Practical lurches towards semantic interoperability:

Standards and mash-ups in production and in development

James Bryce Clark
Director, Standards Development, OASIS

jamie.clark@oasis-open.org
www.oasis-open.org

Ontolog Forum
March 2008

Who's OASIS?

Colophon

RDF: Mandated in the EU

WSDL, SEE and SOA

CAP: Meaning Lite

Identifier Purgatory

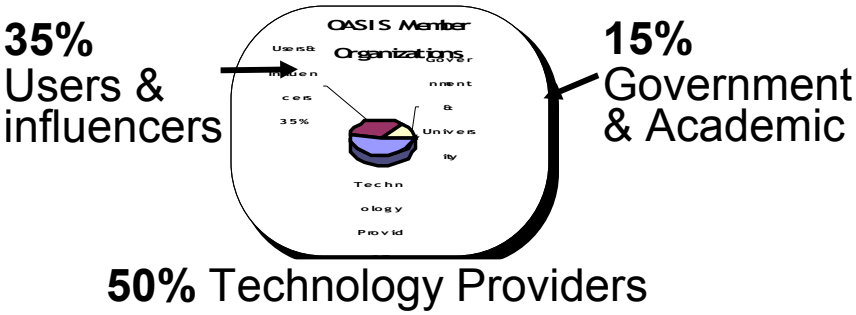
Core Components in SemanticLand

Who is OASIS?

- **OASIS is a member-led, international non-profit standards consortium concentrating on global e-business**
- **Over 650 members**
- **Over 60 technical committees producing royalty-free and RAND standards**

"The largest standards group for electronic commerce on the Web"

The New York Times ON THE WEB

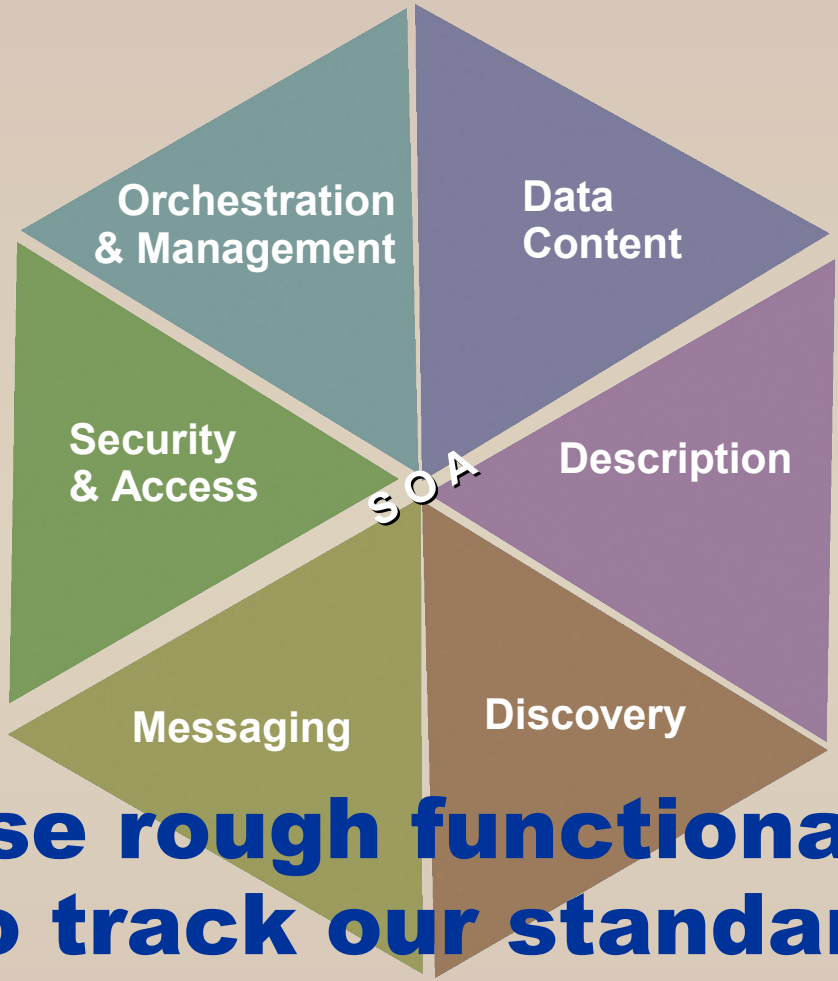


OASIS interoperates with the world:

- **Cooperation, liaison and harmonization with other standards organizations is a first-class OASIS priority**
  - Working to reduce duplication, promote interoperability
  - Gaining sanction/authority & adoption for OASIS Standards
- **Formal working relationships with:**
  - ISO, IEC, ITU, UN-ECE MoU for E-Business
  - ISO/IEC JTC1 SC34, ISO TC154 (Cat. A Liaison)
  - ITU-T A.4 and A.5 Recognition
  - IPTC, LISA, SWIFT, UPU ...
  - ABA, ACORD, AIAG, ANSI, INCITS, HL7, European ICTSB, CEN/ISSS, ETSI, PISCES, LRC ...
  - Asia PKI, Changfeng Alliance, EA-ECA, ECIF, KIEC, PSLX, Standards-AU ...
  - CommerceNet, IDEAlliance, OAGi, OGC, OGF, OMG, GS1 (RosettaNet/UCC), W3C, WfMC, WSCC, WS-I ...

OASIS interoperates with the world: sharing our successes

- ebXML > ISO = ISO TS 15000
- OpenDocument > JTC1 = ISO/IEC TS 26300
- SAML > ITU-T = ITU Rec. X.1141
- XACML > ITU-T = ITU Rec. X.1142
- WebCGM > W3C = Final Recommendation
- UBL v2.0 > UN/CEFACT = in process
- CAP > ITU-T = ITU Rec. X.1303



