Balisage 2014
The Markup Conference

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

Questions:

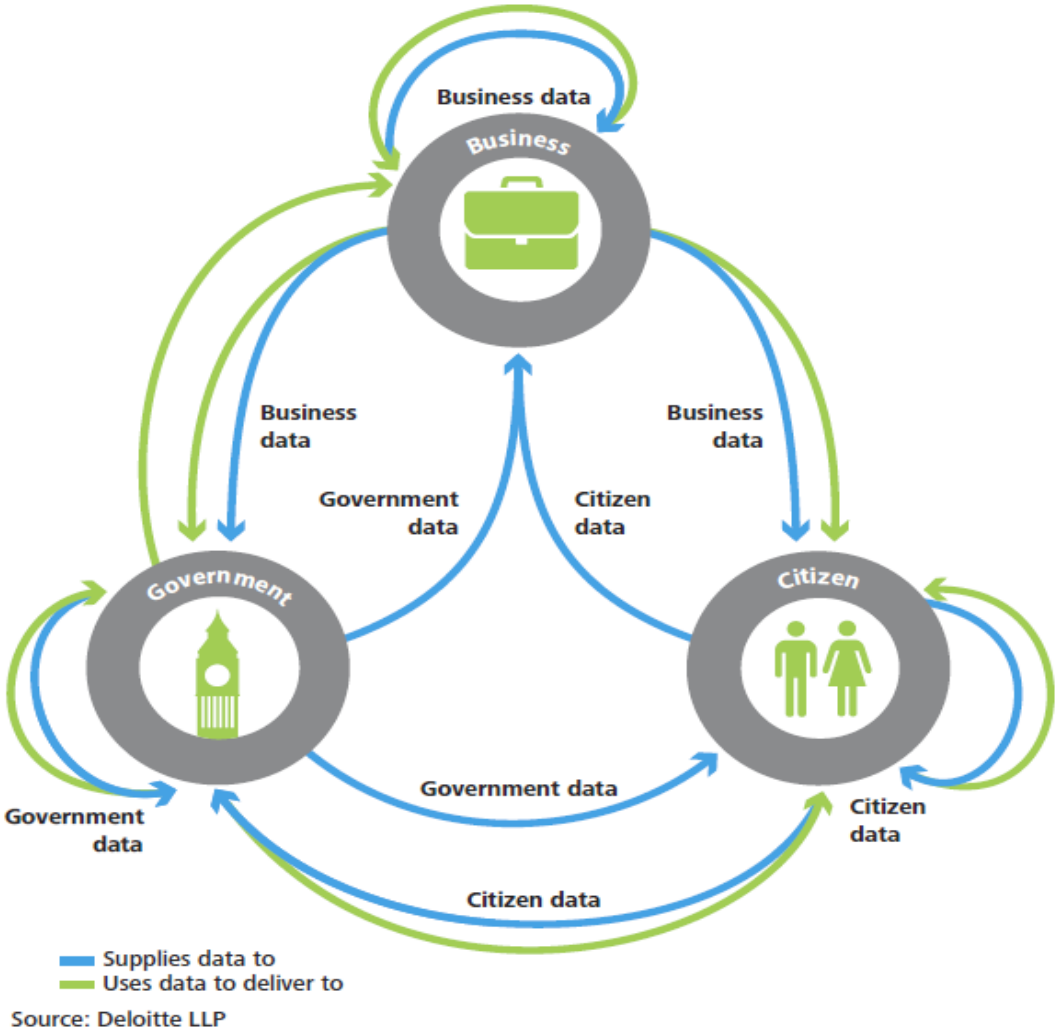
What data standards will be selected / developed?

