

Meeting the Twin Challenges of Open Data for DATA Act compliance and Delivering next generation Industry Services

David Webber

August 2014

DATA Act 2014 (Financial Reporting)

Digital Accountability and Transparency Act

DTC Data Transparency Coalition

Home | About | What is Data Transparency? | News | Issues | Events | Data Transparency in Action | Blog | Contact

Holding Government Accountable.

If more federal data were published online in machine-readable formats, citizens and watchdog groups could keep closer tabs on what their government is doing.

Federal Data Reform

The Data Transparency Coalition advocates on behalf of the private sector and the public interest for the publication of government information as standardized, machine-readable data. Data transparency strengthens democratic accountability, enhances government management, reduces compliance costs, and stimulates innovation.

Federal data reform starts with the Digital Accountability and Transparency Act ([DATA Act](#)), which will open the government's spending information to illuminate waste and fraud. But it won't end there. [Other types of federal information need reform, too.](#)

Use our [contact form](#) to get in touch about joining the Coalition or for any general inquiries you may have. Click [here](#) to learn more about our various membership options.

Current Campaign

SEC [open data enforcement](#) began on July 7.

The third [Data Transparency Breakfast](#), presented by PwC, will be held on July 29. Our topic: [Transforming Government Reporting Around the World](#)

On May 9, President Obama [signed](#) the DATA Act into law. Congress [passed the DATA Act](#) unanimously through both chambers in April 2014.

Follow us on our [blog](#), [Twitter](#) and [Facebook](#) for all the [latest news](#) and information.

Advocating Open Data

DATA Act and federal data reform

U.S. DATA REFORM
SPENDING \$762 B
TRACKING \$762 B

0:00 / 2:48 YouTube

Questions:

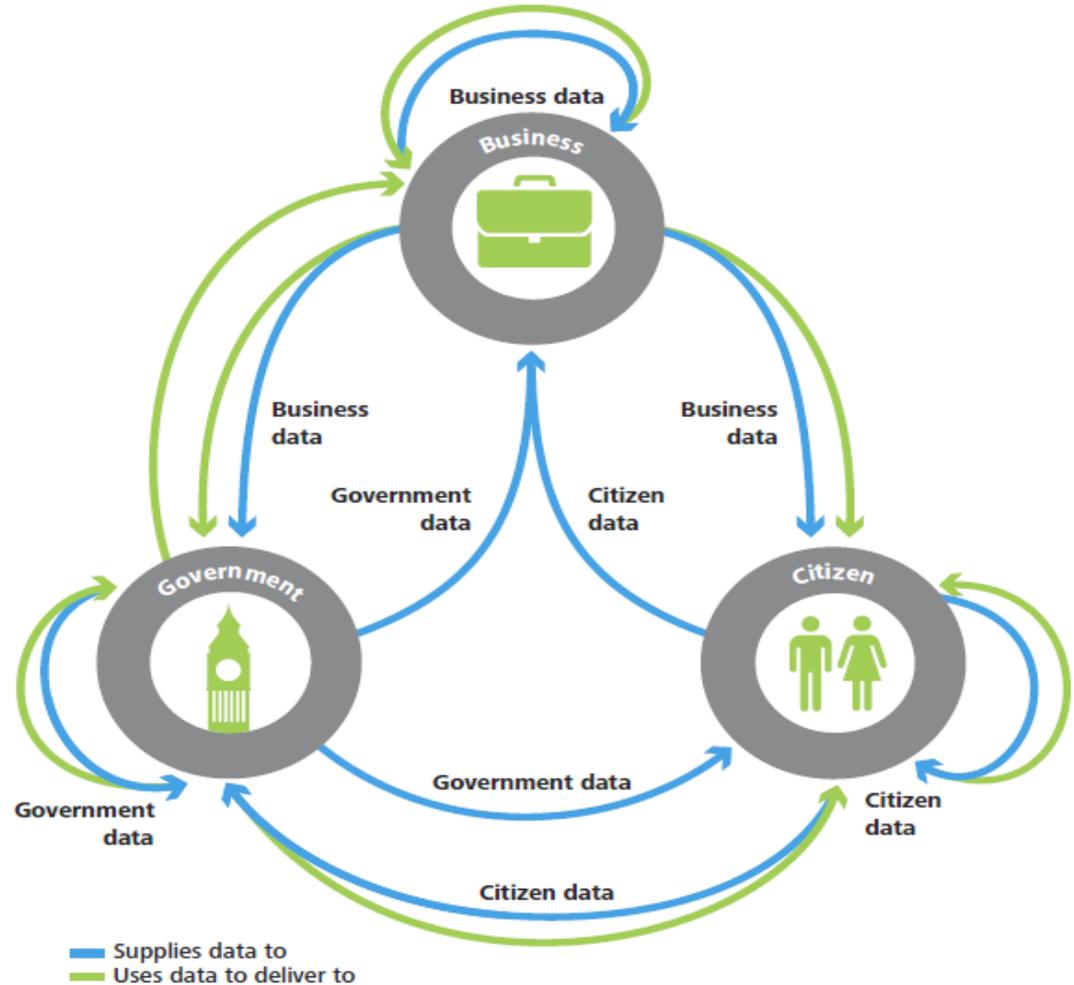
What data standards will be selected / developed?

How will government publish data (where, when, who)?

How will coordination work to share resources and technology?

The Open Data Ecosystem

Challenge:
*managing the
Who, What,
Why, Where and
How*



Source: Deloitte LLP

Delivering Next Generation Industry Services

- **Open Data Vision**
 - Citizen-centric Services Delivery
 - Tourism and Travel facilitation
 - Industry Services facilitation
 - Business Opportunity Enabling
 - Emergency Services
 - Disaster Recovery
 - Government Transparency
 - Government Participation
- **Open Data Semantics and Vocabulary**
 - Alignment at Municipal, State and Federal and International levels
 - Mobile application enabling

Existing Coordination Work

- NIEM for Federal, State, Local, Tribal information exchanges around JPS (Justice and Public Safety)
- Open Data initiatives – Municipal 311
- EU Joinup initiative
- Data.gov
- Schema.org

Of tadpoles and data...

- Tadpole – large fat head – long thin tail
- EDI taught us that 95% of exchange content use just 5% of the components
- For EDI 50% of content is in code values
- Once implemented 5% of the data cause 90% of the edits and changes
- Domain specialization components account for the 95% “long thin tail”
- Lessons learned - information and processes coalesce around core content and exchanges; most of ongoing components definition and maintenance work are on edge conditions and extensions.

Financial Reporting - XBRL

- Pros
 - Mature existing implementations
 - Widely available
 - Well understood by accounting professionals
- Cons
 - Consortium specification
 - Complexity
 - Expense
 - Public accessibility
 - Requires special XBRL-aware software

National Information Exchange Model (NIEM)

- Pros
 - Has concept of common “core components”
 - Widely adopted for JPS applications
 - Mature methods and techniques
 - Well aligned to government needs
- Cons
 - Based on XSD schema technologies
 - Learning curve is significant
 - Complexity of components library
 - Narrowly focused on JPS applications
 - Hardwired Context and Role
 - Not really a model; schema objects collections

OASIS ebXML / UBL

- Pros
 - Core Components conceptual design and abstract representation (CCTS)
 - First to formalize concepts of Context and Role
 - Registry/Repository technology
 - Business Process technology (BPSS)
 - Aligned with ISO 11179 work
- Cons
 - Not widely adopted
 - UBL suffers same complexity issues as NIEM and limitations of W3C XSD schema

Education Industry Consortium - PESC

- Pros
 - Mature adoption in Higher Education community
 - Exchange schema well aligned with college business processes
 - Core Components approach for schema
 - Consistent Naming and Design Rules
- Cons
 - Alignment with NIEM components for similar entities
 - Consistency of implementation approach
 - Leveraging of development tools
 - Lack of dictionary definitions

Supporting New Technologies / Platforms

- Mobile devices
- JSON / RESTful APIs
- NoSQL repositories
- Open Data
- Folksonomies (socially grown)
- Linked Data
- International delivery