OASIS rough project map

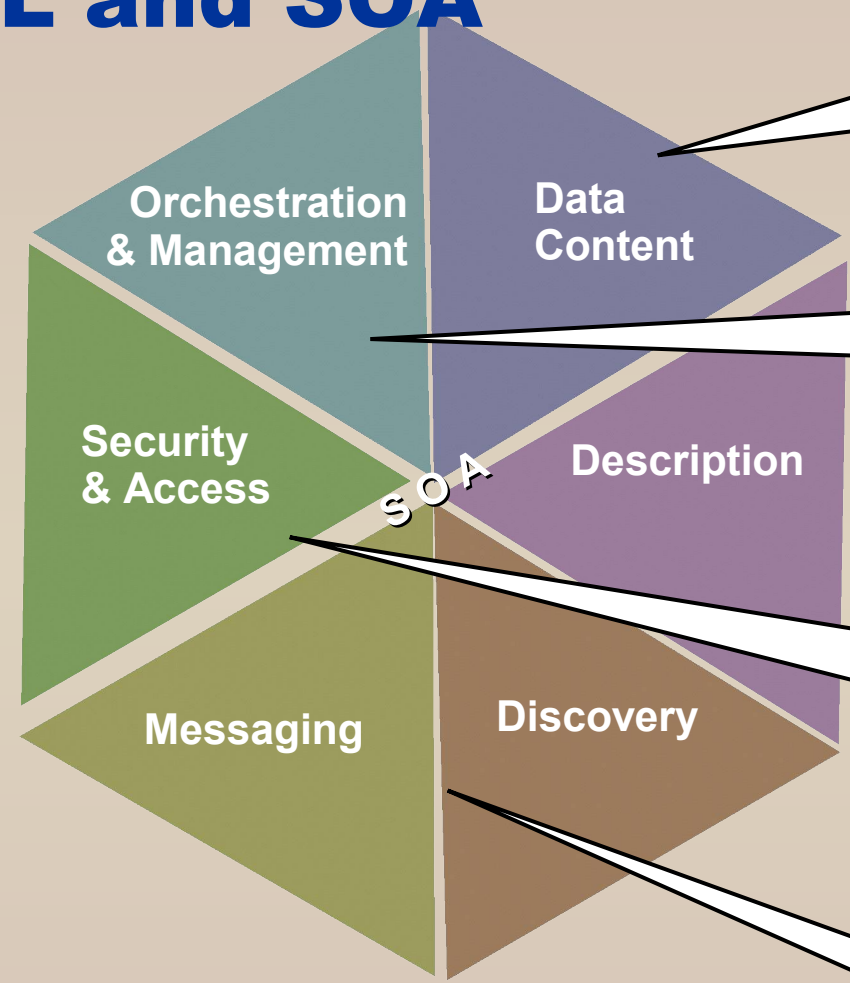
We use rough functional categories to track our standards work

Common language (XML)

Common transport (HTTP, etc.)