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



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

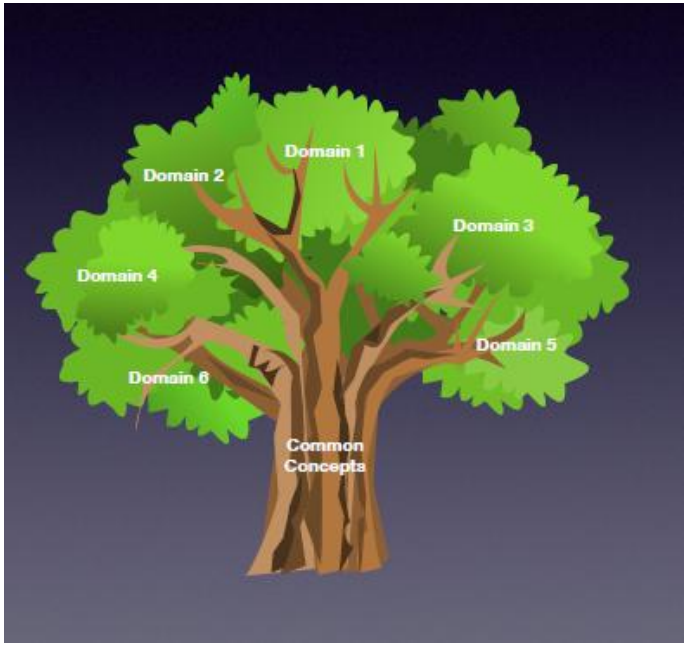
Next Generation Approach

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

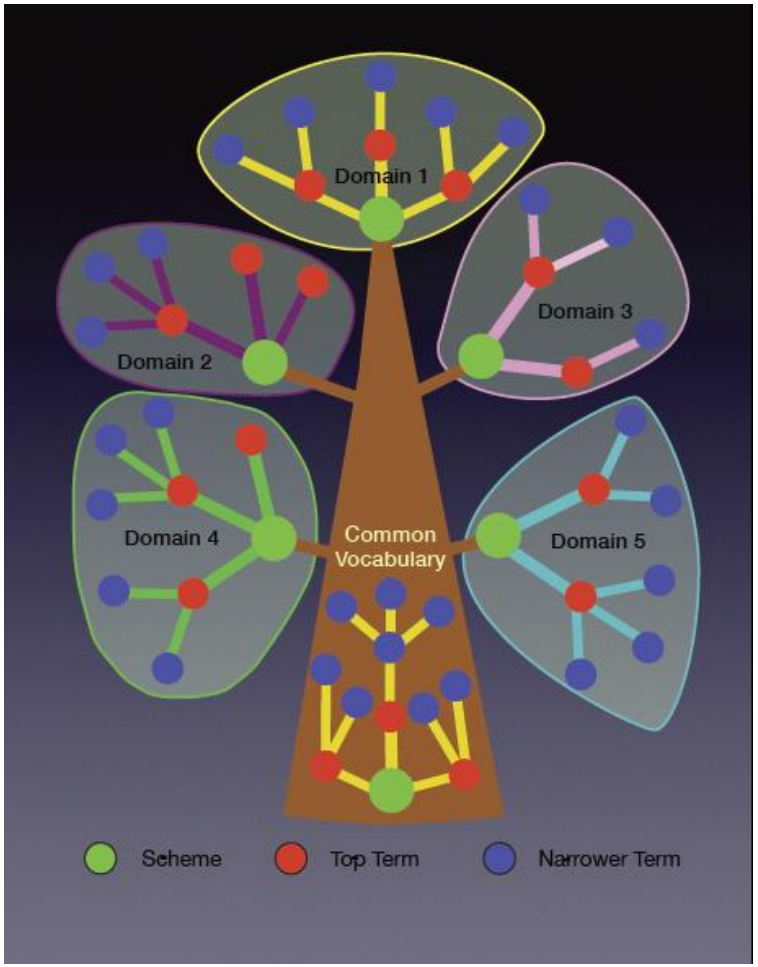
Toward Semantic Interoperability

- 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