Moving Forward: Semantic Interoperability

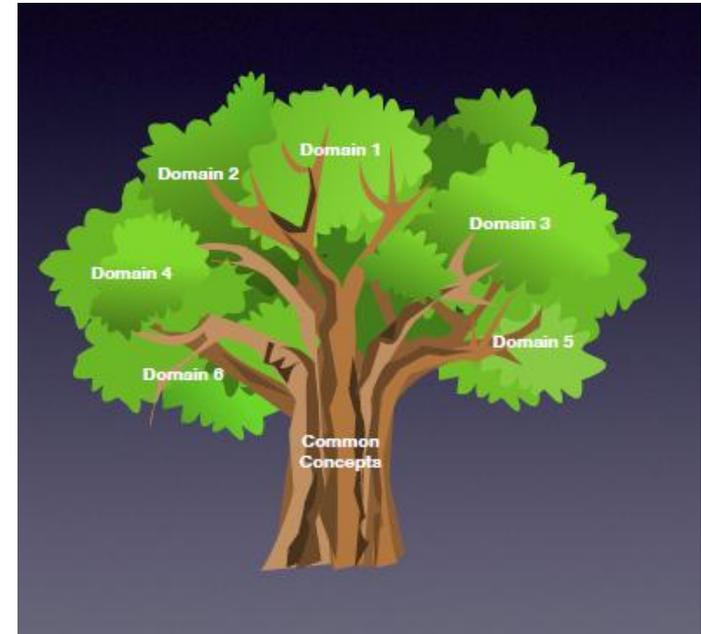
Next Generation Approach

XML and Information Semantics

- Replaced EDI separators with “tags”
- Tags are “human readable” but not always “human understandable!”
- No formal mechanism to support the declaration of semantic integrity constraints.
- No means of validating object semantics even if these are declared formally.
- XML thus formally governs syntax only

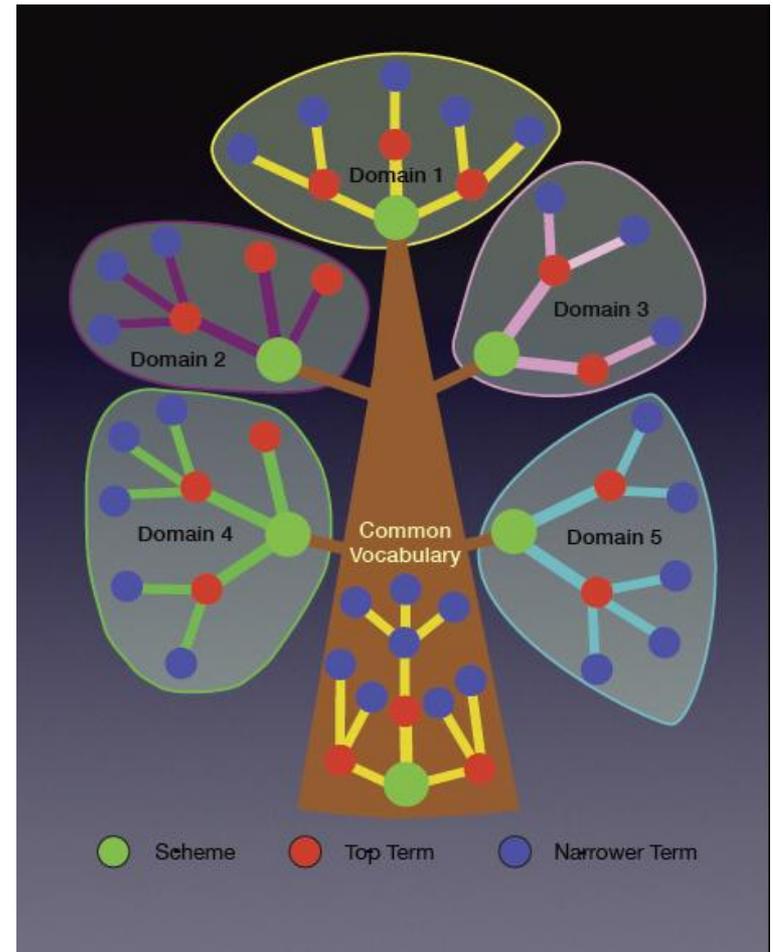
Toward Semantic Interoperability

- No changes to domain specific names and definitions
- Create a new central system with:
 - Multi-lingual shared common vocabulary and concepts expressed in “everyday” natural language
 - Semantic hierarchical and nonhierarchical relations to other concepts
 - Collections, set of concepts that have something in common
 - Concept mappings back to the domain specific concepts indicating the level of equality matching

