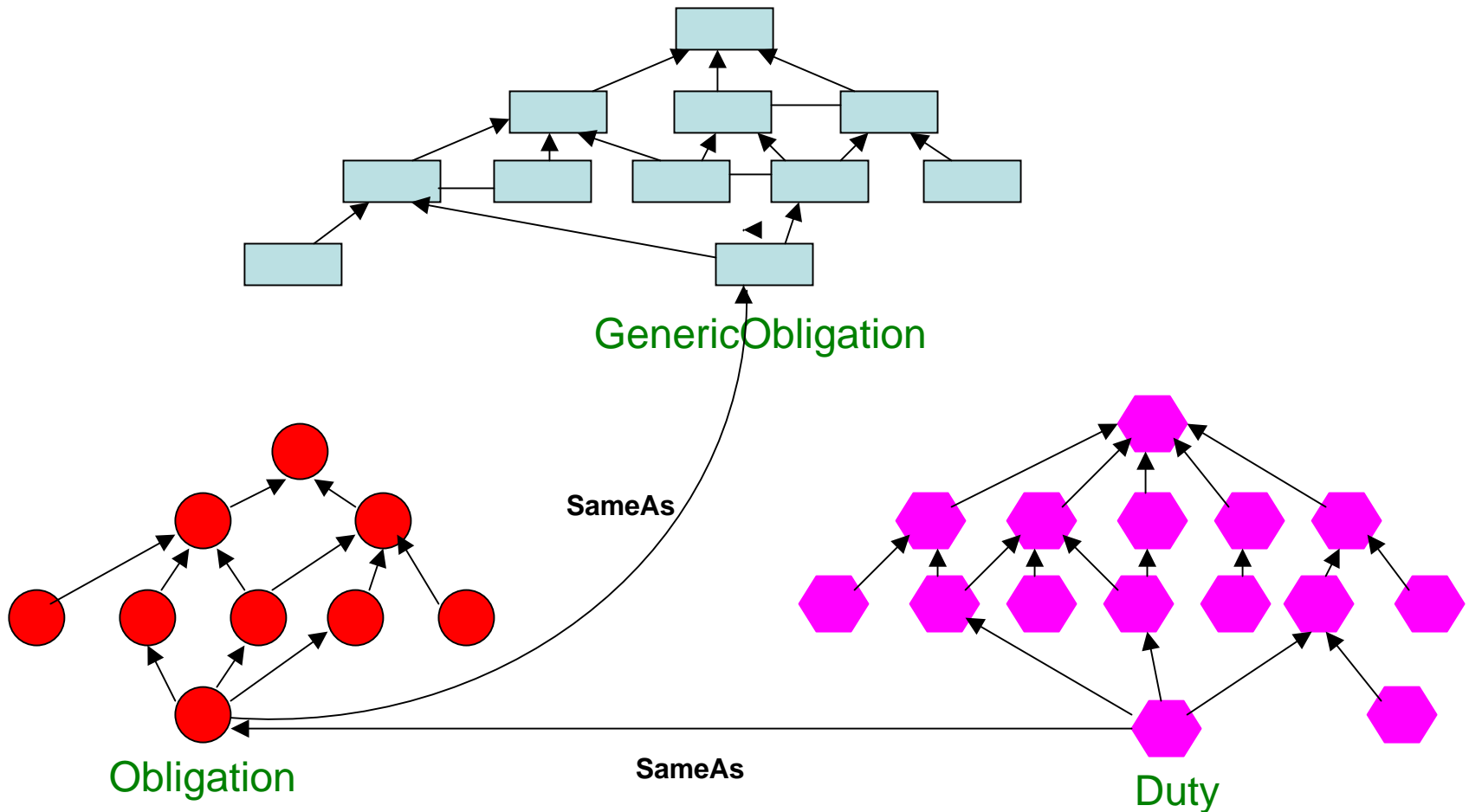
Details of the selected component of the measure:

- 0.3333 -> CODE{1} contributes 33% to name, and is the same symbol used by source
- + 0.3333 -> SEX{1} contributes 33% to name, and has similarity of 1.0 with synonym GENDER{1} - both are synonyms

Close

Relating Ontologies

The Integrating Function of the Common Semantic Model – via Domain-level Mapping



MatchIt Semantic Matching Tool

- A way to use ontologies in a world where nearly 100% of what already exists is not in an ontology.
- Map connections between ontologies that are being built and artifacts currently in use:
 - RDBMs schemas
 - XML and XSD files
 - Spreadsheet data
 - More coming, including ontologies!
- Map an imported model to a Vocabulary, and a Vocabulary to an Ontological structure

Summary of Data Services ROI

- Achieve SOA architecture with existing data assets
- Access integrated data from multiple, disparate sources in real-time
 - Fresh data
 - Consistent data from authoritative data sources
- Information in the format & semantics required
 - Semi-automated semantic matching
- Metadata & Model Driven
 - More adaptable to change than code
 - Greater re-use
 - Faster time to deployment
 - Easier change management of data
 - Significantly lower cost