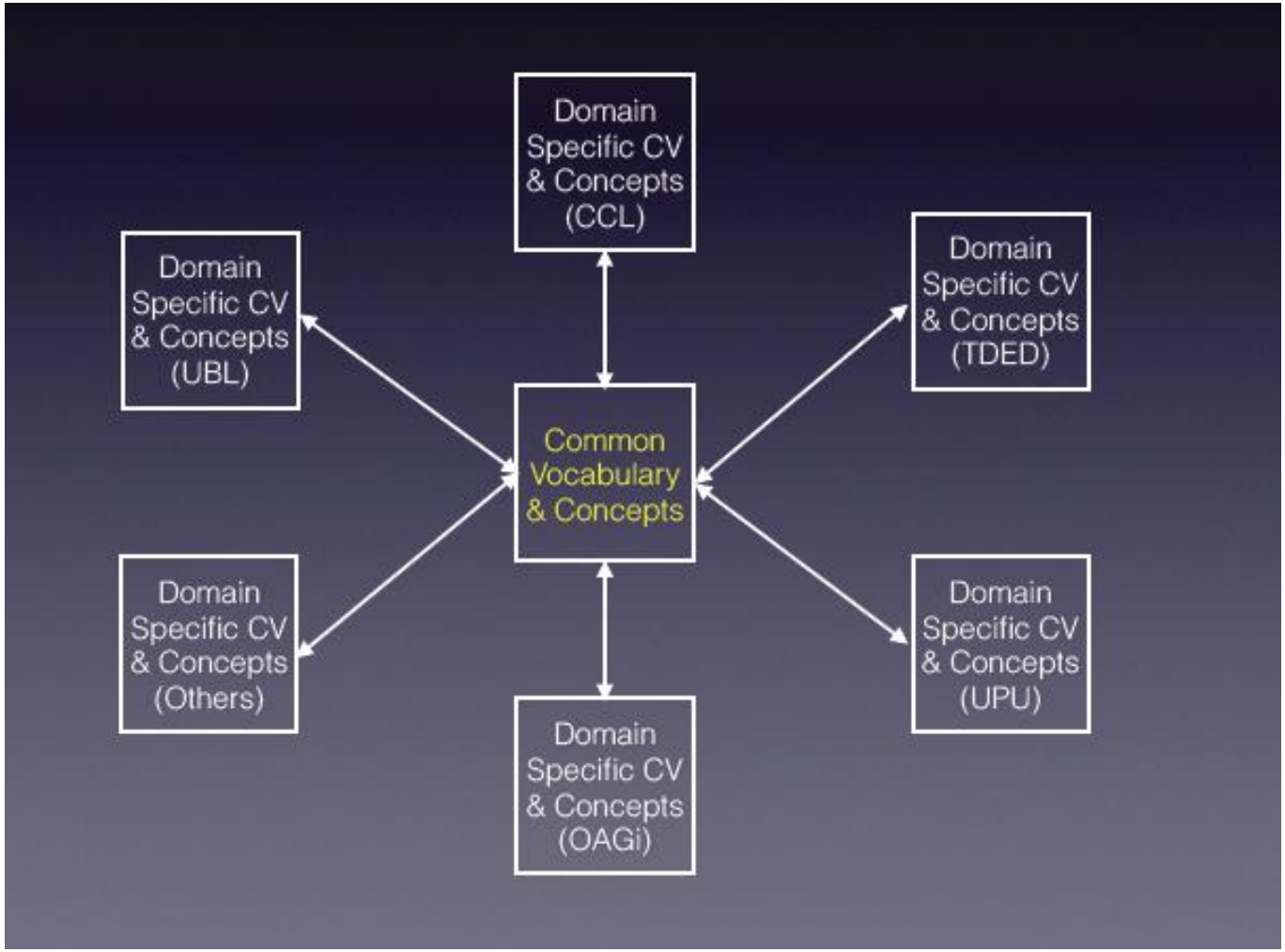


Semantic Concepts Tree

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



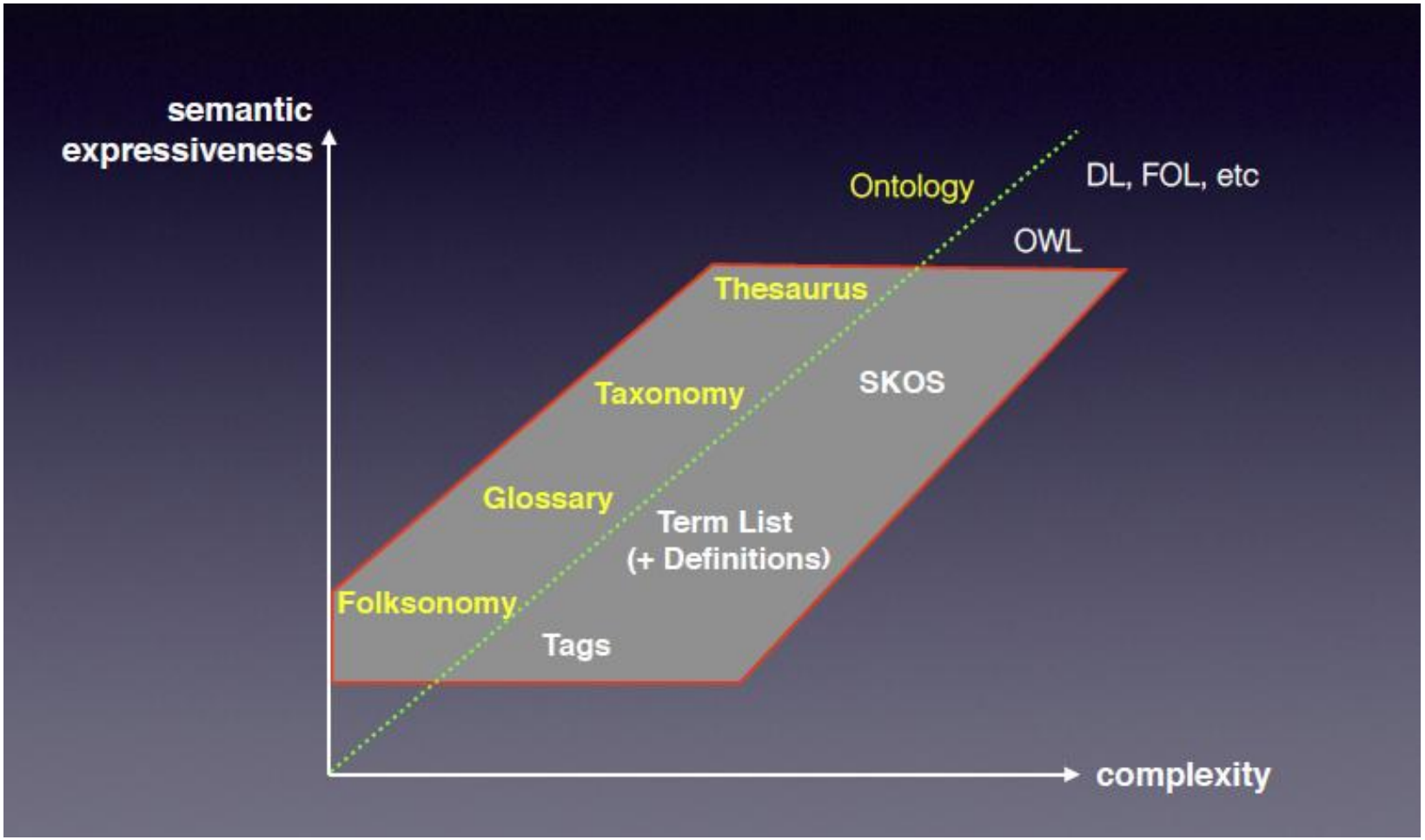
Linking Concepts using Web and SKOS



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

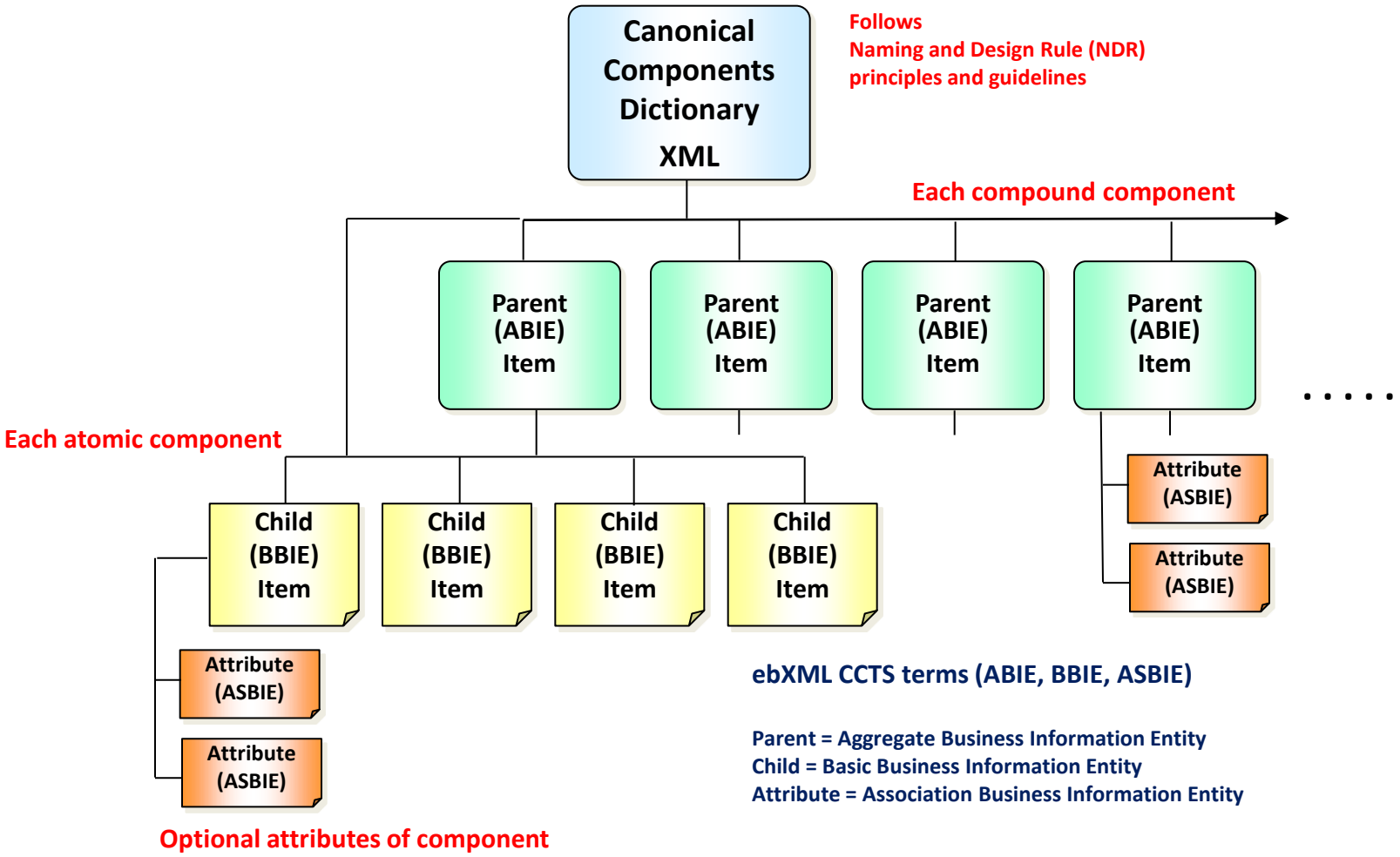
Knowledge Organization Systems (KOS)