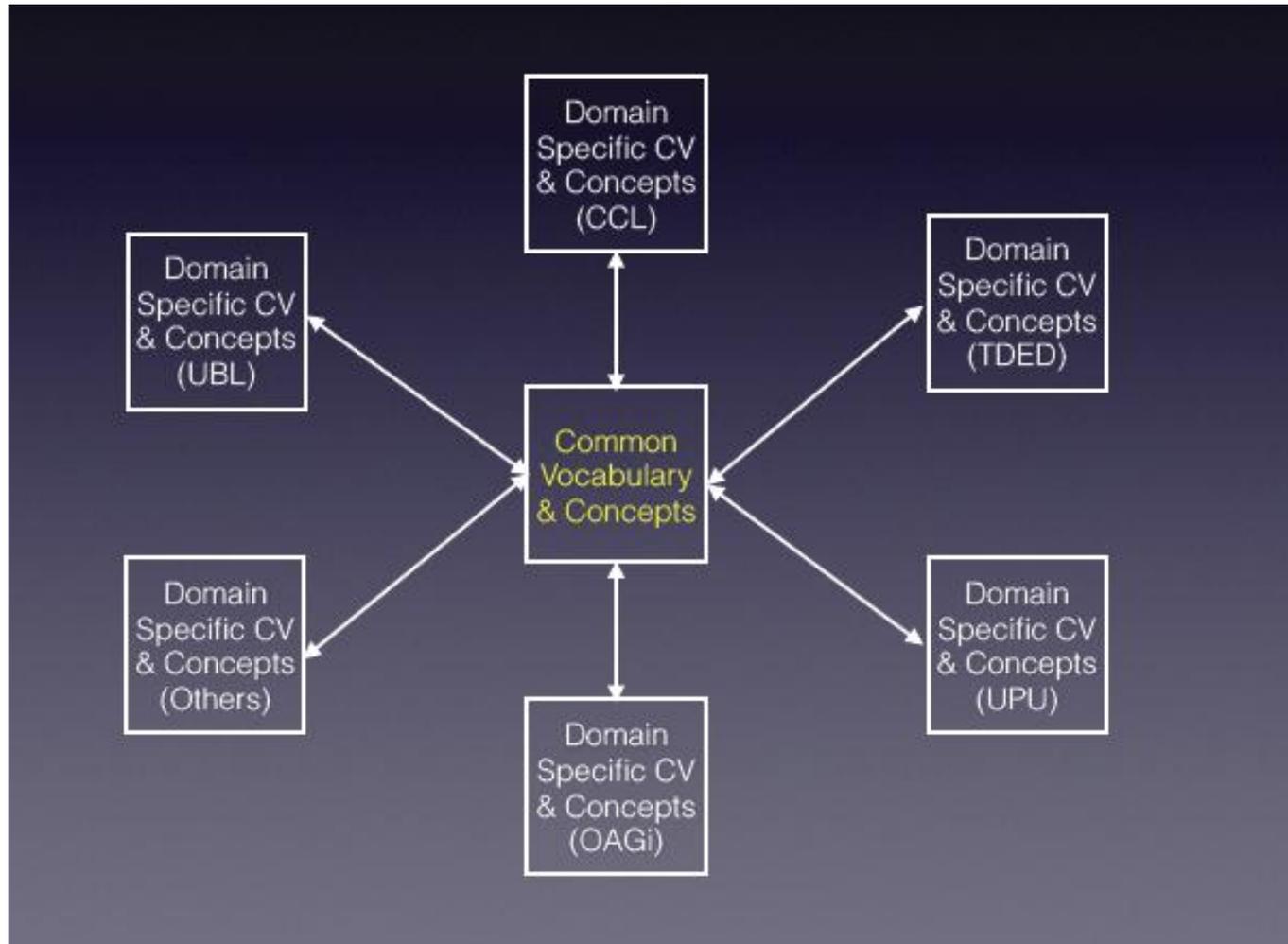


Semantic Concepts Tree

- Top Terms
- Narrower Terms
- Conceptual Terms
- Scheme and Components



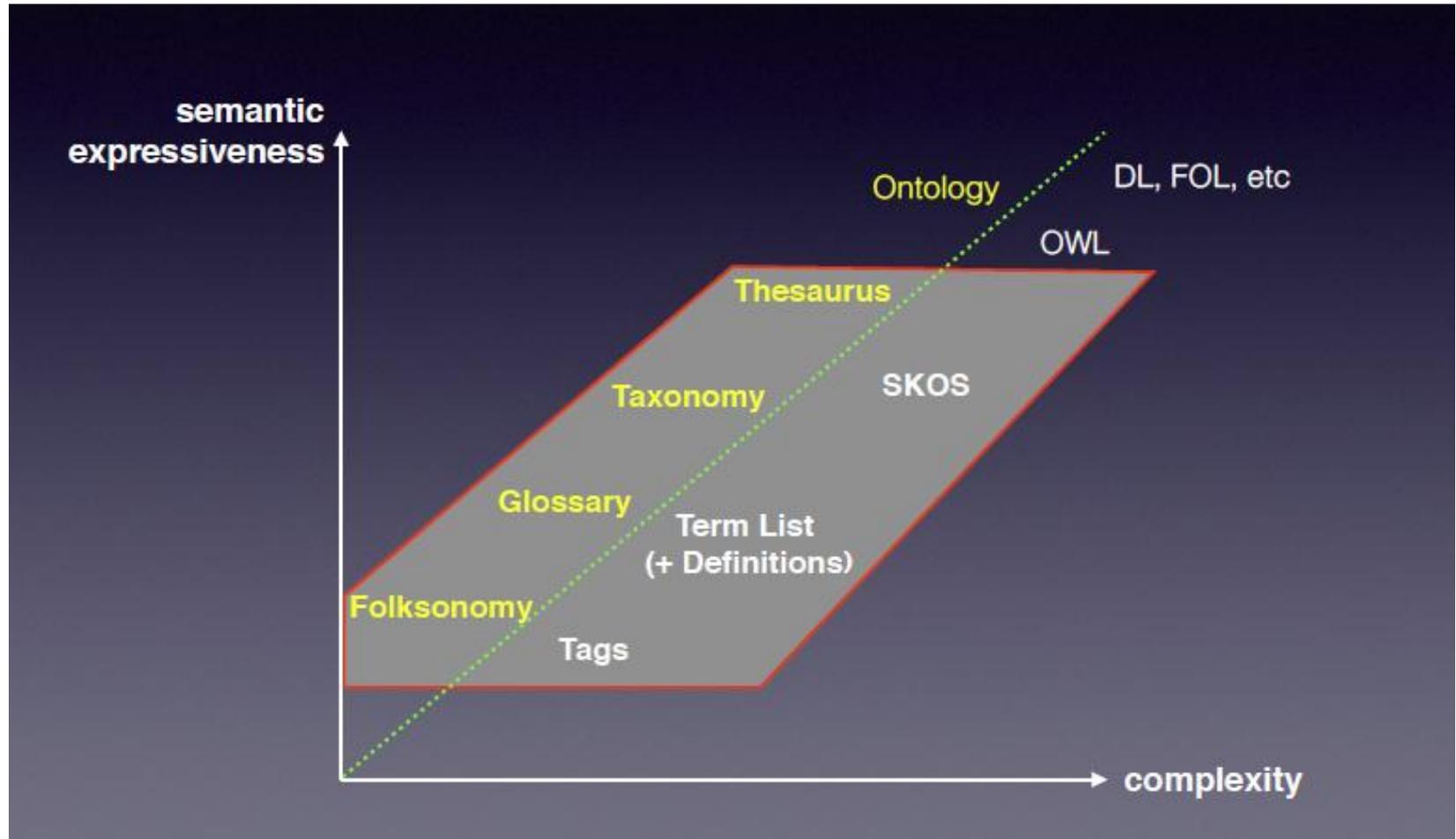
Linking Concepts using Web and SKOS



What is SKOS?

- *Simple Knowledge Organization System*
- A model for expressing the basic structure and content of concept schemes such as **thesauri**, **classification schemes**, **taxonomies**, **folksonomies**, and other similar types of controlled vocabularies
- W3C specification
- Available as open source implementation - iQvoc

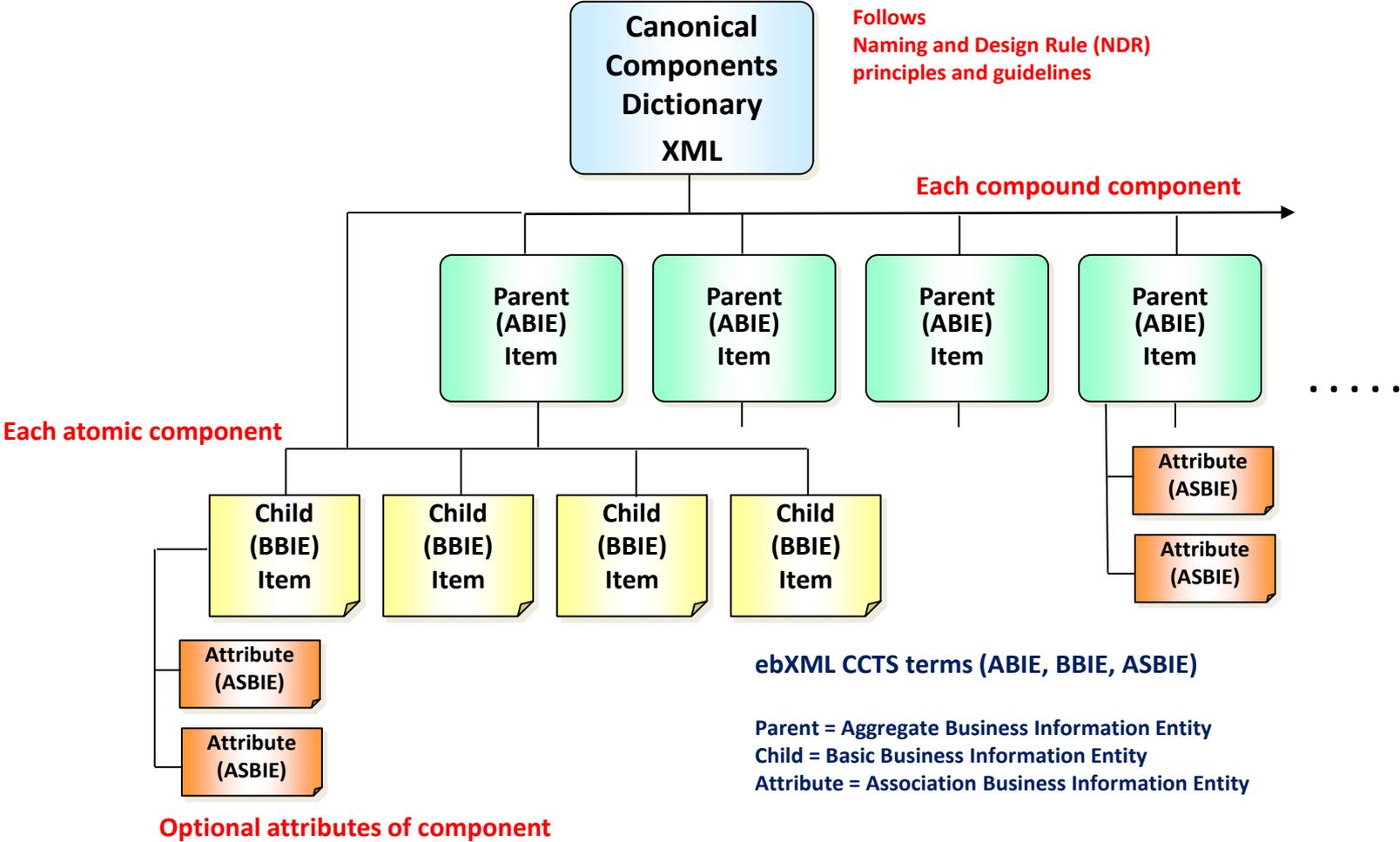
Knowledge Organization Systems (KOS)



ebXML CCTS Dictionary

- XML-4-CCTS schema for representing core components with XML
- OASIS Content Assembly Mechanism (CAM) dictionaries
 - Subset of XML-4-CCTS using 8 key components
 - CAM dictionaries compatible with:
 - Eclipse tree viewer tool
 - Excel spreadsheets
 - MindMap viewers
 - UML/XMI physical models