Developing standards for XML and SOA

OASIS

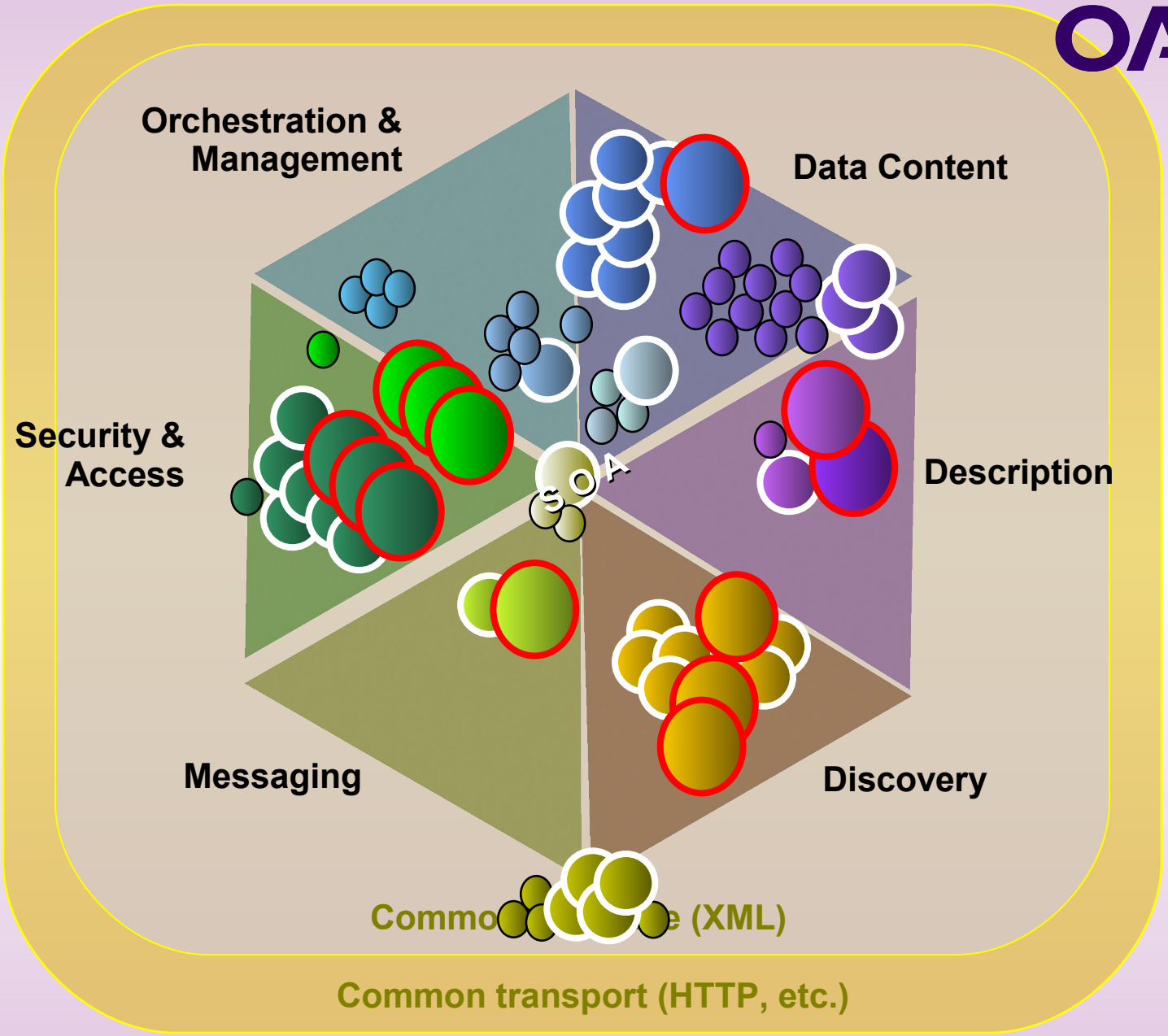


Business data formats

Common transactional methods

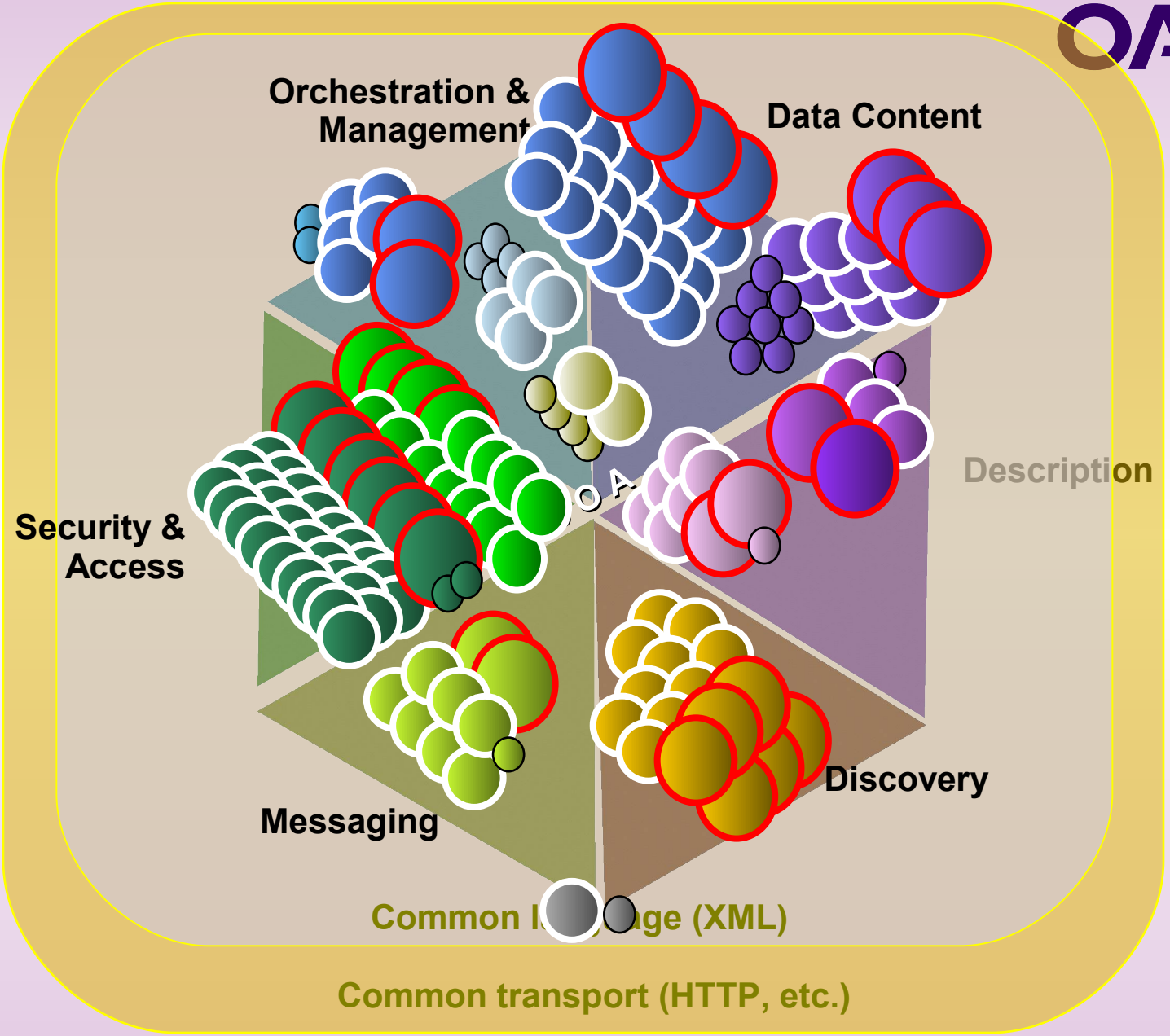
Compliance, security & risk management

Infrastructure & "plumbing"

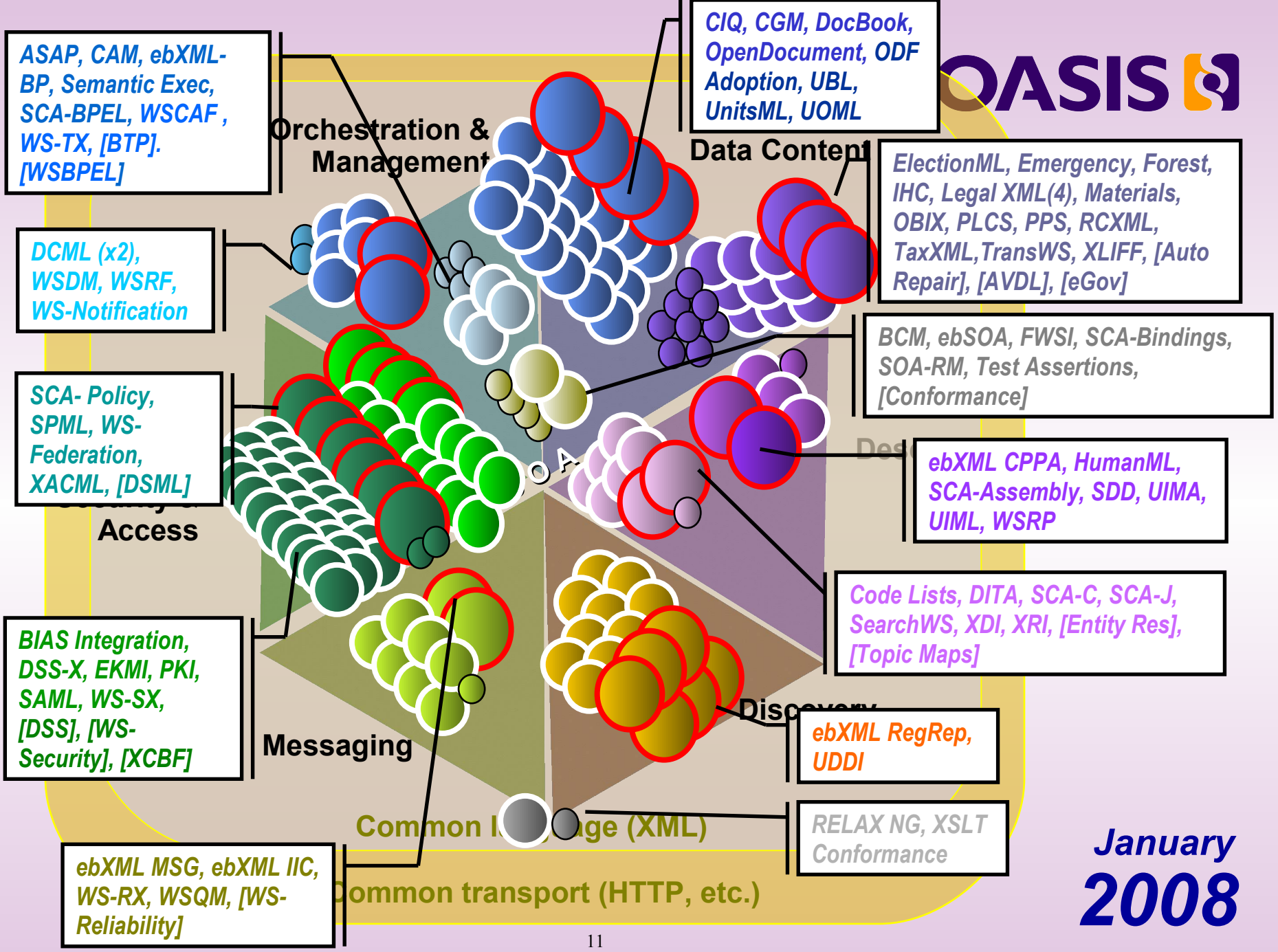


Approval levels

June 2004



January
2008



January
2008

Colophon:

Today's virtual chalk talk

Salient features of today's talk

- A report from some of the battlefronts where meaning is being retrofitted, shoe-horned or cajoled into standardized electronic data exchanges.
- Perspective from the KR layman: from those upon whom semantics are visited, not those who invent systems. Realpolitik, not research.
- Intended to be informative and provocative, but not comprehensive.
- Probably will more raise questions than answers.
- May evince a common theme about *bottom-up* vs. *top-down* creation of meaning.
- Still, it's early days on the fields of KR battle.

(This presentation provided in the ISO/IEC standard formats ODF and PDF.)

Information was born free ...
… but everywhere, it is in silos.

(Apologies to F. Hayek)

- Some communities of transactors understand each other quite well.
- Some have the shared knowledge, and trust, to do so, but lack a communication method. (CAP)
- Some have the shared knowledge … not so much trust … so need structure for communication and reliability both. (Auto Repair Info)
- Some may be from different perspectives entirely, lacking a shared frame of reference as well. (Core Components?)
- All of these are customers for practical KR.