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

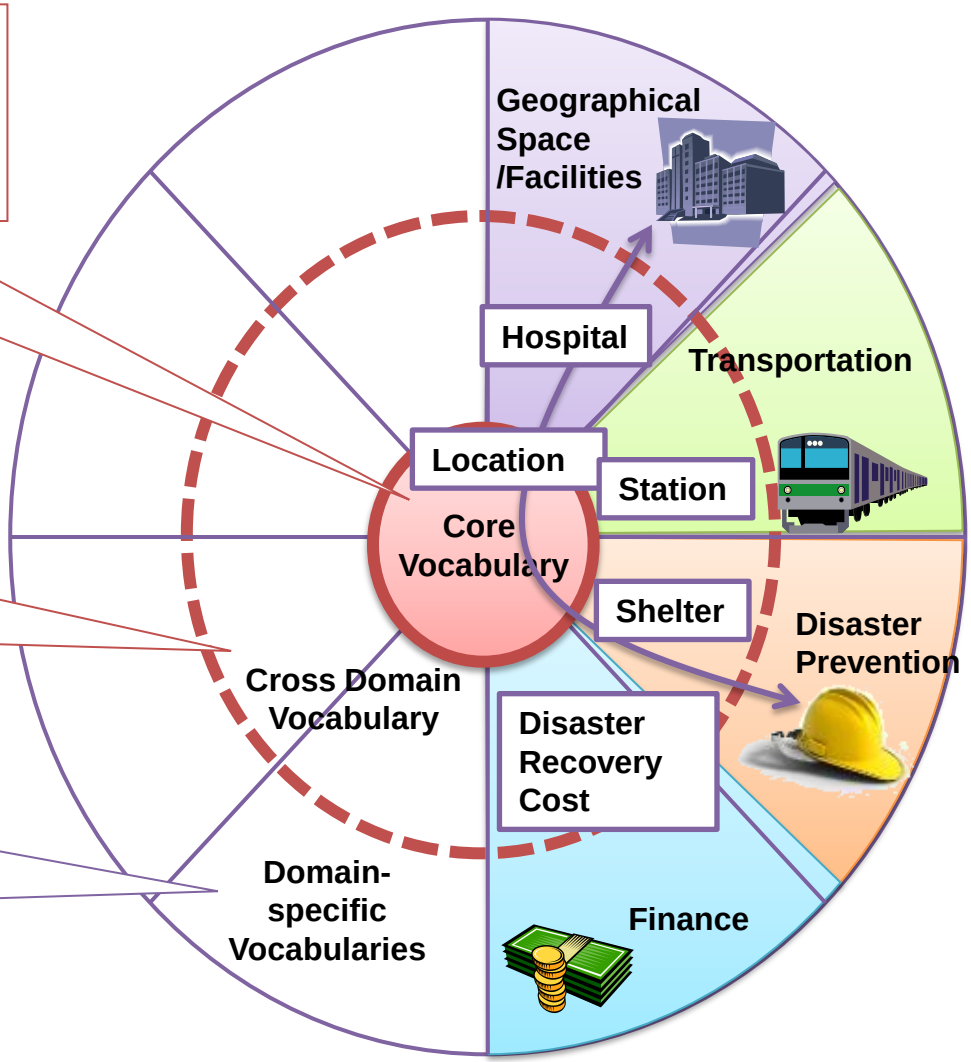
SKOS, CCTS, Dictionaries, NIEM,
Open Data, FOSS, XML, JSON

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.

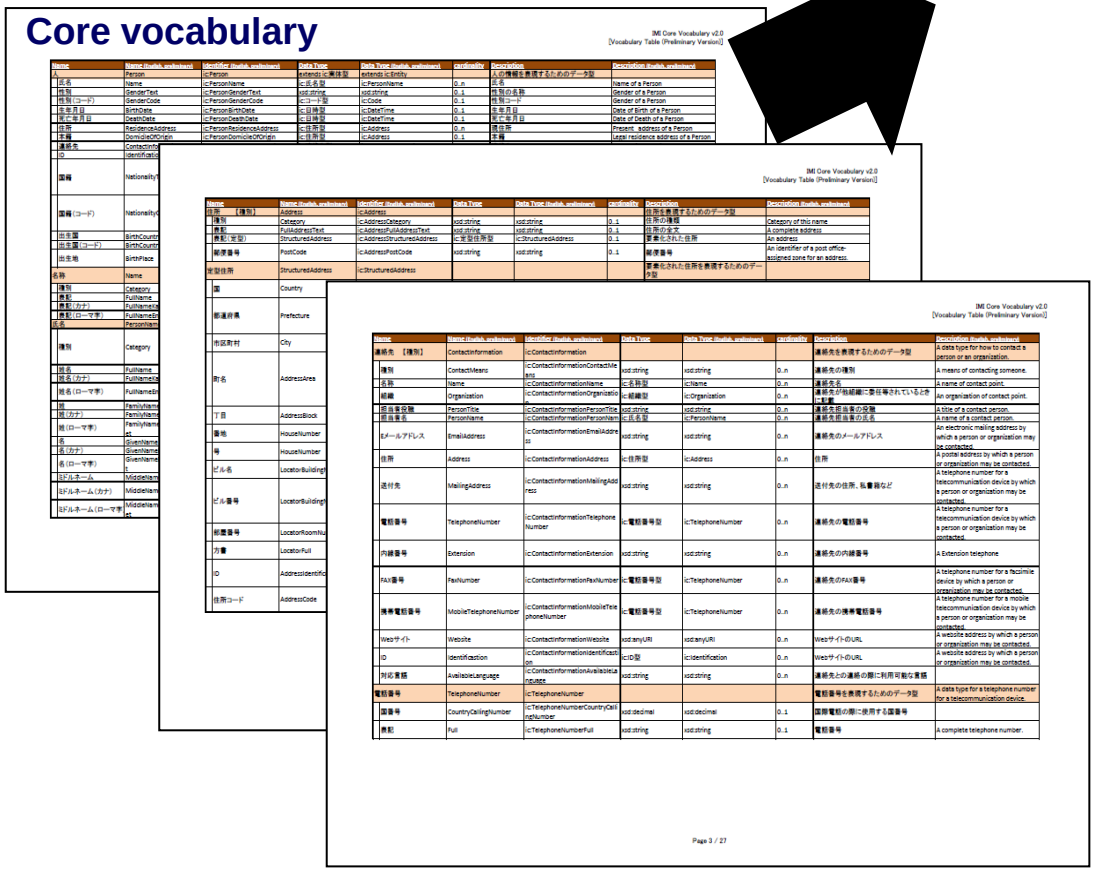
Domain-specific Vocabularies
Vocabularies that are specialised for the use in each domain.
E.g. Number of beds, Schedule.



Vocabulary Example

- Top Terms

Person	Location	Organization	Facility
Person Name	Address	Organization	Facility
Person	Spot	Name	Building
			Equipment