Conceptual Information Model



* CCTS – Core Components Technical Specification

Putting all the pieces together

SKOS, CCTS, Dictionaries, NIEM,
Open Data, FOSS, XML, JSON

Conceptual Vocabulary Structure

Core Vocabulary

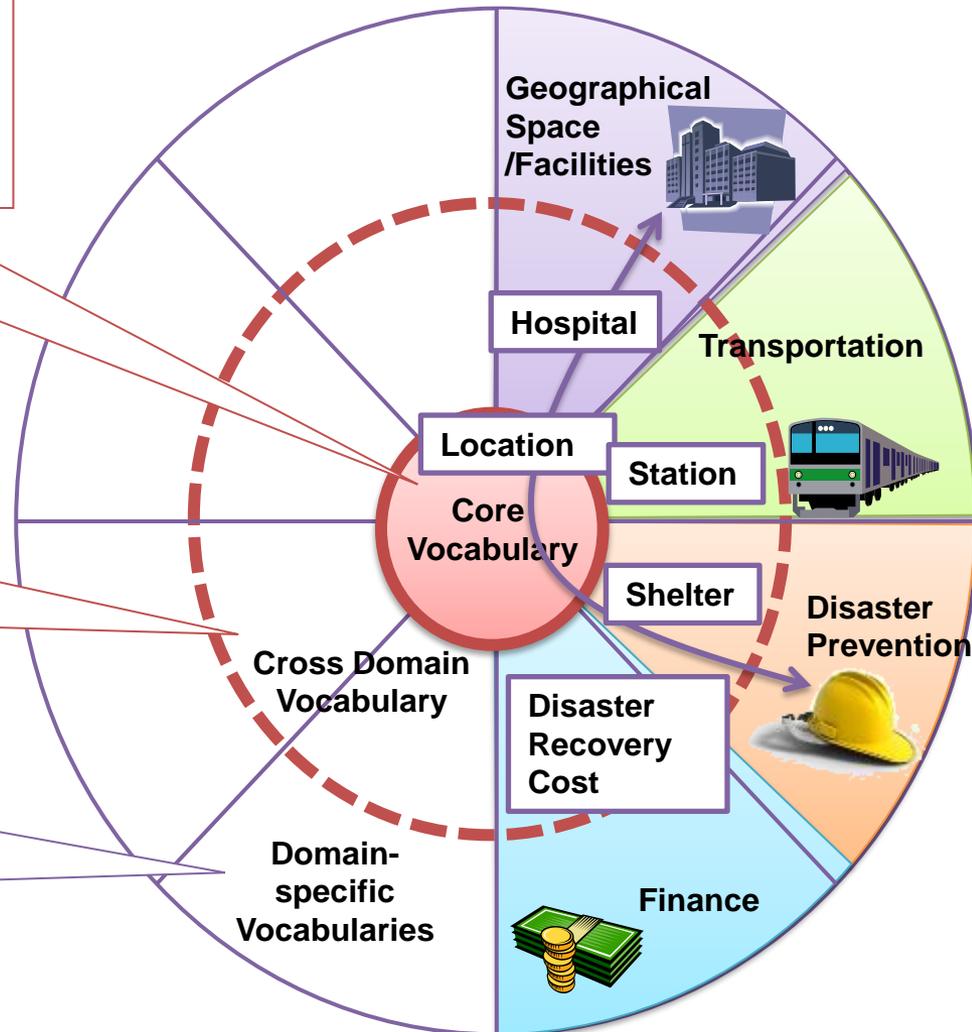
Universal vocabularies that are widely used in any domain (top terms).
E.g. person, address, place, date.

Cross Domain Vocabulary

Key vocabularies among domain-specific vocabularies that are referenced in other domains.
E.g. hospital, station, shelter.

Domain-specific Vocabularies

Vocabularies that are specialised for the use in each domain.
E.g. Number of beds, Schedule.



Vocabulary Example

■ Top Terms

Person

Location

Organization

Facility

Person Name

Address

Organization

Facility

Person

Spot

Name

Building

Equipment

Core vocabulary

Name	Name (English)	Identifier (English)	Data Type	Cardinality	Description
氏名	Name	PersonName	string	0..1	個人を識別する1つのデータ型
姓	Surname	PersonSurname	string	0..1	個人を識別する1つのデータ型
姓(ローマ字)	Surname (Roman)	PersonSurname (Roman)	string	0..1	個人を識別する1つのデータ型
名	GivenName	PersonGivenName	string	0..1	個人を識別する1つのデータ型
名(ローマ字)	GivenName (Roman)	PersonGivenName (Roman)	string	0..1	個人を識別する1つのデータ型
性別	Sex	PersonSex	enum	0..1	個人を識別する1つのデータ型
性別(ローマ字)	Sex (Roman)	PersonSex (Roman)	enum	0..1	個人を識別する1つのデータ型
生年月日	BirthDate	PersonBirthDate	date	0..1	個人を識別する1つのデータ型
生年月日(ローマ字)	BirthDate (Roman)	PersonBirthDate (Roman)	date	0..1	個人を識別する1つのデータ型
住所	Address	PersonAddress	string	0..1	個人を識別する1つのデータ型
住所(ローマ字)	Address (Roman)	PersonAddress (Roman)	string	0..1	個人を識別する1つのデータ型
連絡先	Contact	PersonContact	string	0..1	個人を識別する1つのデータ型
連絡先(ローマ字)	Contact (Roman)	PersonContact (Roman)	string	0..1	個人を識別する1つのデータ型

Name	Name (English)	Identifier (English)	Data Type	Cardinality	Description
住所	Address	Address	string	0..1	住所を識別する1つのデータ型
住所(ローマ字)	Address (Roman)	Address (Roman)	string	0..1	住所を識別する1つのデータ型
郵便番号	PostalCode	AddressPostalCode	string	0..1	郵便番号を識別する1つのデータ型
郵便番号(ローマ字)	PostalCode (Roman)	AddressPostalCode (Roman)	string	0..1	郵便番号を識別する1つのデータ型
郵便局	PostOffice	AddressPostOffice	string	0..1	郵便局を識別する1つのデータ型
郵便局(ローマ字)	PostOffice (Roman)	AddressPostOffice (Roman)	string	0..1	郵便局を識別する1つのデータ型
国	Country	AddressCountry	enum	0..1	国を識別する1つのデータ型
国(ローマ字)	Country (Roman)	AddressCountry (Roman)	enum	0..1	国を識別する1つのデータ型
都道府県	Prefecture	AddressPrefecture	enum	0..1	都道府県を識別する1つのデータ型
都道府県(ローマ字)	Prefecture (Roman)	AddressPrefecture (Roman)	enum	0..1	都道府県を識別する1つのデータ型
市区町村	City	AddressCity	enum	0..1	市区町村を識別する1つのデータ型
市区町村(ローマ字)	City (Roman)	AddressCity (Roman)	enum	0..1	市区町村を識別する1つのデータ型
町名	AddressArea	AddressArea	string	0..1	町名を識別する1つのデータ型
町名(ローマ字)	AddressArea (Roman)	AddressArea (Roman)	string	0..1	町名を識別する1つのデータ型
丁目	AddressBlock	AddressBlock	string	0..1	丁目を識別する1つのデータ型
丁目(ローマ字)	AddressBlock (Roman)	AddressBlock (Roman)	string	0..1	丁目を識別する1つのデータ型
建物	HouseNumber	AddressHouseNumber	string	0..1	建物を識別する1つのデータ型
建物(ローマ字)	HouseNumber (Roman)	AddressHouseNumber (Roman)	string	0..1	建物を識別する1つのデータ型
ビル名	LocationBuilding	LocationBuilding	string	0..1	ビル名を識別する1つのデータ型
ビル名(ローマ字)	LocationBuilding (Roman)	LocationBuilding (Roman)	string	0..1	ビル名を識別する1つのデータ型
建物番号	LocationRoomNumber	LocationRoomNumber	string	0..1	建物番号を識別する1つのデータ型
建物番号(ローマ字)	LocationRoomNumber (Roman)	LocationRoomNumber (Roman)	string	0..1	建物番号を識別する1つのデータ型
ID	AddressIdent	AddressIdent	string	0..1	IDを識別する1つのデータ型
ID(ローマ字)	AddressIdent (Roman)	AddressIdent (Roman)	string	0..1	IDを識別する1つのデータ型