RDF: Mandated in the EU

The limits of legislating standards.

A hopeful sign.

Information problems with known dimensions and boundaries.

Bottom-up or top-down?

The Auto Repair Information Project Saga

- In 2002-03, a group of European automobile manufacturers (OEMs), repair industry representatives and regulators from the EU Enterprise Directorate, convened an OASIS TC.
- Objective: Define data exchange specifications for data about certain vehicle repairs & parts, to make it broadly available to all repair shops. [1]
- TC defined and issued a mutually acceptable data structure. [2] But they declined to approve it by final vote, over disagreement over bearing the cost of provisioning that data. [3]

The Auto Repair Information Project Saga

- A number of incumbents also feared losing their business as information intermediaries. So the issue submarined for several years.
- Eventually, seeing no voluntary resolution of the cost sharing issue, the European Parliament passed legislation mandating its use nevertheless, in a resolution amending its Directive 72/306/EEC. [4]

[1] http://www.oasis-open.org/committees/autorepair/

[2] http://www.oasis-open.org/committees/ download.php/2412/Draft %20Committee%20Specification.pdf

[3] See Appendix C to the draft specification.

[4] http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA +P6-TA-2006-0561+0+DOC+XML+V0//EN

RDF under the hood

- The OASIS Auto Repair Information TC draft specification, now made law, relies principally on W3C's RDF.
- The spec also defines and consumes several other namespaces, including some industry specific ones (such as vehicle identification number), and some common concepts from other general schemes (OASIS' UBL for 'price', 'currency', etc.; Dublin Core for resource 'creator', 'title', 'subject' and 'date'; W3C's SWAP Personal Information Markup for personal address data like 'phone', 'address' and 'city'). Spec also permits 'local' namespaces & taxonomies (such as parts catalogs for one OEM).
- Note: Industry-specific data, a well defined user group, and stakeholders who helped design the meaning structure. Arguably, this was a paradigmatic top-down scheme.

OASIS

WSDL, SEE and SOA:

The Post-it Note problem.

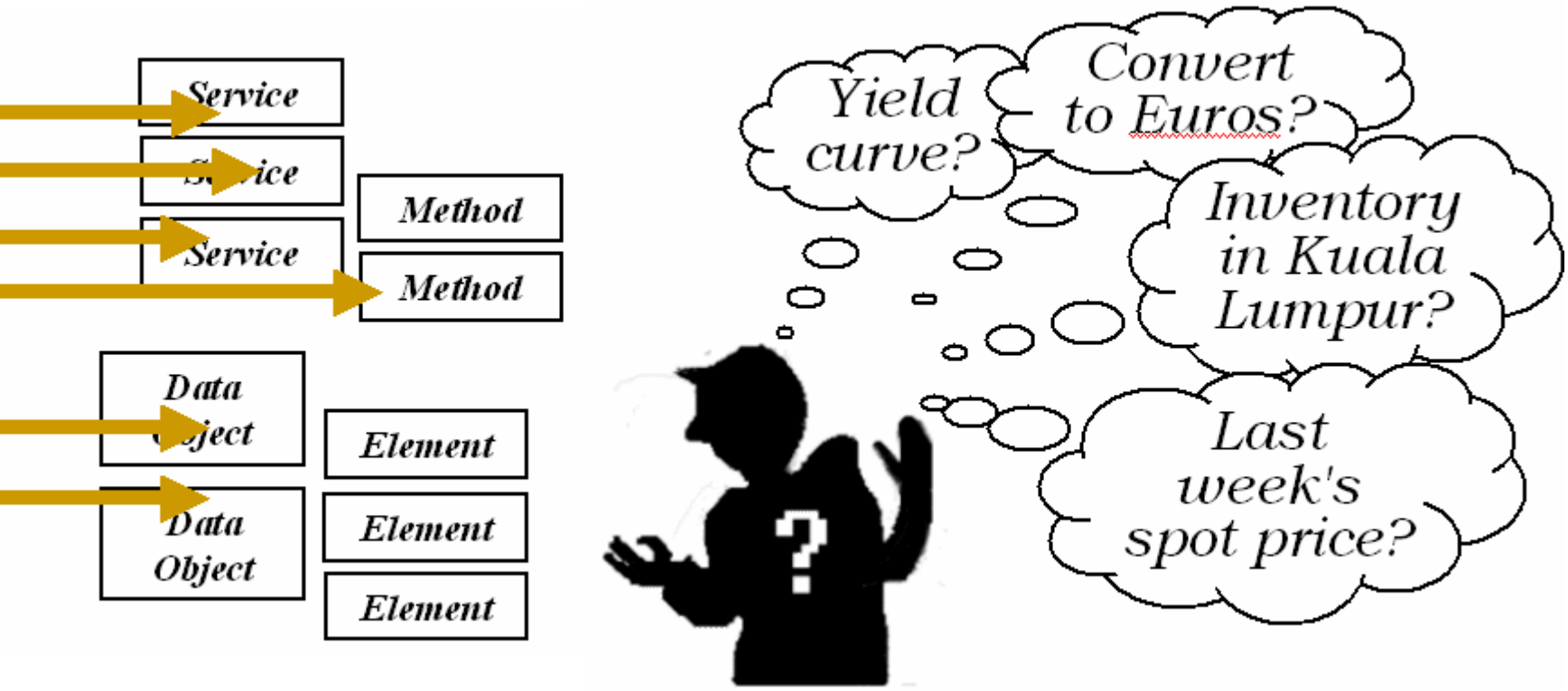
What's in your WSDL?

Beating service swords into plowshares: the Semantic Execution Environment project.

Mash-ups: the speaker reveals his biases.