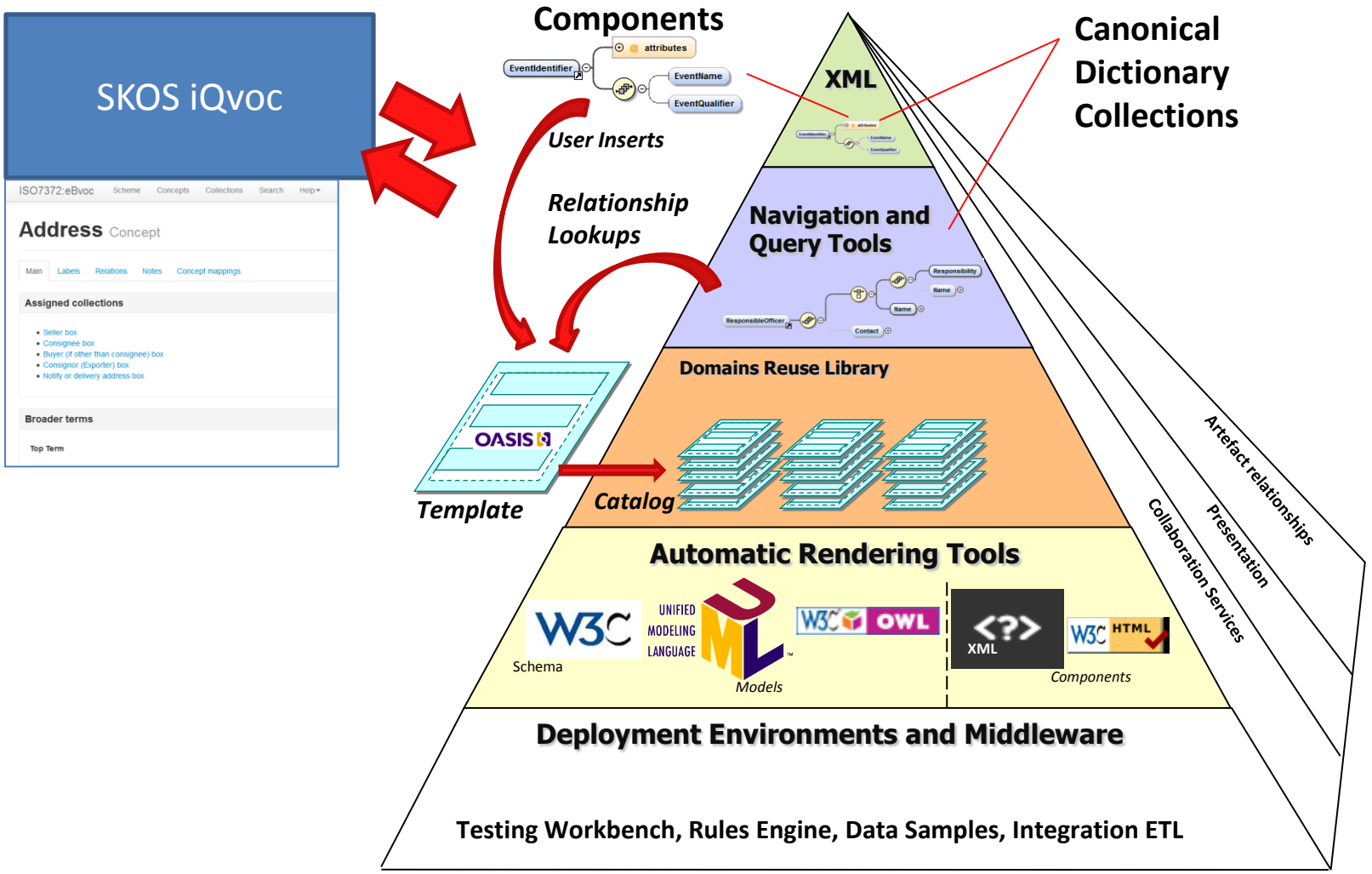


Vocabulary Item

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

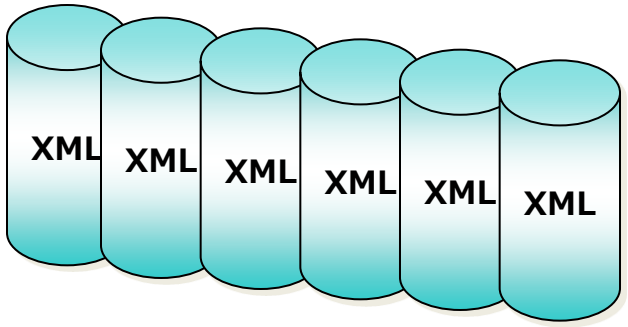
- CBRN dictionary
- Emergency dictionary
- Family dictionary
- Immigration dictionary
- Infrastructure dictionary
- Intelligence dictionary
- Justice dictionary
- Maritime dictionary
- Screening dictionary
- Trade dictionary
- Immigration blueprint
- **NIEM core dictionary**

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

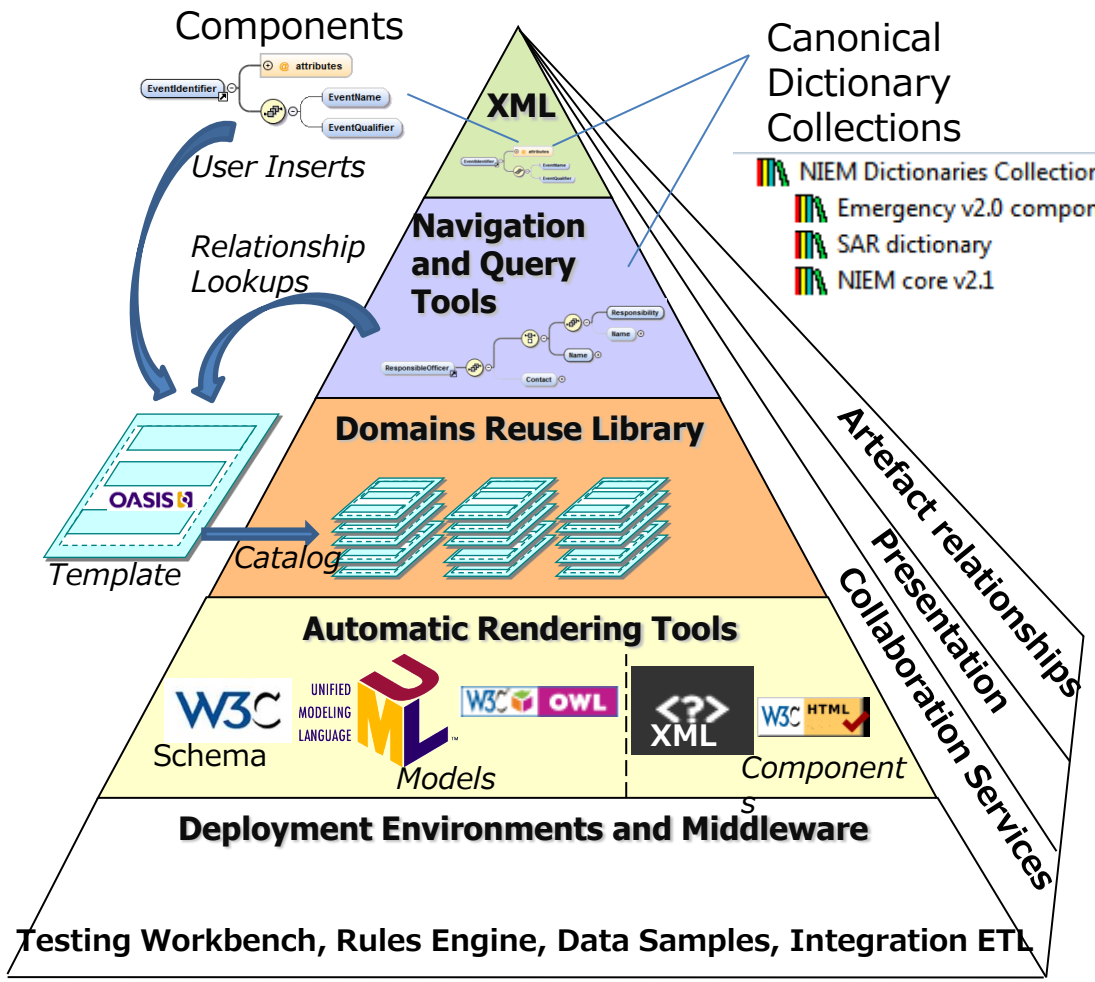
NIEM IEPD Process

Repeatable, Reusable Process
(Exchange Specification Lifecycle)