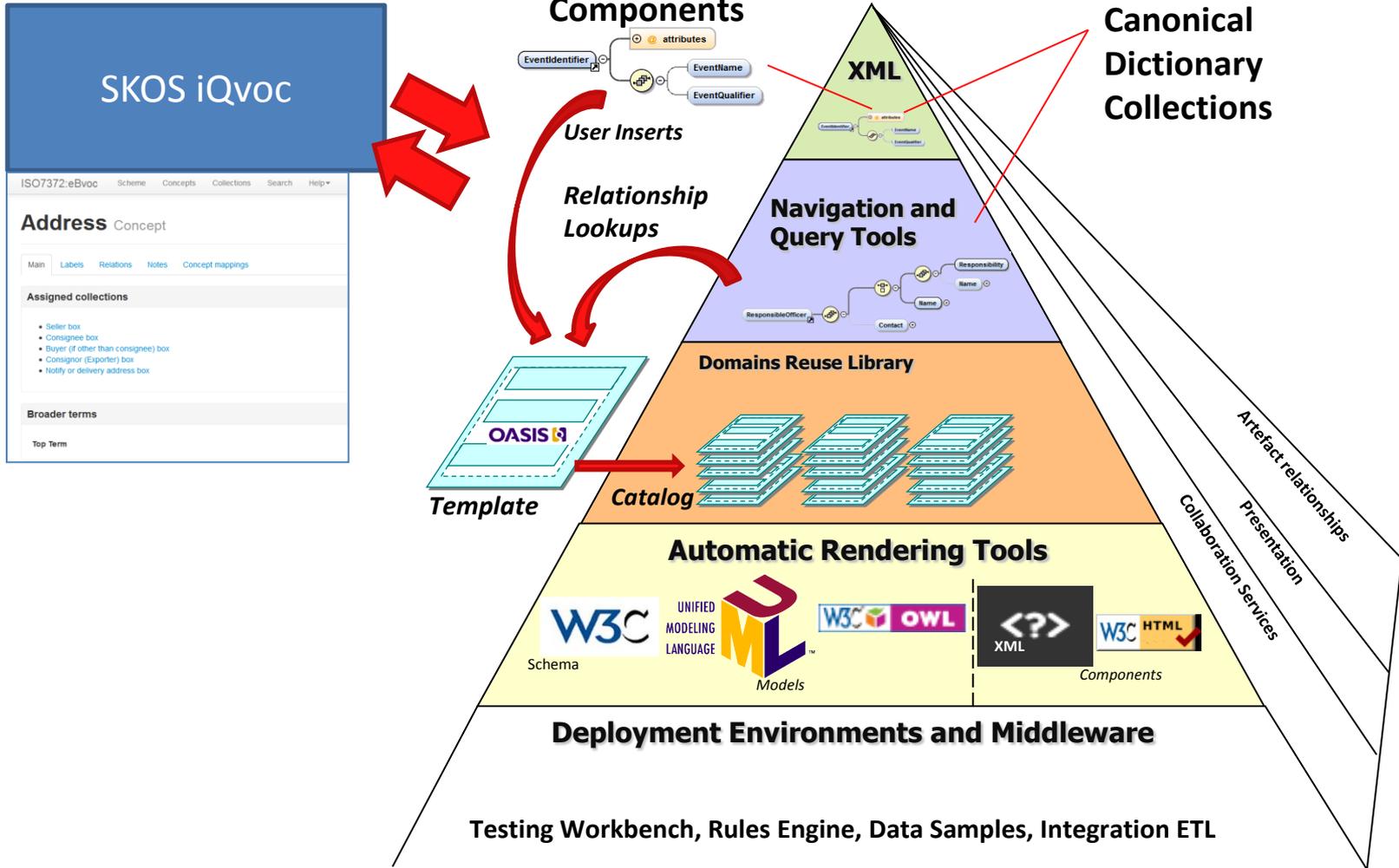
Name	Name (English)	Identifier (English)	Data Type	Cardinality	Description
連絡先 (連絡)	ContactInformation	ContactInformation	string	0..1	連絡先を識別する1つのデータ型
連絡先	ContactMeans	ContactInformationContactMeans	string	0..1	連絡先の種類
名前	Name	ContactInformationName	string	0..1	連絡先の名前
組織	Organization	ContactInformationOrganization	string	0..1	連絡先の組織
電話番号	Telephone	ContactInformationTelephone	string	0..1	連絡先の電話番号
住所	Address	ContactInformationAddress	string	0..1	連絡先の住所
メールアドレス	EmailAddress	ContactInformationEmailAddress	string	0..1	連絡先のメールアドレス
住所	Address	ContactInformationAddress	string	0..1	連絡先の住所
送付先	MailingAddress	ContactInformationMailingAddress	string	0..1	送付先の住所、郵便番号
電話番号	TelephoneNumber	ContactInformationTelephoneNumber	string	0..1	連絡先の電話番号
内線番号	Extension	ContactInformationExtension	string	0..1	連絡先の内線番号
FAX番号	FaxNumber	ContactInformationFaxNumber	string	0..1	連絡先のFAX番号
携帯電話番号	MobileTelephoneNumber	ContactInformationMobileTelephoneNumber	string	0..1	連絡先の携帯電話番号
ウェブサイト	Website	ContactInformationWebsite	string	0..1	連絡先のウェブサイト
ID	Identification	ContactInformationIdentification	string	0..1	連絡先のID
利用言語	AvailableLanguage	ContactInformationAvailableLanguage	string	0..1	連絡先の連絡の際に利用可能な言語
電話番号	TelephoneNumber	ContactInformationTelephoneNumber	string	0..1	電話番号を識別する1つのデータ型
国番号	CountryCallingNumber	ContactInformationCountryCallingNumber	string	0..1	国際電話番号に使用する国番号
前綴	Prefix	ContactInformationPrefix	string	0..1	電話番号

Vocabulary Item

Person

Name (Class/Property)	Name (English)	Identifier (English)	Data Type (English)	cardinality	Description (English)
人	Person	ic:Person	extends ic:Entity		
氏名	Name	ic:PersonName	ic:PersonName	0..n	Name of a Person
性別	GenderText	ic:PersonGenderText	xsd:string	0..1	Gender of a Person
性別(コード)	GenderCode	ic:PersonGenderCode	ic:Code	0..1	Gender of a Person
生年月日	BirthDate	ic:PersonBirthDate	ic:DateTime	0..1	Date of Birth of a Person
死亡年月日	DeathDate	ic:PersonDeathDate	ic:DateTime	0..1	Date of Death of a Person
住所	ResidenceAddress	ic:PersonResidenceAddress	ic:Address	0..n	Present address of a Person
本籍	DomicileOfOrigin	ic:PersonDomicileOfOrigin	ic:Address	0..1	Legal residence address of a Person
連絡先	ContactInformation	ic:PersonContactInformation	ic:ContactInformation	0..n	Contact information of a Person
ID	Identification	ic:PersonIdentification	ic:Identification	0..n	Identification of a Person
国籍	NationalityText	ic:PersonNationalityText	xsd:string	0..n	A county that assigns rights, duties, and privileges to a person because of the birth or naturalization of the person in that country.
国籍(コード)	NationalityCode	ic:PersonNationalityCode	ic:Code	0..n	A county that assigns rights, duties, and privileges to a person because of the birth or naturalization of the person in that country.
出生国	BirthCountry	ic:PersonBirthCountry	xsd:string	0..1	A location where a person was born.
出生国(コード)	BirthCountryCode	ic:PersonBirthCountryCode	ic:Code	0..1	A location where a person was born.
出生地	BirthPlace	ic:PersonBirthPlace	ic:Location	0..1	A location where a person was born.