Your business defines your services and data

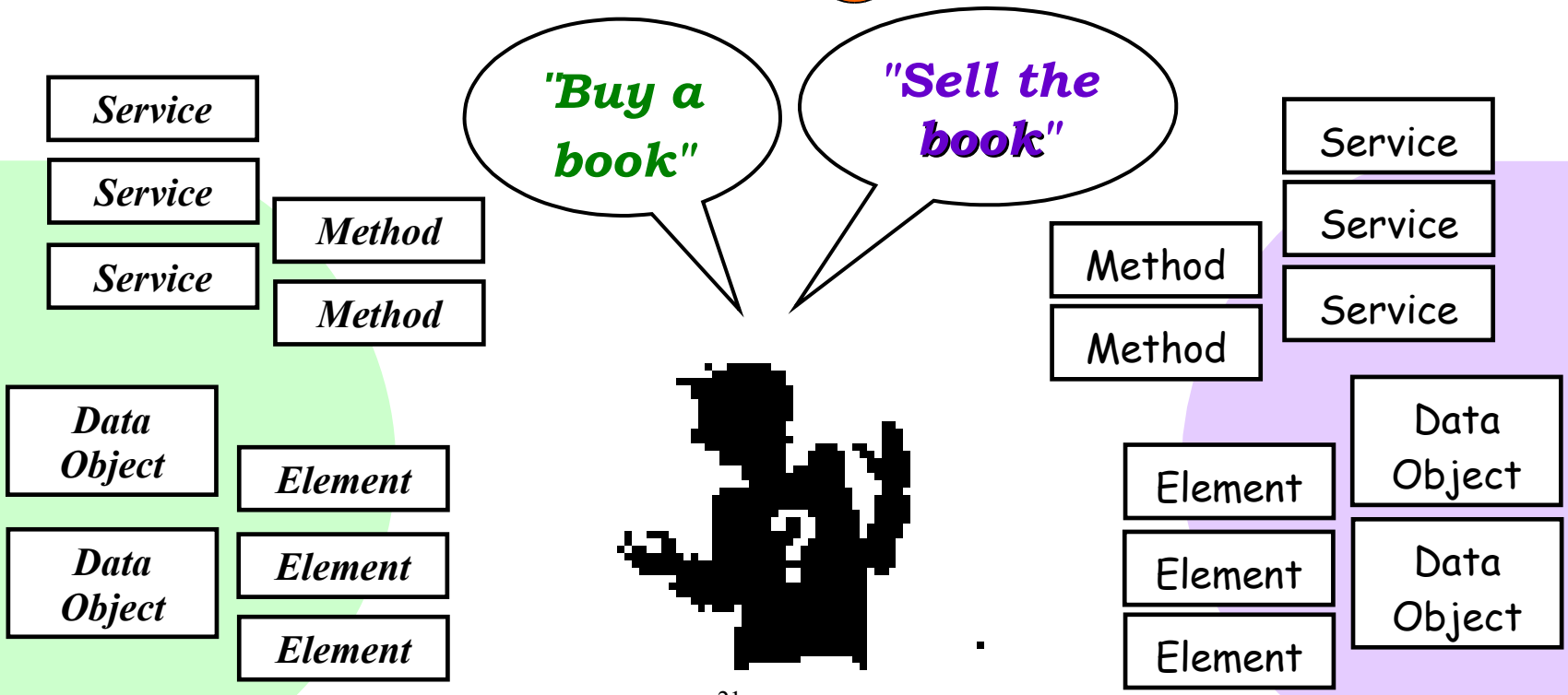
What data do you want to deploy for re-use? With what meaning? What computing functions should, and should not, be available to outside counterparties?



"Heidi" services

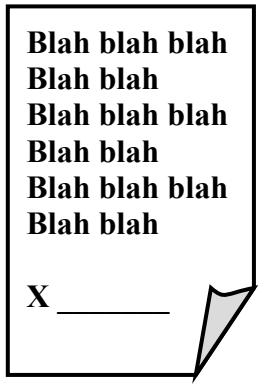
The Banff problem:

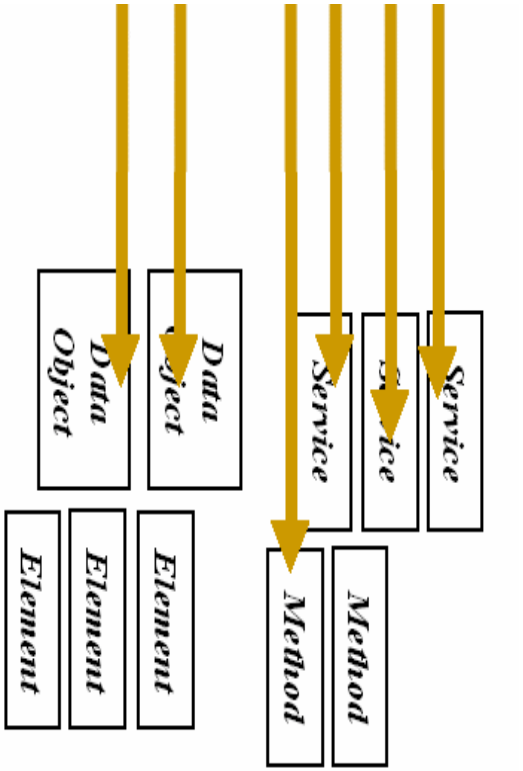
Amazon methods ≠ eBay methods

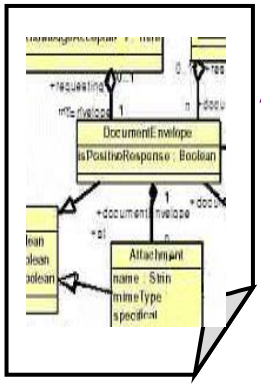


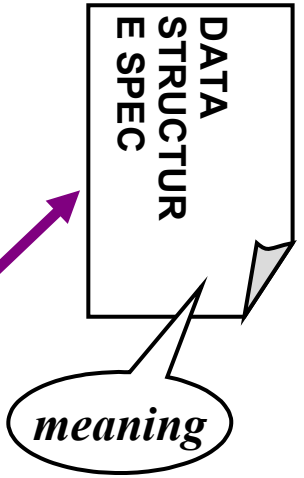
No agreement = no deal

Robust electronic transactional automation: remember who bears the risk of misinterpretation when everyone repudiates and ends up in court









Harmonized data components are expected to resolve data meaning disputes. What will resolve service definition disputes?

- WSDL + WS-Addressing?
- What gets written in the sticky note attached to the service endpoint?
- OAGI's WSDL concept in OAGIS v9:

  http://www.openapplications.org
- Interesting take from the 2005 W3C Workshop on Semantics in Web Services:

  http://www.w3.org/2005/01/ws-swsf-cfp.html
- OASIS SEE TC

OASIS Semantic Execution Environment TC

- One way to provide a structure for semantically meaningful service descriptions
- And, importantly, a mediation method for matching and interpolating

TC pages:

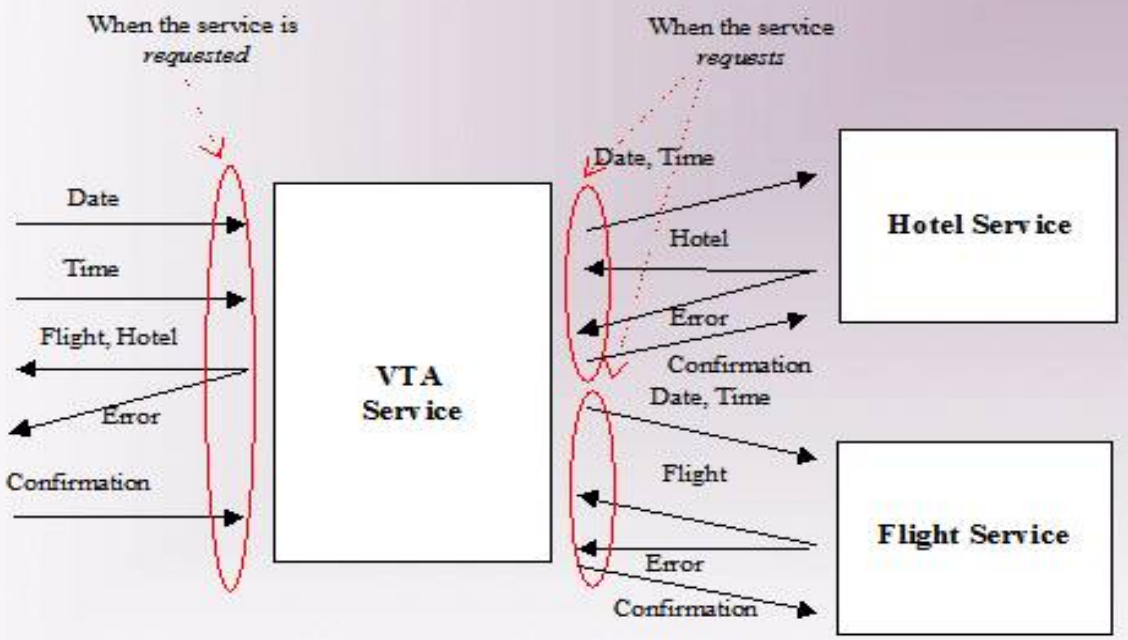
http://www.oasis-open.org/committees/semantic-ex

