

Design of an Organization Authority Database with classification principles

Dagobert Soergel

Department of Library and Information Studies
Graduate School of Education, University at Buffalo

Denisa Popescu

World Bank Group, Washington, DC, USA

UDC 2015 Seminar Lisbon 2015 October 29-30
Proceedings Ergon p. 69 - 81

Outline

- 1 Introduction**
- 2 The use case**
- 3 Design**
 - 3.1 Data structure:
Beginnings of the conceptual data schema
 - 3.2 User interface and search
 - 3.3. System implementation
- 4 Populating an Organization Authority Database**
- 5 Conclusions**

1 Introduction

Theme: Unification

To unify =

1. to recognize common (abstract) structures

2. and exploit for

- **sharing software modules** across applications
- **common user interface** across applications

2 The use case

Many data systems of the World Bank Group deal with organizations in different roles, for example:

- **suppliers** to the WBG, including consulting companies
- **suppliers** or potential suppliers **for projects funded by the WBG**
- **customers**
- loan recipients
- partners,
- For an business: **competitors** (competitive intelligence),
- **authors or subjects of documents** (library and several systems that manage internal and external documents)
- **search terms** when searching for texts (including Web search) by or about an organization, or any of a group of organizations
- **sub-units of the organization are themselves organizations** that can occur in some of these roles plus additional roles, such as organization where a person works