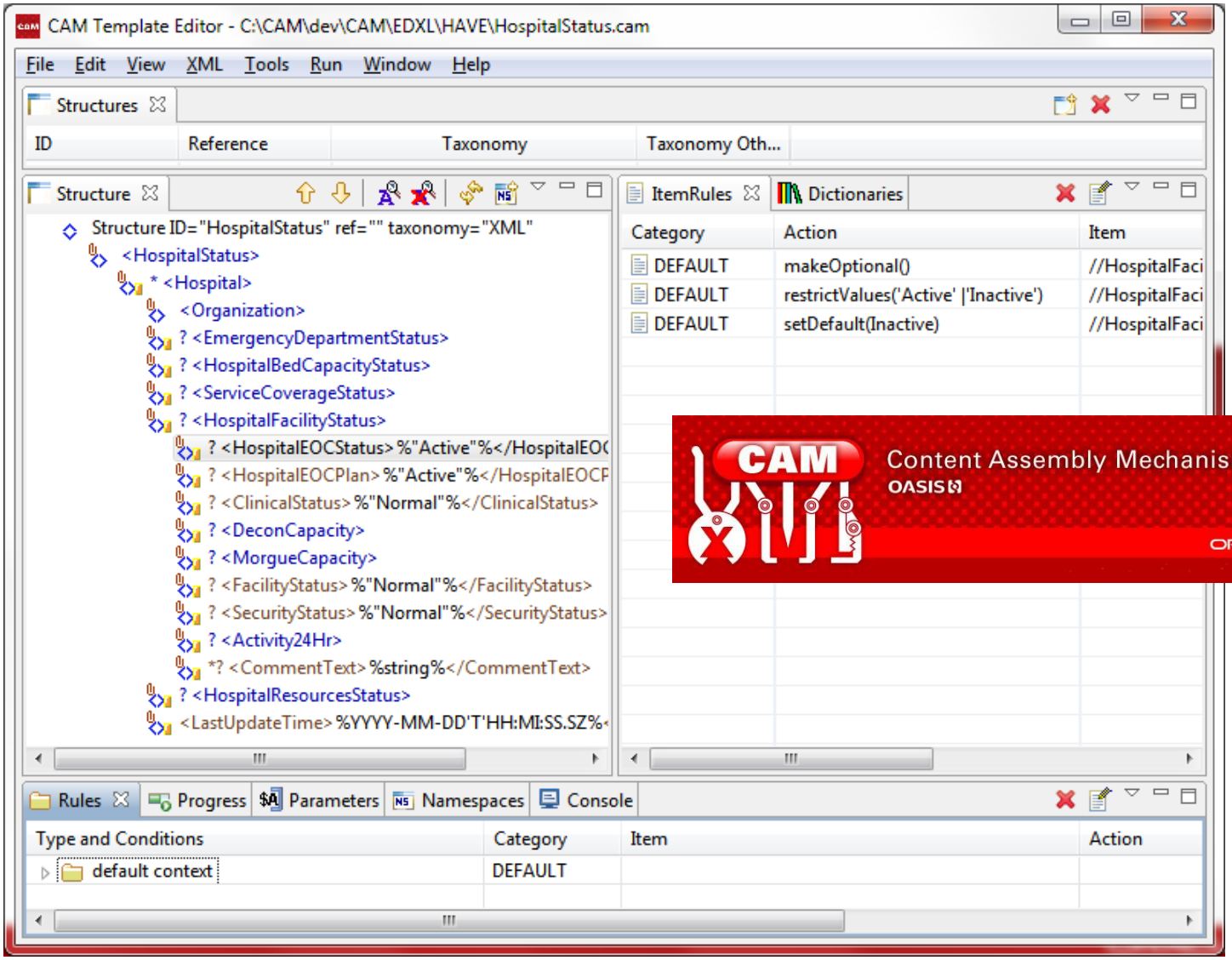
*IEPD - Information Exchange Package Documentation

Vision: Top Down Exchange Assembly

1. Canonical XML Components Dictionary
2. Component Associations and Couplings
3. Exchange Templates and Rules
4. W3C Schema and Model Representations
5. Delivery Control, Messaging, Security
6. Implementation Artifacts and Examples