Background paper:

http://www.oasis-open.org/committees/download.php/25706/ SEE-background-and-related-work_11.doc

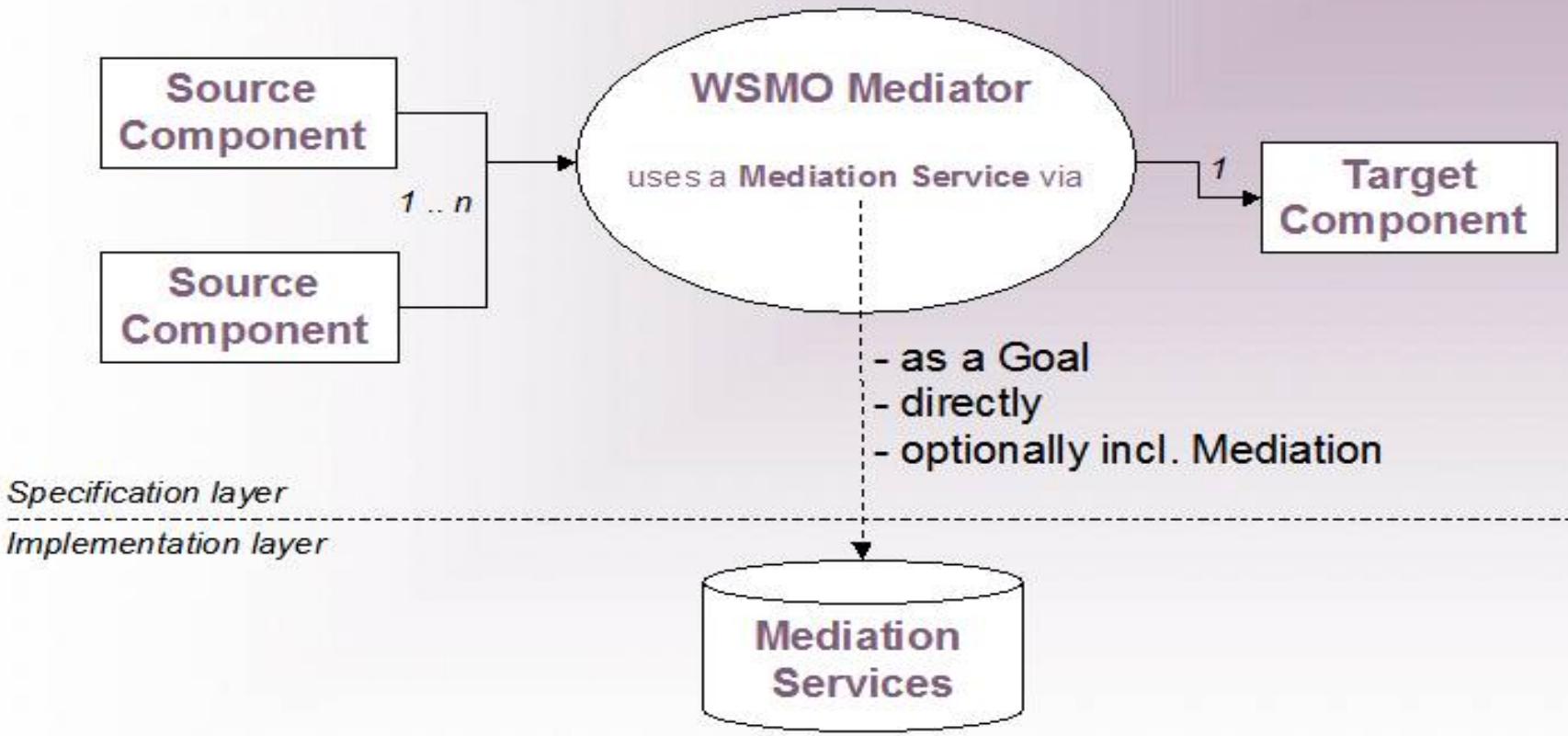
Choreography & Orchestration

VTA example:



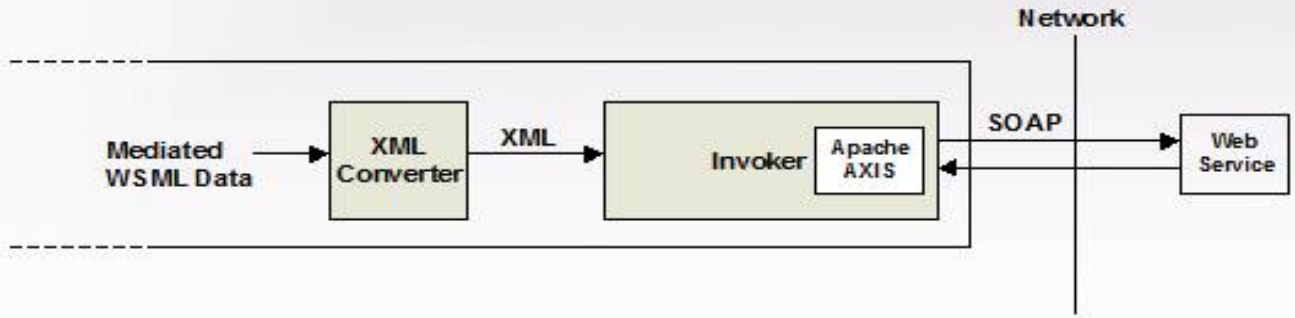
- **Choreography** = how to interact with the service to consume its functionality
- **Orchestration** = how service functionality is achieved by aggregating other Web services