Dictionary Technology



Available NIEM XML Dictionaries

NIEM 2.1 dictionaries

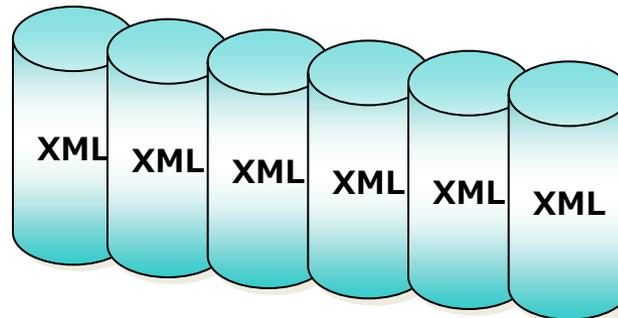
- CBRN dictionary
- Emergency dictionary
- Family dictionary
- Immigration dictionary
- Infrastructure dictionary
- Intelligence dictionary
- Justice dictionary
- Maritime dictionary
- Screening dictionary
- Trade dictionary
- Immigration blueprint
- **NIEM core dictionary**

Note: Those marked in bold are model style dictionaries with recursive components.

Available from download site *direct link:*

<http://sourceforge.net/projects/camprocessor/files>

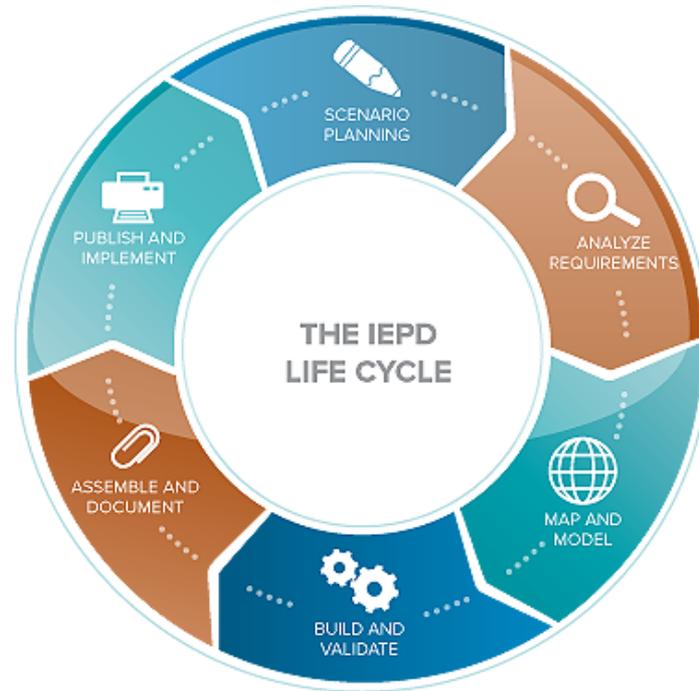
+ includes spreadsheets and sample models



The 8 "D"s and NIEM

- Design
- Develop
- Deploy
- Document
- Dictionaries
- Discovery
- Differentiate
- Diagnose

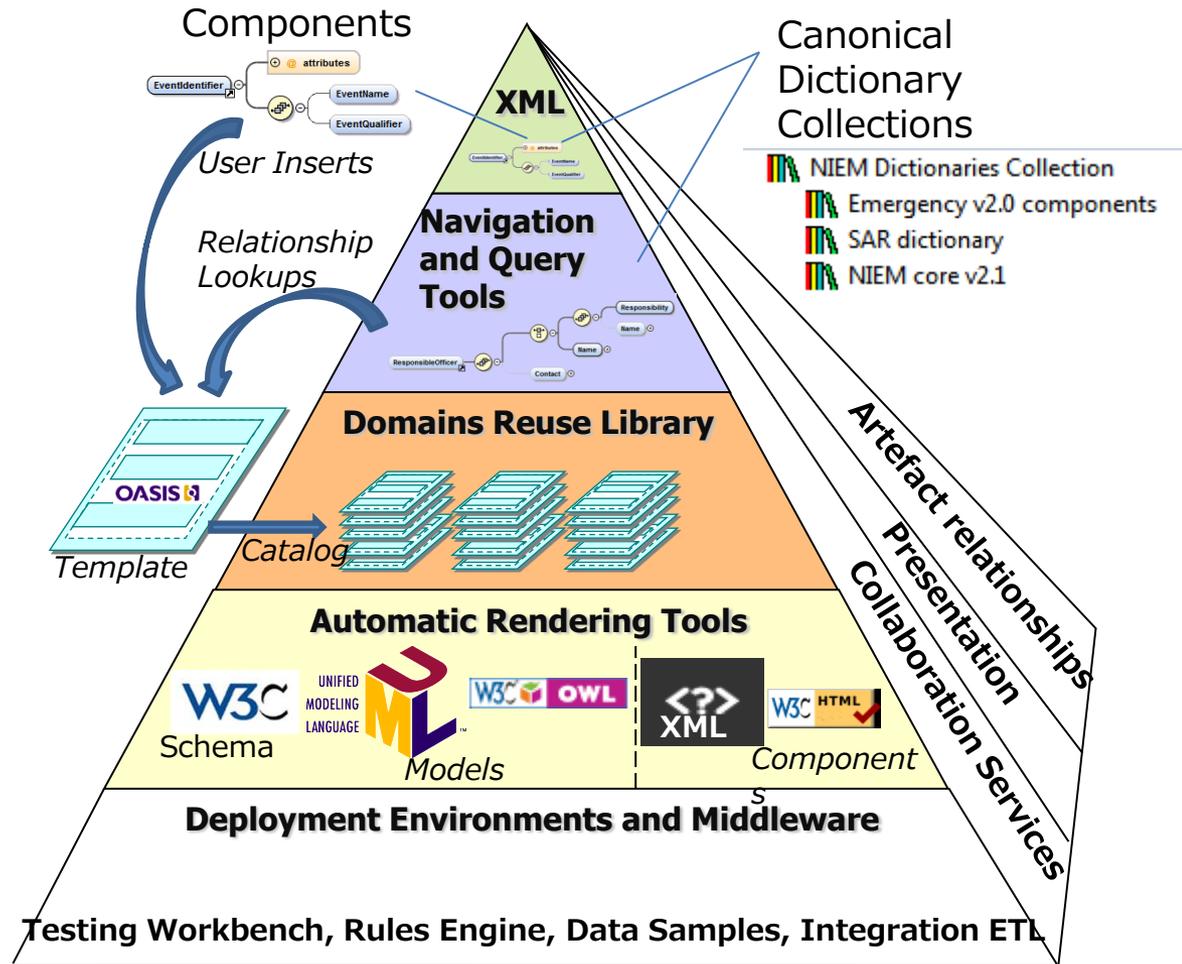
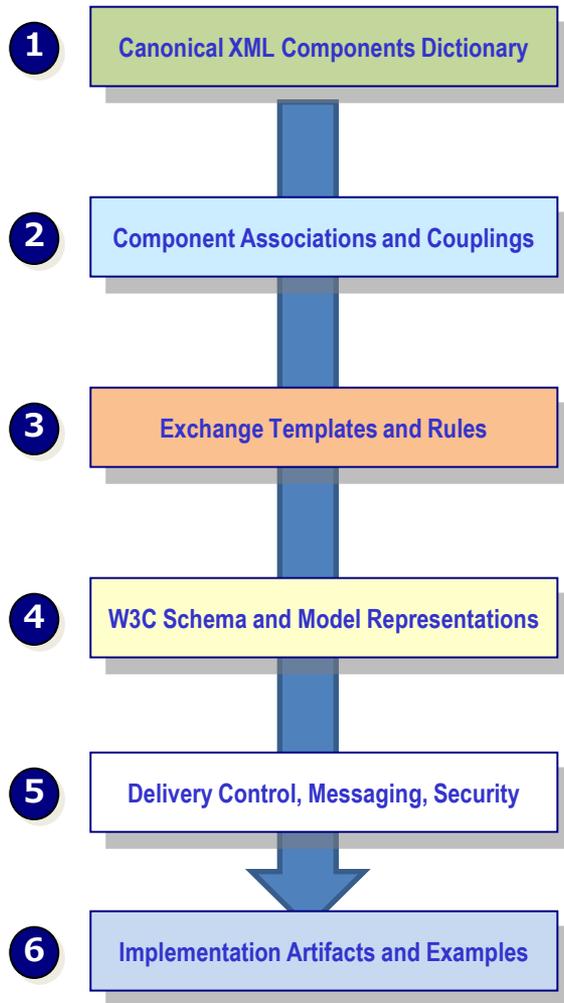
NIEM IEPD Process