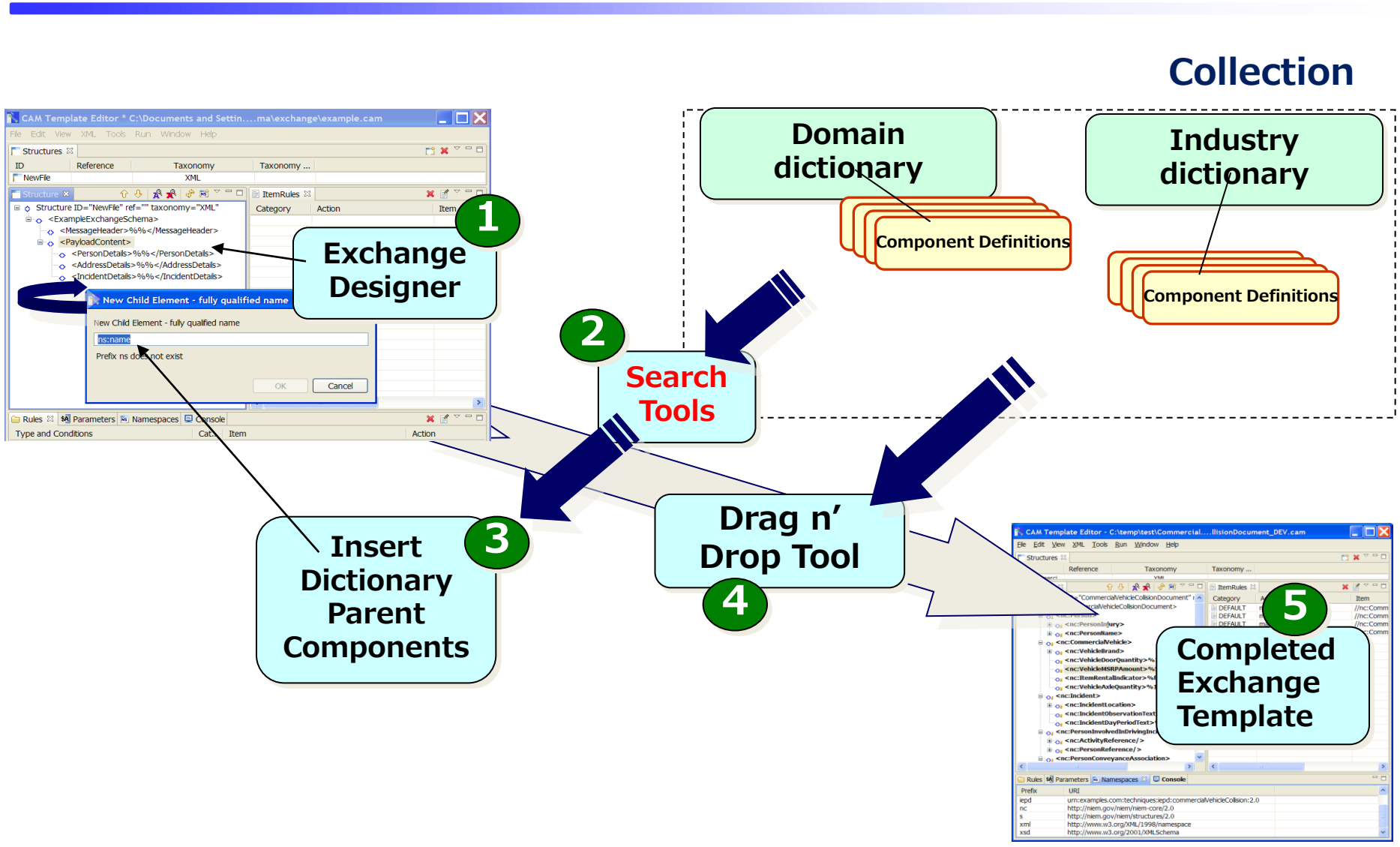


CAM Toolkit

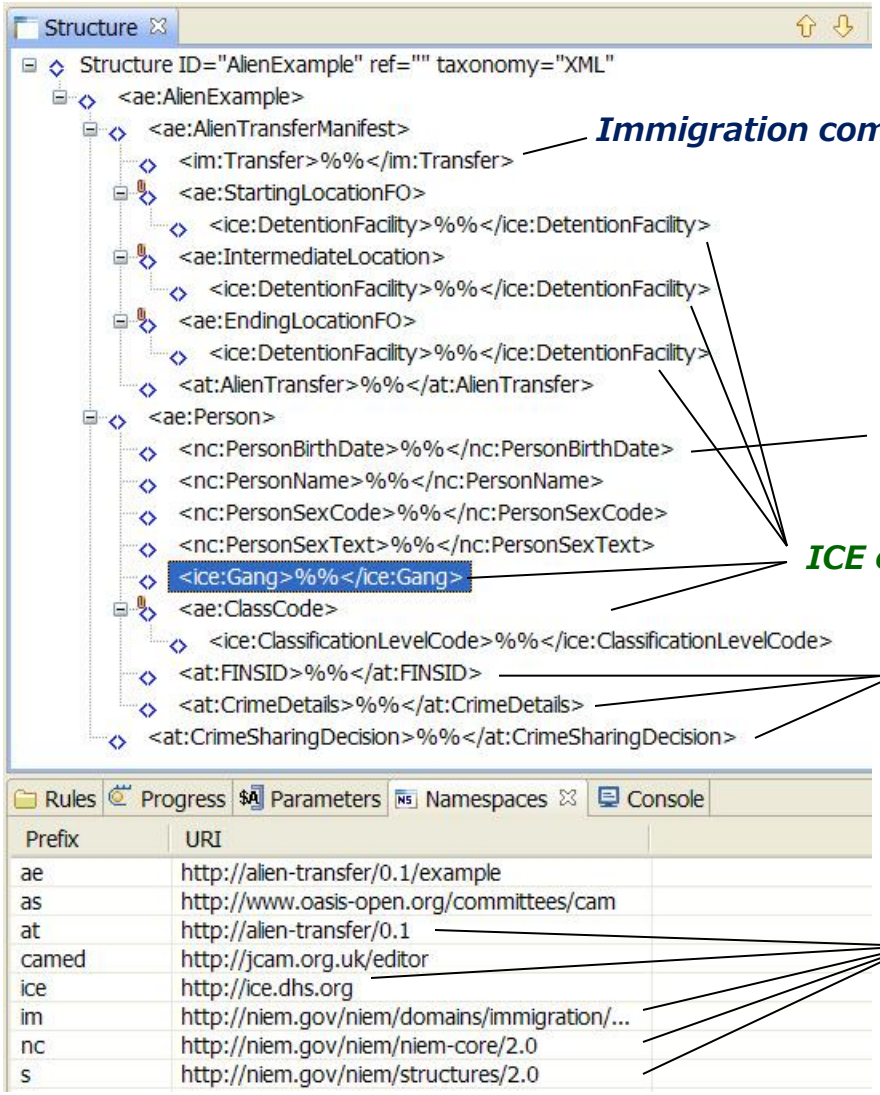


http://www.cameditor.org

Visual Designer with Dictionary Collection



Alien Transfer Manifest – Mock Up



Immigration component

NIEM components

ICE components referenced

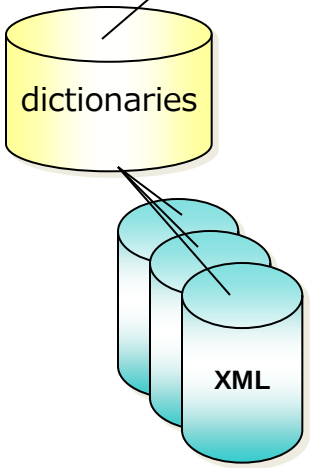
EID pre-defined collections

based on NIEM + ICE + new