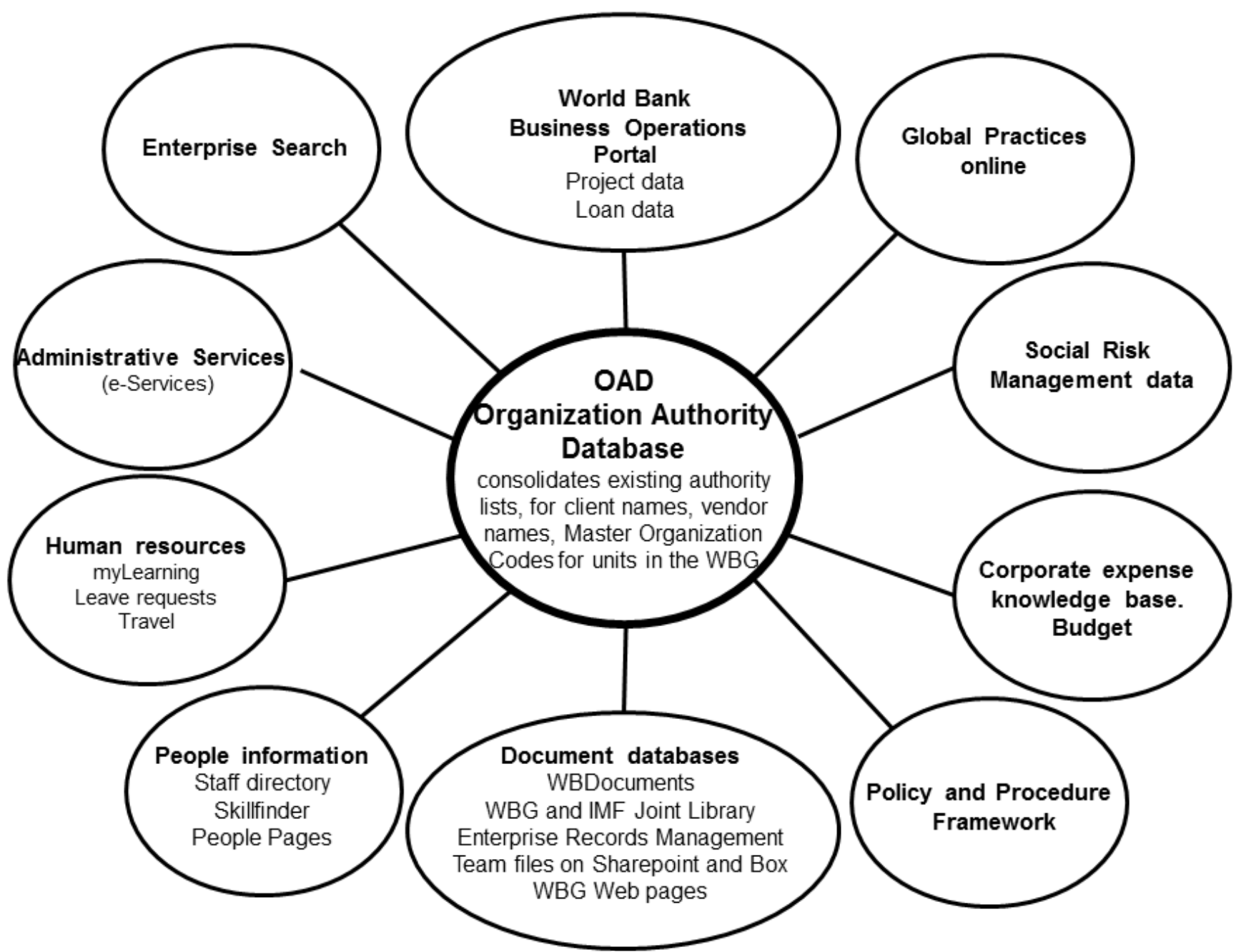


Figure 1. A Central Organization Authority Database

2 The use case, slide 2

Needed

An **Organization Authority Data Base (OAD)** that gives for each organization

1.a unique URI that can be used to link information across all WBG systems

2.all names and acronyms in many languages

3.more basic information that is useful in itself and that can be used to search for organizations, including **hierarchical relationships** between organizations

2 The use case, slide 3

Efficiencies and usage advantages of a central OAD for the WBG

- 1.A single system for maintaining and serving organization data**
- 2.Acquiring data about organizations from external sources saves maintenance effort and gives a more complete database
- 3.Accessing all data about an organization available in any of the WBG data systems through the unique URI**
- 4.Accessing data about an organization available in external sources, including the Web
- 5.Providing superior support for searching .**

3 Design

Much in common between

**an Organization Authority Database and
a hierarchically structured thesaurus:**

- **Organizations form a hierarchy**
- **Organizations may have many names**
- Both the hierarchy and the multiple names can be used for **query term expansion** to support search

3 Design, slide 2

3.1 **Data structure:**

Beginnings of the conceptual data schema

3.2 **User interface and search**

3.3. **System implementation**

3.1 Data structure: Beginnings of the conceptual data schema

Entity	<i><isa></i> ~ <i><hasInstance></i>	EntityType	For organizations: OrganizationType org: <i>classification</i>
Entity	<i><hasName></i>	(Name, NameStatus)	skos: <i>label</i> . NameStatus examples: PreferredName, AlternateName, OfficialLegalName, DoingBusinessAs
Entity	<i><hasStartTime></i>	PointInTime	
Entity	<i><hasEndTime></i>	PointInTime	
Entity	<i><hasSuccessor></i> ~ <i><hasPredecessor></i>	Entity	See org:5.6
Entity ,	<i><isPartOf></i> ~ <i><hasPart></i>	Entity	org: <i>unitOf</i> ~org: <i>hasUnit</i>
Entity	<i><isAbout></i> ~ <i><coveredIn></i>	Entity	Narrower <i><hasPurpose></i>
Entity	<i><coveredIn></i>	Document	E.g., the home page
Entity	<i><hasPurpose></i>	Entity	org: <i>purpose</i> Broader <i><isAbout></i>
Entity	<i><hasDescription></i>	Text	

Figure 2. A partial organization ontology for illustration

org: the W3C Organization Ontology

skos: Simple KOS ontology

Notes

1. Use multiway relationships for adequate or more efficient representation. Avoid the limitations of RDF.
2. All statements in the database can be qualified by TimeSpan.
3. All string values (Name, Text) have a language indicator (such as @fr)
4. Many relationships apply to all kinds of entities, including organizations.
5. LegalEntity includes Person and Organization, approx. = foaf: Agent.
6. Entity instances identified by a URI used across the Web.
7. ~ means inverse relationship

Figure 2. A partial organization ontology. Notes

foaf: the Friend Of A Friend Ontology

Organization	<i><hasHeadquarterLoc.></i>	Location	Could be as specific as address org:5.4 has more detail
Organization	<i><hasOfficialLanguage></i>	Language	
Entity	<i><hasNarrower></i> ~ <i><hasBroader></i>	Entity	skos:narrower ~skos:broader org: <i>hasSubOrganization</i> ~ org: <i>subOrganizationOf</i> Narrower <i><hasPart></i> , <i><hasOrgFamMember></i> , <i><owns></i> , <i><hasSubsidiary></i>
Organization	<i><hasOrgFamMember></i>	Organization	Broader Rel: <i><hasNarrower></i>
Organization	<i><owns></i>	Organization	Broader Rel: <i><hasNarrower></i>
Organization	<i><hasSubsidiary></i>	Organization	Broader Rel: <i><hasNarrower></i>
Organization	<i><org:linkedTo></i>	Organization	
Organization	<i><org:hasMember></i> ~ <i><org:memberOf></i>	LegalEntity	
Organization	<i><hasStaffMember></i>	(Person, InOrgRole)	In org: the artificial class <i>membership</i> special case: org:headOf
Organization	<i><org:hasPost></i> ~ <i><org:PostIn></i>	Post	In US English: Position