Repeatable, Reusable Process
(Exchange Specification Lifecycle)

*IEPD - Information Exchange Package Documentation

Vision: Top Down Exchange Assembly



CAM Toolkit

<http://www.cameditor.org>

The screenshot displays the CAM Template Editor interface. The main window title is "CAM Template Editor - C:\CAM\dev\CAM\EDXL\HAVE\HospitalStatus.cam". The menu bar includes File, Edit, View, XML, Tools, Run, Window, and Help. The interface is divided into several panes:

- Structures:** A tree view showing the XML structure for "HospitalStatus". The root is "Structure ID='HospitalStatus' ref='' taxonomy='XML'". It contains a "Hospital" element with sub-elements: "Organization", "EmergencyDepartmentStatus", "HospitalBedCapacityStatus", "ServiceCoverageStatus", "HospitalFacilityStatus", "HospitalEOCStatus", "HospitalEOCPlan", "ClinicalStatus", "DeconCapacity", "MorgueCapacity", "FacilityStatus", "SecurityStatus", "Activity24Hr", "CommentText", "HospitalResourcesStatus", and "LastUpdateTime".
- ItemRules:** A table listing rules for the "HospitalFacilityStatus" item.
- Item Rules Table:**

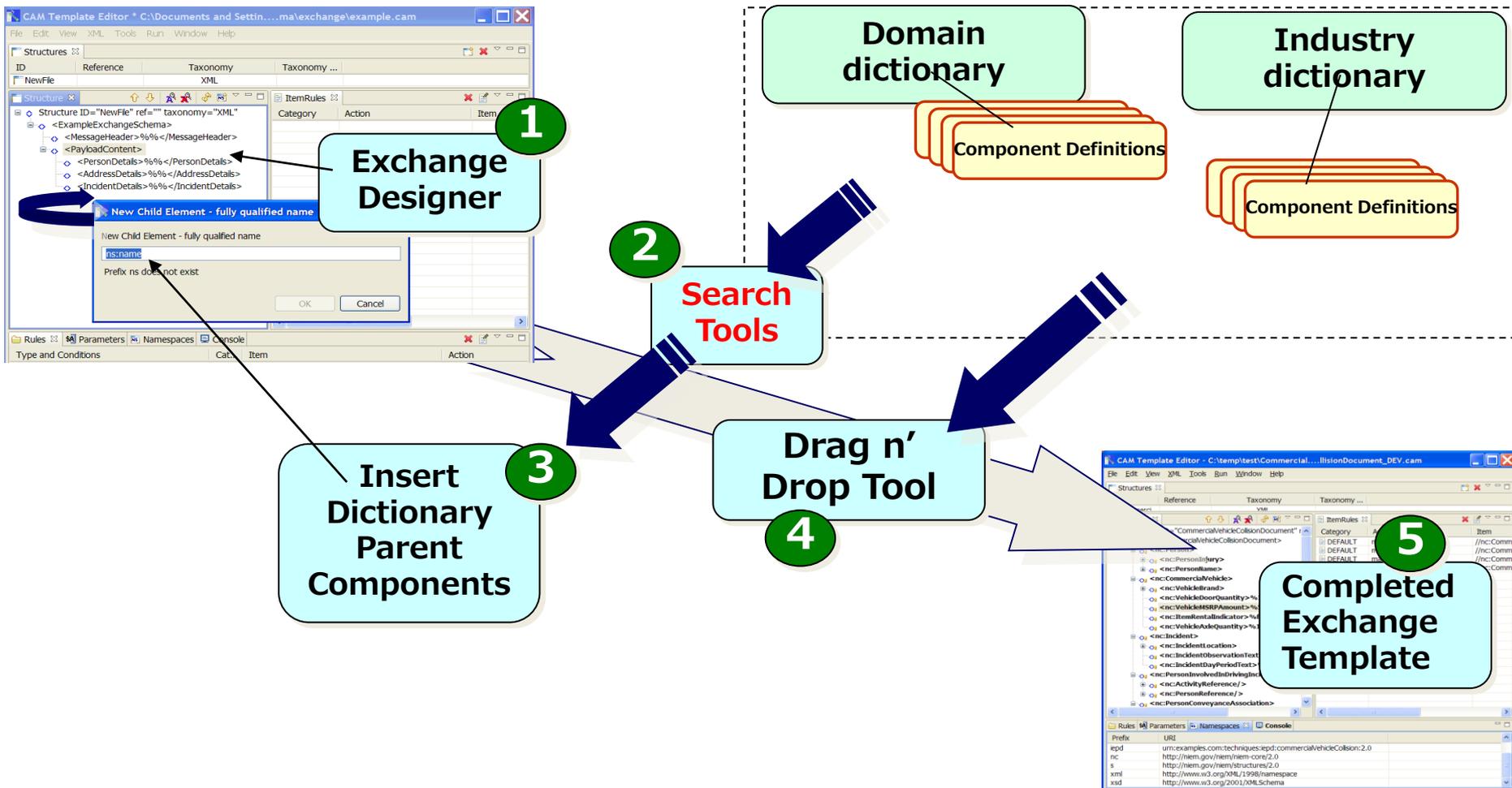
Category	Action	Item
DEFAULT	makeOptional()	//HospitalFaci
DEFAULT	restrictValues('Active' 'Inactive')	//HospitalFaci
DEFAULT	setDefault(Inactive)	//HospitalFaci
- Rules:** A table showing the "default context" rule.
- Rules Table:**

Type and Conditions	Category	Item	Action
default context	DEFAULT		

At the bottom right, there is a red banner with the CAM logo (a stylized robot head) and the text "Content Assembly Mechanism OASIS" and "ORACLE".