Namespaces of dictionary components

Collection

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



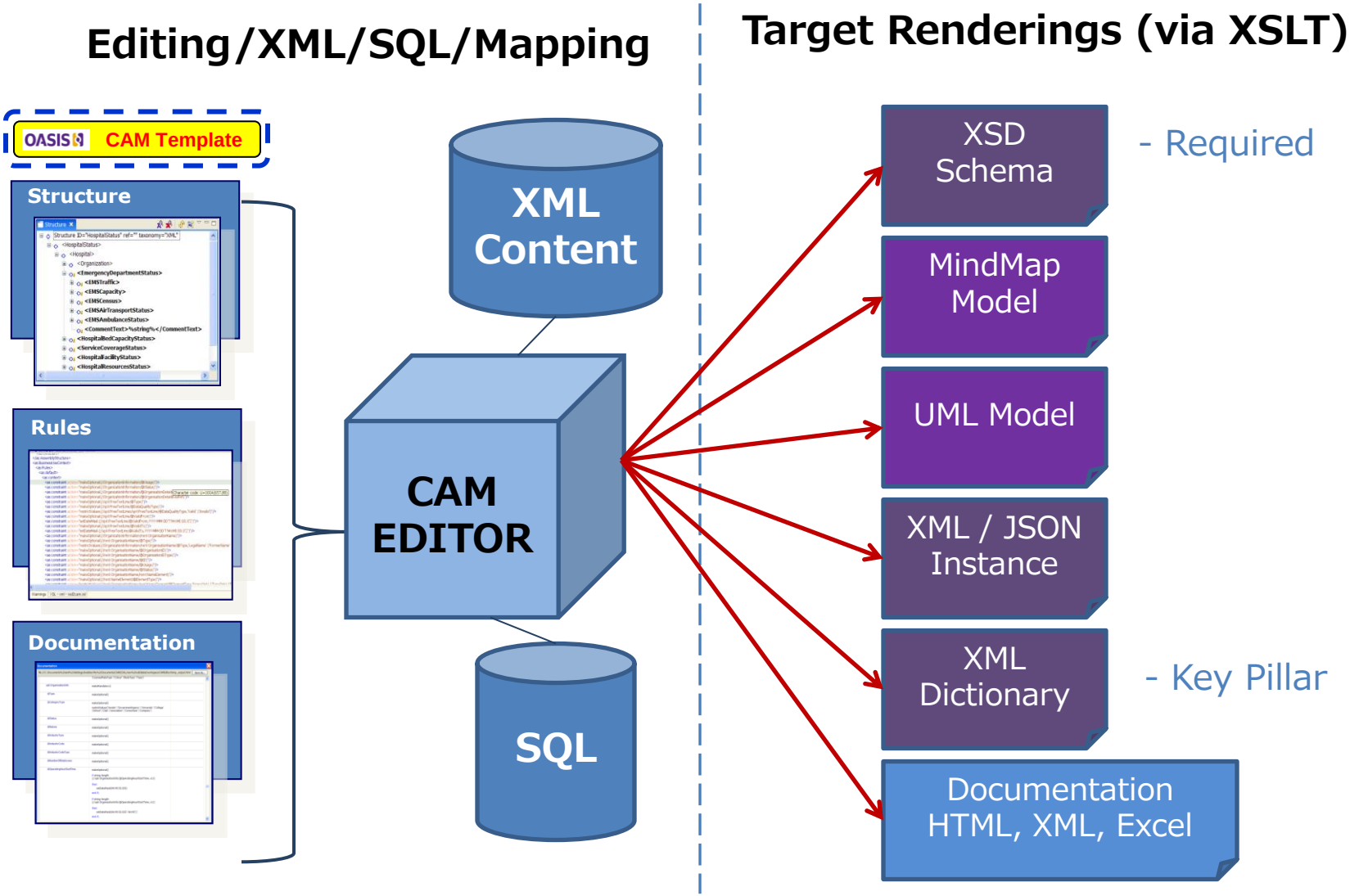
Definitions stored as syntax neutral canonical XML

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

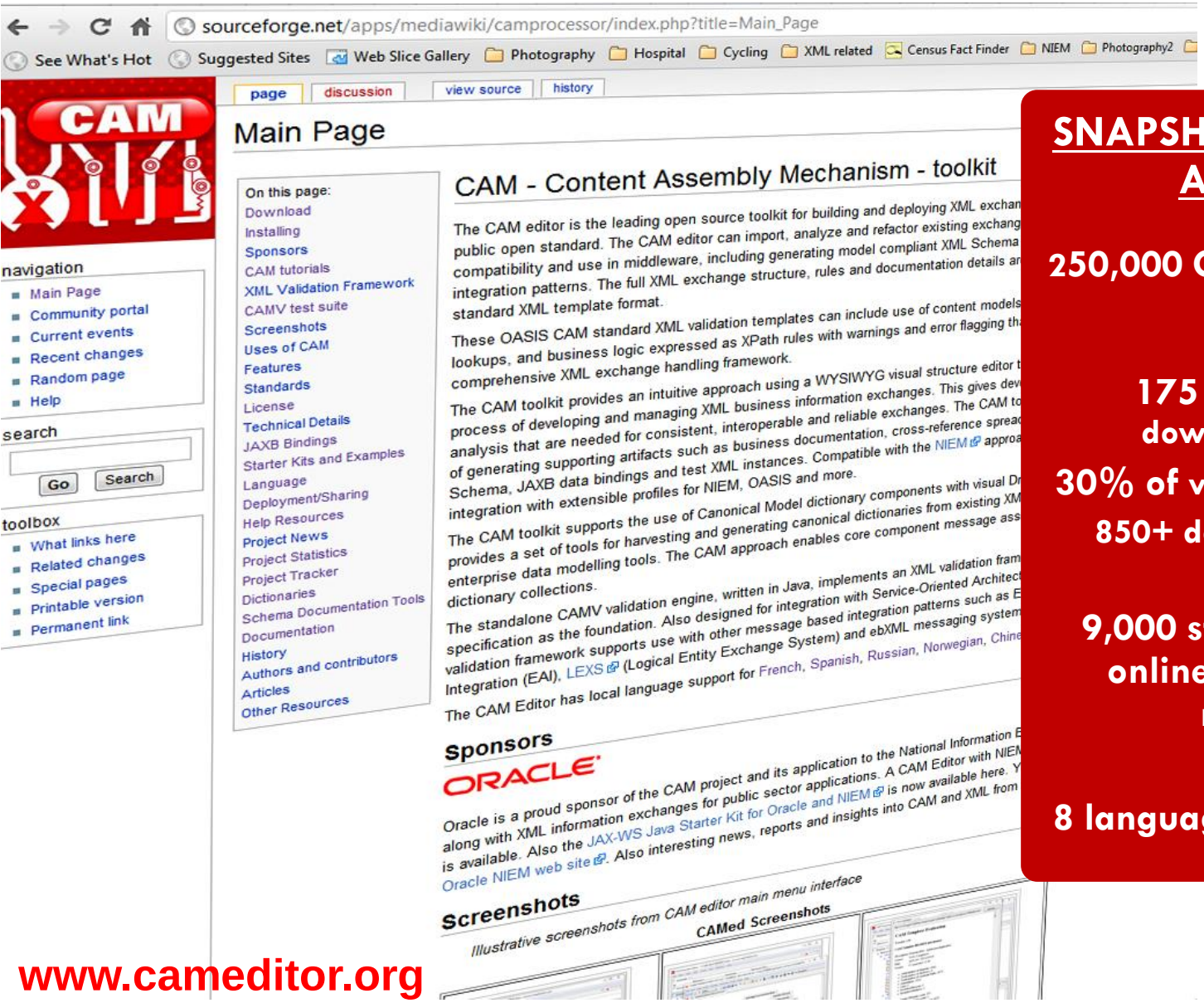


Example – Chicago Pharmacy Vaccines Search

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