Mediator Structure

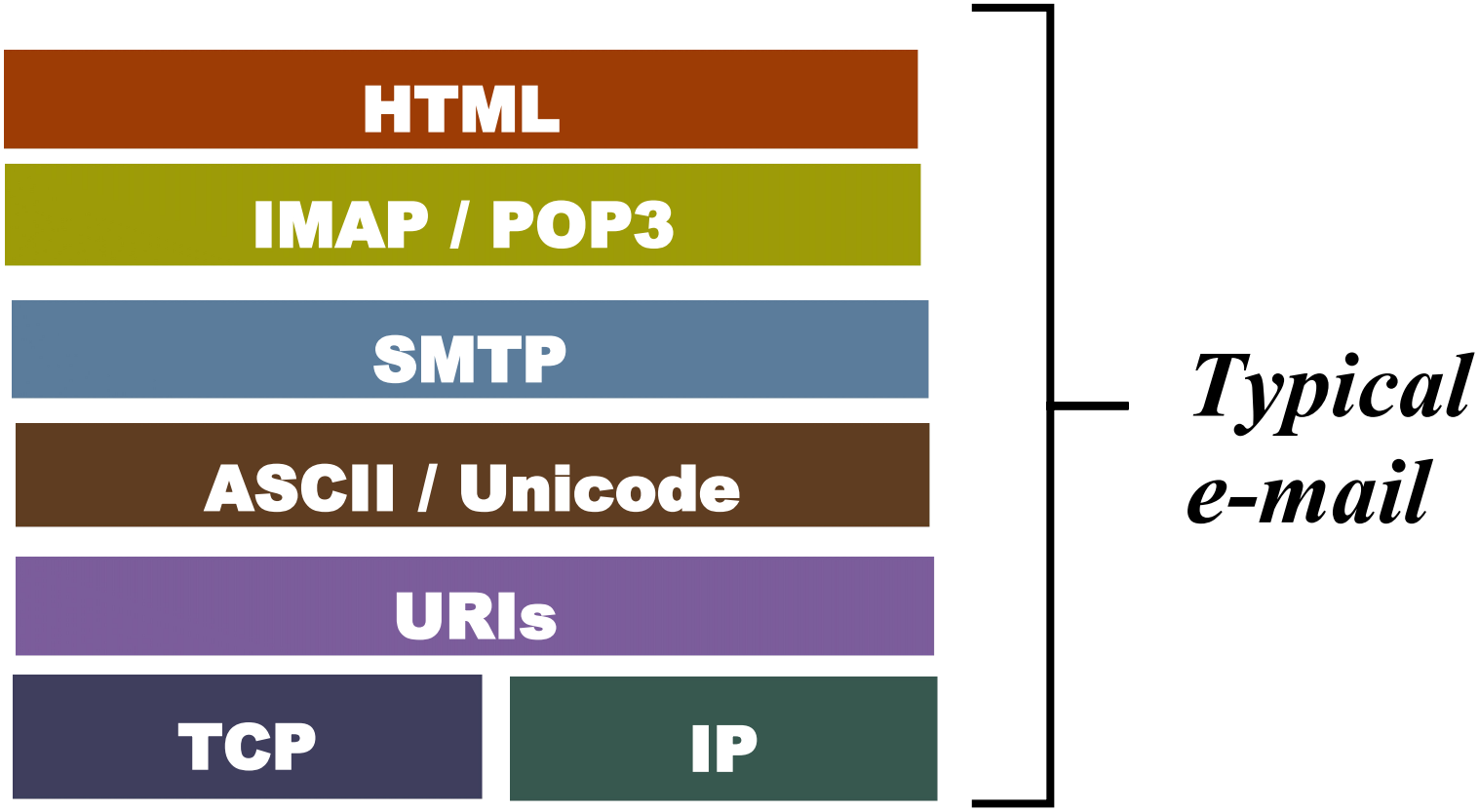


Communication Manager – Invoker

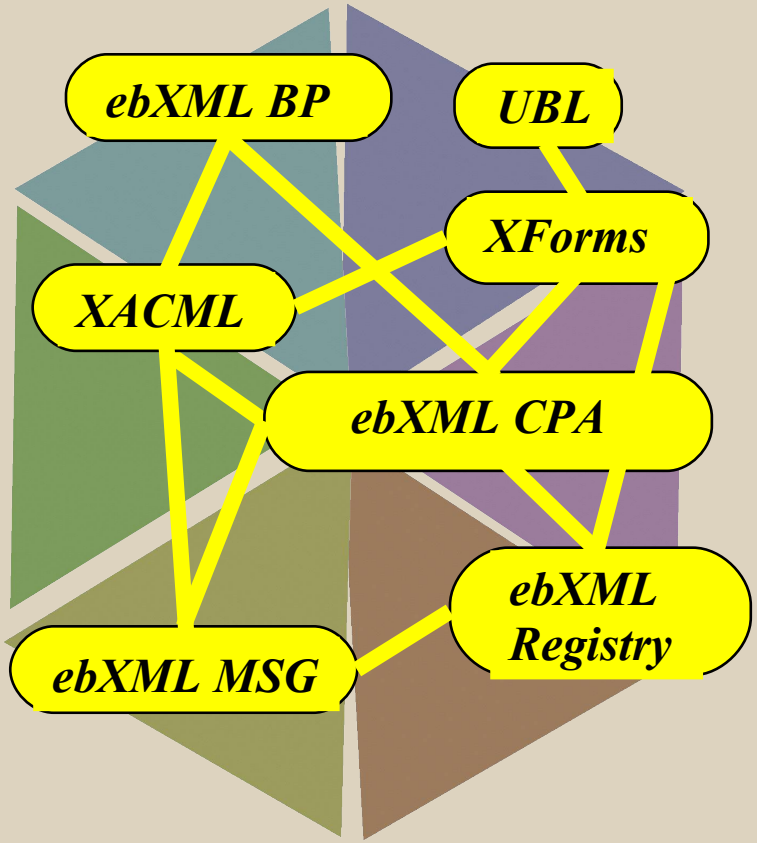
- WSMX uses
  - The SOAP implementation from Apache AXIS
  - The Apache Web Service Invocation Framework (WSIF)
- WSMO service descriptions are grounded to WSDL
- Both RPC and Document style invocations possible
- Input parameters for the Web Services are translated from WSML to XML using an additional XML Converter component.