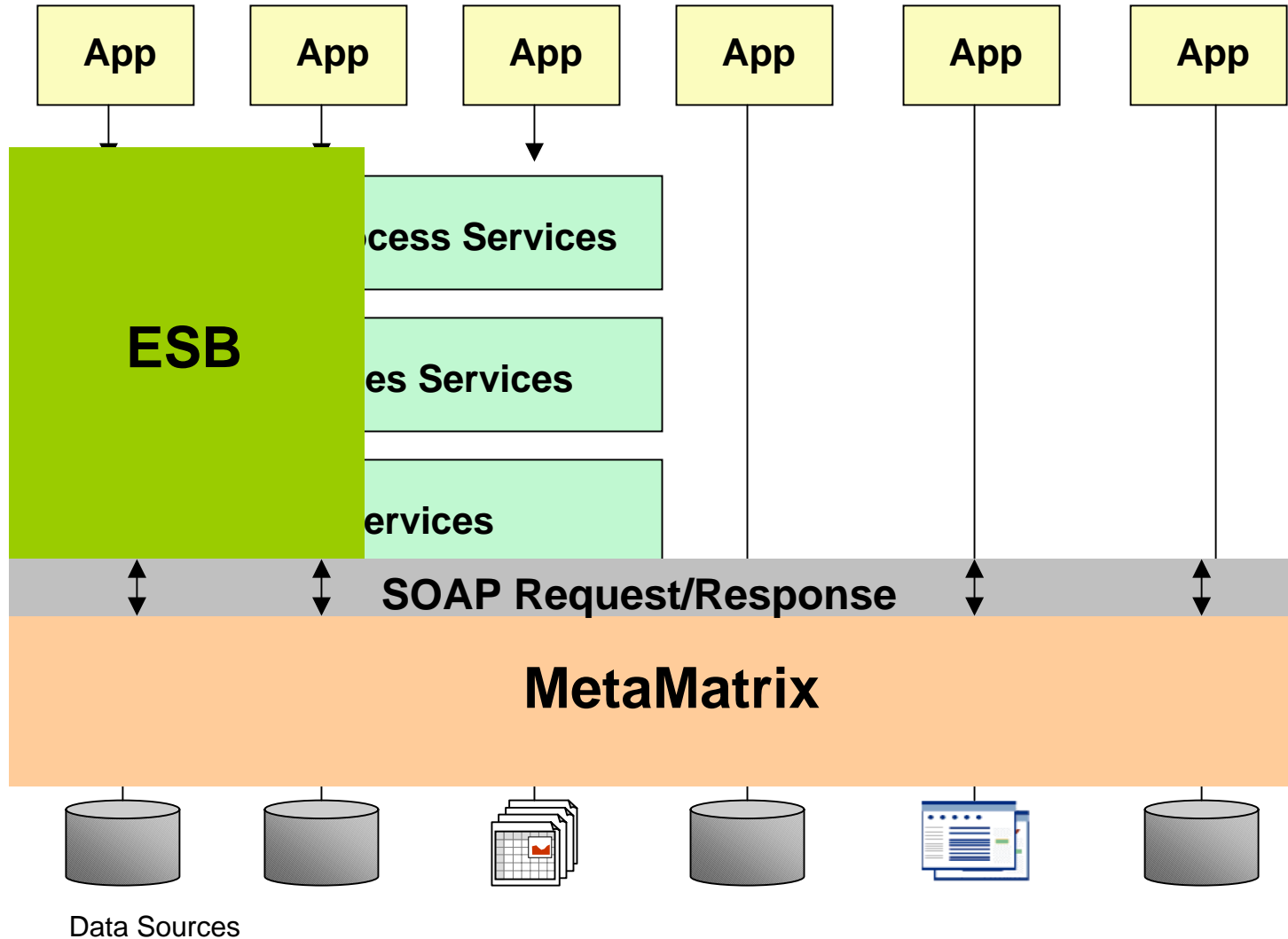
References

- www.metamatrix.com
 - White papers
 - Webinars
 - Company information
- devcentral.metamatrix.com
 - Downloads
 - Documentation
 - Developer forums, blogs
- Whitepaper: “An SOA Data Strategy for Federal Agencies: Data Services Provide Data Access and Interoperability - *Implications for DoD's Net-Centric Enterprise Services*”:
http://www.metamatrix.com/pdfs/SOA_Data_Strategy_Federal_WP.pdf

Backup Slides

Data Service Layer in SOA

Client Process & Applications



Data Service Substitutes

- **Hard coding**
 - No data or integration re-use
 - Longer time & cost
 - No change management
- **Enterprise Service Bus (ESB)**
 - Message mediation only
 - No data abstraction
 - Semantic mediation must be coded
 - Limited transformation & access performance enhancement
- **Data Integration Software**
 - No data abstraction & modeling
 - No semantic mediation
 - Limited data and integration logic re-use
 - Many are only batch processes

Components of an SOA framework

- **Use MetaMatrix for**
 - Service-enablement of data sources
 - Metadata-based, model-driven abstraction layer
 - Semantic and syntactic mediation & federation
 - High performance transformation and access
 - Data and metadata management
- **Use an ESB for**
 - Reliable delivery with QoS levels
 - Message management
 - Service orchestration
 - Business process modeling

MetaMatrix Enterprise Differentiators

- Extensible connector framework
 - Metadata driven
 - Wrappers-based connectors problematic in real-world – requires professional services
- Management & Auditing
 - Dynamic configuration of connectors
 - Seamless versioning of Data Services
 - Audit all data use
- Distributed Architecture
 - Failover, Load-balancing, and Linear scalability
- Interoperable Security
 - Single sign-on for data sources
 - Granular access control and credential management
 - Leverage LDAP and external authentication
 - SSL for secure communication
- Caching and Staging
 - Result set caching and materialized views
- Metadata Management
 - Comprehensive metadata management repository
 - Multi-user support with check-in/check-out, versioning, reporting
- Transaction Management
 - Comprehensive XA & two phase commit support

Conclusion

- Using Metamatrix eliminates custom coding and allows for an agile infrastructure that is built on standards and is highly adaptive to change
- MetaMatrix provides best-of-breed components for rapidly deploying data services to any SOA
- New products give flexibility in applying MetaMatrix technology to projects of varying size
- MetaMatrix is a proven entity within the DoD & IC, and civilian agencies.
- Unique value-add for semantic matching and enterprise search.