Visual Designer with Dictionary Collection

Collection



Alien Transfer Manifest – Mock Up

Structure ID="AlienExample" ref="" taxonomy="XML"

```

    <ae:AlienTransferManifest>
      <im:Transfer>%%</im:Transfer>
      <ae:StartingLocationFO>
        <ice:DetentionFacility>%%</ice:DetentionFacility>
      <ae:IntermediateLocation>
        <ice:DetentionFacility>%%</ice:DetentionFacility>
      <ae:EndingLocationFO>
        <ice:DetentionFacility>%%</ice:DetentionFacility>
      <at:AlienTransfer>%%</at:AlienTransfer>
      <ae:Person>
        <nc:PersonBirthDate>%%</nc:PersonBirthDate>
        <nc:PersonName>%%</nc:PersonName>
        <nc:PersonSexCode>%%</nc:PersonSexCode>
        <nc:PersonSexText>%%</nc:PersonSexText>
        <ice:Gang>%%</ice:Gang>
      <ae:ClassCode>
        <ice:ClassificationLevelCode>%%</ice:ClassificationLevelCode>
      <at:FINSID>%%</at:FINSID>
      <at:CrimeDetails>%%</at:CrimeDetails>
      <at:CrimeSharingDecision>%%</at:CrimeSharingDecision>
  
```

Prefix	URI
ae	http://alien-transfer/0.1/example
as	http://www.oasis-open.org/committees/cam
at	http://alien-transfer/0.1
camed	http://jcam.org.uk/editor
ice	http://ice.dhs.org
im	http://niem.gov/niem/domains/immigration/...
nc	http://niem.gov/niem/niem-core/2.0
s	http://niem.gov/niem/structures/2.0

Immigration component

NIEM components

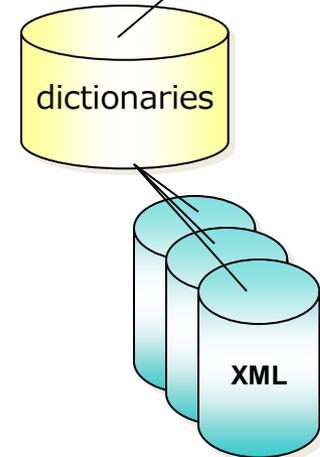
ICE components referenced

EID pre-defined collections based on NIEM + ICE + new

Namespaces of dictionary components

Collection

- ICE EDM dictionary
- NIEM Immigration dictionary
- EID extensions dictionary



Definitions stored as syntax neutral canonical XML

Technical Note: Inline child elements for ice: components shown for illustrative purposes only; actually are type reuse references (paperclip icon on parent node).

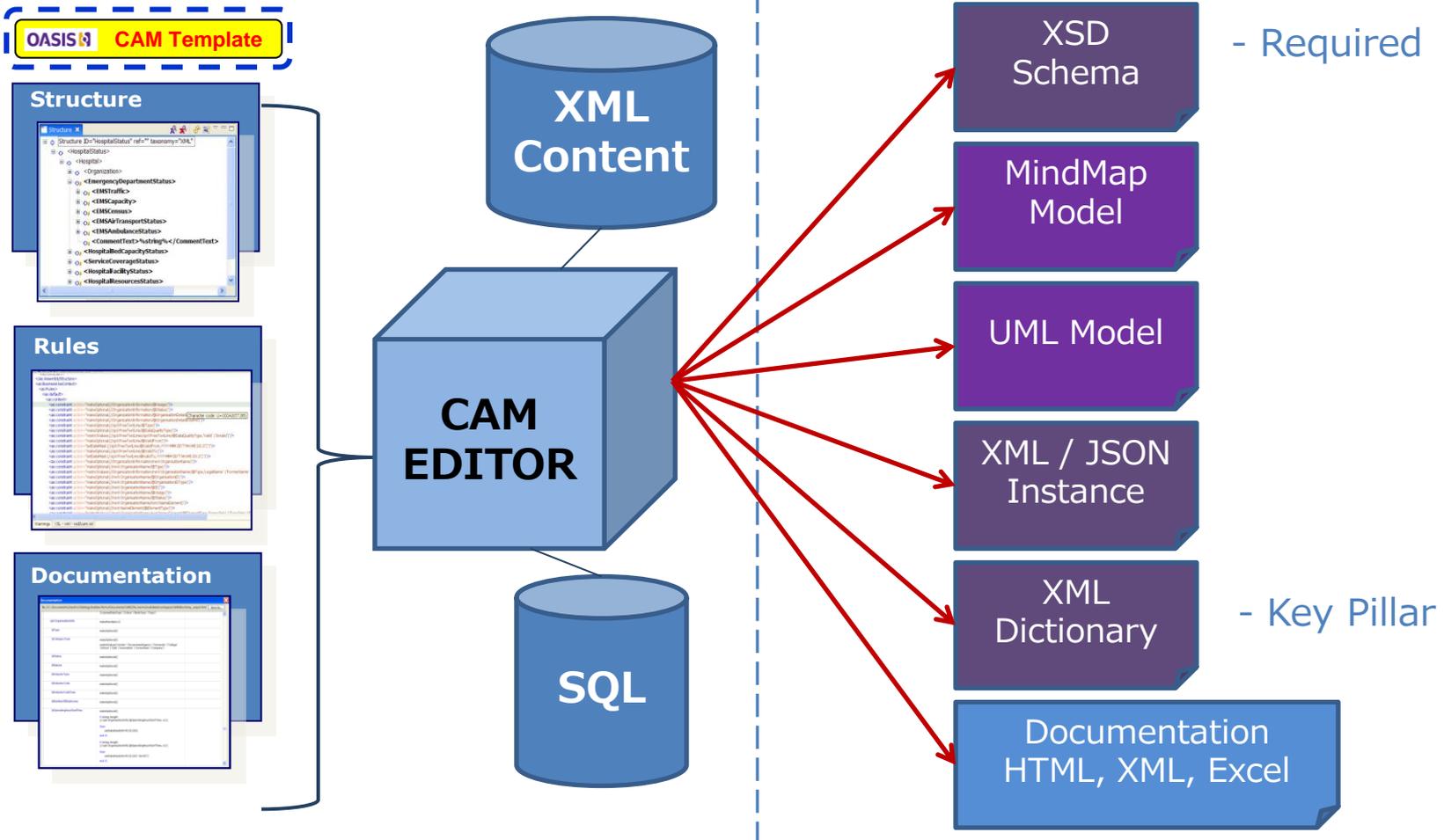
Differentiate

- This step includes building in deployment specific details and rules and usage policy determinations
 - Add additional XPath rules for local integration needs
 - Constrain code lists to local use
 - Limit and restrict content based on policy and role of exchange partners
 - **Contextually exclude structure components based on rules**
 - Create other integration artifacts for middleware
- *Can configure these aspects through the CAM template editor*

Delivery – CAM Editor Developer Tools

Editing / XML / SQL / Mapping

Target Renderings (via XSLT)



Example – Chicago Pharmacy Vaccines Search

Pharmacy Search Demo

<http://www.verifyxml.org>

WALGREEN CO 09001
191 NORTH CLARK STREET
CHICAGO IL 60601-6232

Vaccine	In Stock	Rx Required
Flu Shot	Y	N
Flu nasal spray	Y	N
Hepatitis A	Y	Y
Hepatitis B	Y	Y
HPV	Y	N

Business Hours

Day	Hours
Su	9:00am to 7:00pm
Mo	9:00am to 7:00pm
Tu	9:00am to 7:00pm
We	9:00am to 7:00pm
Th	9:00am to 7:00pm
Fr	9:00am to 7:00pm
Sa	9:00am to 7:00pm

Providers

- BOND DRUG COMPANY OF ILLINOIS
- CVS PHARMACY 4781
- OSCO DRUG 5514
- WALGREEN CO 09001
- WALGREENS 10558

Vaccine Search

Vaccines

- Hepatitis
- Shingles
- TD
- Flu
- Coccid
- HPV
- Varicella
- MMR

Location

Enter ZIP code: 60601

Radius: 0 miles

Search

Copyright ©2013 VerifyXML.org

1. Socrata data extract
2. MySQL database
3. Glassfish
4. Open-XDX
5. Prime Faces
6. Hosting site

Combines: rapid development paradigm; open data approach; open source tools; RESTful web services; community based resources and delivery; NIEM-aligned information feeds.

CAMeditor.ORG Project Statistics



SNAPSHOT OF PROJECT ACTIVITIES

250,000 CAMeditor.org page visits

175 countries have downloaded tools;

**30% of visitors are from U.S.;
850+ downloads weekly**

**9,000 students views of
online video training
resources**

8 languages now available

www.cameditor.org