Mashups: most real-world installations are composed of multiple standards

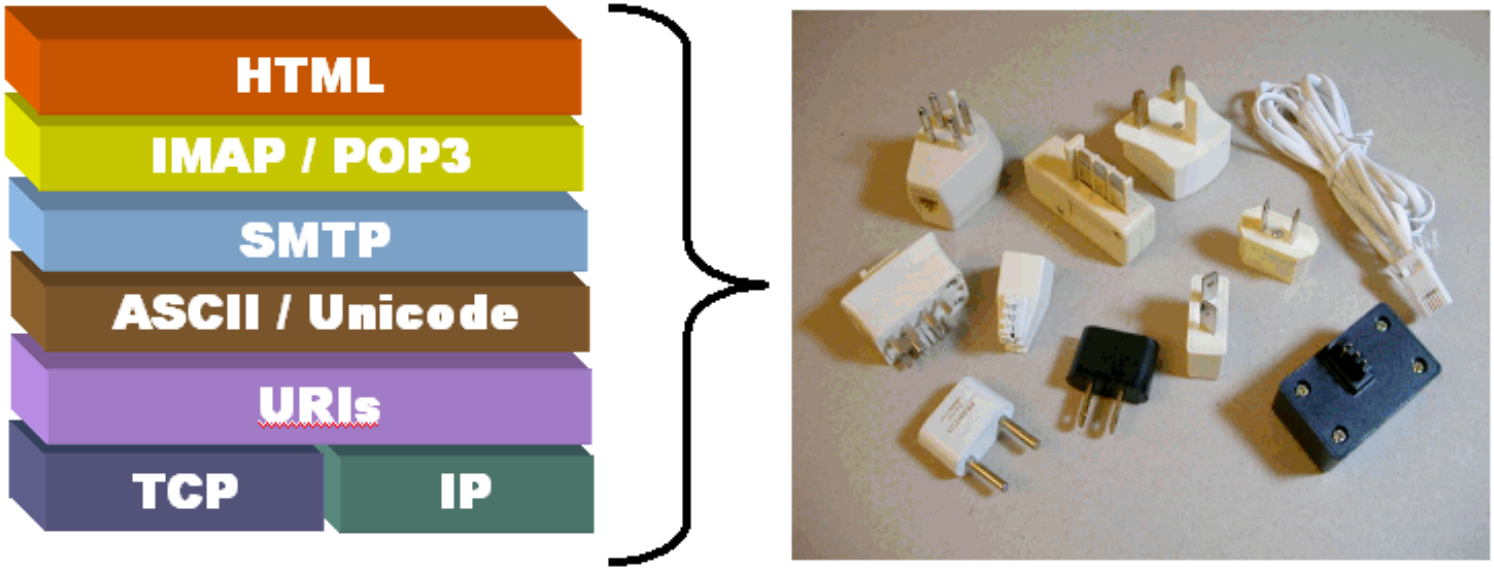


Real-world installations are composed of multiple standards



Example: The OASIS Disease Control Interoperability Demo at XML 2003

Mash-ups: Interoperability requires flexibility



You are going to plug lots of pieces together. Modularity good. Conformance good. Alternatives good. Exclusive dependencies bad.

OASIS

Meaning Lite: the Common Alerting Protocol

Sometimes the amount of interoperability needed is high … and the amount of information needed is really, really small

Extensions to loosely organized data: the OASIS Common Alerting Protocol

- Lack of technical interoperability has been one of the most challenging aspects of emergency and incident management, transmitting notices and assistance in catastrophic weather, hazard or security conditions.
- Historically, siloed and disparate communication systems, often can't intercommunicate, or even share a single message. Teams often form *ad hoc* across dep'ts.
- The OASIS Standard "Common Alerting Protocol" (CAP) was developed by OASIS' Emergency Management Technical Committee [5] to enable public warning information exchange over a wide variety of data networks and systems.
- CAP specifies a common, very light, XML-based data structure for warning messages.

Extensions to loosely organized data: the OASIS Common Alerting Protocol

CAP remains simple, so as to remain fully compatible with existing heterogeneous legacy public warning systems, legacy data structures, and multiple transport methods. The document model is composed of a few simple categories of metadata that practically any system can parse:

- an <alert> element, containing basic message identifying data such as time-stamping, recipients, and containers to pass other implementation-specific instructions;
- <info> elements to contain the core details about the alert event (such as category, urgency, severity, source, event codes, etc.);
- <resource> elements to contain pointers & descriptions (or serializations) of relevant data sources: images, audio, etc.;
- <area> elements to specify geographic application of the alert data, using a specified geospatial reference systems. (GML)

Under the hood of CAP: more tolerance than taxonomy

- **CAP v1.0 was approved as an OASIS Standard in May of 2004, and implemented by US NOAA (weather reporting) and USGS (earthquake, volcanic and landslide events). CAP v1.1 added several functions, and after final approval at OASIS, was cross-contributed to ITU-T for a joint workshop in 2006 [6] and obtained global approval as ITU Recommendation x.1303 in 2007 [7].**

[5] http://www.oasis-open.org/committees/emergency/

[6] http://www.oasis-open.org/committees/download.php/15135/ emergency-CAPv1.1-Corrected_DOM.pdf

[7] http://www.oasis-open.org/events/ITU-T-OASISWorkshop2006/ proceedings.php, and http://www.itu.int/ ITU-T/worksem/ictspw/index.html

Under the hood of CAP: more tolerance than taxonomy

- A multiplicity of responders to a hazardous waste emergency (for example) may have different needs. It's essential that all of them readily can parse the basics of a warning message. But their more detailed data need may diverge.
- The "ground truth" native form of disaster event data often has more value than a transformed version.
- A highly heterogenous base of users and necessarily interoperable systems … thus, arguably, an extreme case of tolerant, bottom-up systems and low levels of meaning mark-up constraints.

Identifier Purgatory

A brief observation and lament

The good Lord must have liked identifier schemes …
he made so many of them.

(Apologies to A. Lincoln)

- **UPC/ UCC/ Extended codes: See GS1**
- **XRI / XDI: See OASIS**
- **UUIDs: See ISO/IEC JTC1**
- **URNs: See IETF**
- **ASN1: See ITU-T**
- **UDEF: See the Open Group**
- **and so on**

Not a semantics problem per se, but still an obstacle to widespread standardization of data exchanges. Also, this issue shares the governance issues of many KR systems. Who assigns IDs? What assures accuracy and uniqueness? Who pays what to who?

Alice the Core Component in SemanticLand

A brief re-cap.
A thousand flowers blooming.
How to get them in one garden?
What role for semantics & KR?
Bottom-up versus top-down, redux.

ebXML Core Components (1999) > ISO TS 15000-5 (2004) > Multiple implementations > UN/CEFACT harmonization

- The Core Components project and its history is well known to Ontolog participants generally, so it's not re-capped here.
- The CCTS methodology, based on ISO/IEC TS 11179's scheme, contemplated robust contributions of production-proven data, followed by harmonization.
- Arguably the timeline was somewhat reversed. A number of implementer communities developed their own component sets using the methodology first: OASIS UBL, SWIFT's ISO TS 20022, OAGI's BODs &c.
- Then, beginning in mid-2006, they're contributed and weighed against early incumbent draft material.

ebXML Core Components (1999) > ISO TS 15000-5 (2004) > Multiple implementations > UN/CEFACT harmonization

- To some degree the 'harmonization' process is juried; some see this as positive, some as FIFO, others as more discretionary than deterministic. Probably it's too early to judge.
- Application of KR methods to CC seems mostly to come from the outside. Several EU companies have suggested applying RDF to CEFACT CC material; Prof. Dogac's Ontolog presentation last week (6 March 2008) demonstrates ontological management of CC data.
- What does this tell us about top-down versus bottom-up? How diverse are the stakeholders in the case of the contents of (e.g.) an invoice? How much acceptance of semantic methodology can we expect?

James Bryce Clark *(jamie.clark@oasis-open.org)* is the Director of Standards Development at OASIS. He joined the consortium's professional staff in 2003, and is responsible for supervising the operations and output of OASIS's 60+ technical committees.

Jamie chaired the Electronic Commerce Subcommittee of the American Bar Association's business law section for four years. He has served as the U.S. State Department's expert delegate in the e-commerce working group of the United Nations Commission on International Trade Law (UNCITRAL), working on treaties and global model laws for electronic commerce, since 2000.

Jamie started practicing law with Shearman & Sterling on Wall Street in New York City in 1988. Before joining OASIS, he was general counsel to a healthcare e-commerce company, and a partner in a Los Angeles corporate law firm, handling finance and regulatory issues. He has been an active member and editor of business process standardization committees, himself, and chaired the global ebXML joint coordinating committee in 2001-02. He holds JD and BSc degrees from the University of Minnesota.