Figure 2. A partial organization ontology for illustration

3.2 User interface and search

- One interface: Hierarchy browse
- Works just like a hierarchy browse for a classification

- ▼ **United Nations Family**
- ▶ **UN General Assembly**
- ▶ **Security Council**
- ▶ **Secretariat**
- ▶ **Economic and Social Council**
- ▶ **International Court of Justice**
- ▶ **Trusteeship Council**

- ▼ **US Government Agencies**
- ▶ **Departments**
- ▼ **Independent agencies (selected)**
- ▶ **Civil Service agencies**
- ▶ **Education agencies**
- ▶ **Energy and science agencies**
- ▶ **Interior agencies**
- ▶ **Labor agencies**
- ▶ **Monetary and financial agencies**
- ▶ **Retirement agencies**
- ▶ **Transportation agencies**
- ▶ **Volunteerism agencies**
- ▶ **Defense and Security agencies**
- ▶ **Civil Rights**

Figure 3a. A Tree Browse Window with limited drill-down

- ▼ **United Nations Family**
- ▶ **UN General Assembly**
- ▶ **Security Council**
- ▶ **Secretariat**
- ▼ **Economic and Social Council**
- ▶ **Funds and Programmes**
- ▼ **Specialized Agencies (listing just a few)**
- ▶ **FAO, Food and Agriculture Organization of the UN**
- ▶ **WHO, World Health Organization**
- ▶ **UNESCO, UN Educational, Scientific and Cultural Org.**
- ▶ **IMF, International Monetary Fund**
- ▼ **World Bank Group**
- ▼ **World Bank**
- ▶ **IBRD, Internat. Bank for Reconstruction & Dev.**
- ▶ **IDA, International Development Association**
- ▶ **IFC, International Finance Corporation**
- ▶ **MIGA, Multicultural Investment Guarantee Agency**
- ▶ **ICSID, Internat. Ctr f. Settlement of Investment Disputes**
- ▶ **International Court of Justice**
- ▶ **US Government Agencies**

Figure 3b. A Tree Browse Window with drill-down to WBG and below

3.2 User interface and search 2

- Another interface: Show record for an organization
- The following records show just variant names

World Bank	
<i>permalink</i> :	http://lccn.loc.gov/n79043403
<i>Variant(s):</i>	<p>International Bank for Reconstruction and Development <i>Acronym</i> IBRD</p> <p>World Bank Group. World Bank</p> <p>Banque internationale pour la reconstruction et le développement <i>Acronym</i> B.I.R.D. ; BIRD</p> <p>Banque mondiale</p> <p>Mezhdunarodnyĭ bank dliã`rekonstruktsii i razvitiã` <i>Acronym</i> MBRR</p> <p>Internationale Bank für Wiederaufbau und Entwicklung <i>Acronym</i> IBWE</p> <p>Welt Bank</p> <p>Weltbank</p> <p>Banco Internacional de Reconstrucción y Fomento <i>Acronym</i> BIRF</p> <p>Banco Mundial</p> <p>hanākhān Lōk</p>

Figure 4a. World Bank and variants (LC Authorities, selected)

World Bank. Agriculture and Natural Resources Department	
<i>permalink:</i>	http://lccn.loc.gov/nr95045186
<i>Variant(s):</i>	AGR World Bank. Agriculture & Natural Resources Department World Bank. Agriculture and Natural Resources Dept.
<i>See also:</i>	World Bank. Rural Development Department Hierarchical superior: World Bank
<i>Found in:</i>	World Bank Group dir., May 1996: (Agriculture & Natural Resources Department (AGR))
Note the historical information	The World Bank website, Archives, viewed May 4, 2012: International standard archival authority record – Agriculture and Rural Development sector (Agriculture and rural development department, 2002-; Rural development department, 1997-2002; Agriculture and natural resources department 1993-1997)

Figure 4b. Authority record from LC

3.2 User interface and search 3

- Would also provide for standard faceted search with a search box and facets to limit results
- Organizations found can be shown
 - alphabetically
 - grouped by location, type, or other criterion
 - in their organization hierarchy context

3.2 User interface and search 3

- The **organization hierarchy can be used for hierarchic query expansion**. Examples:
 - Search for all documents **from any WBG member organization** dealing with **Uganda**
 - Search for all documents **from any WBG member organization** on **irrigation projects in Africa** (using hierarchic expansion for Location as well).
- **Organization name variants can be used for synonym expansion**

3.2 User interface and search 4

Organizations as the search target

For example, find **potential partners for a project in Africa**

Organization *<hasPurpose>* Economic development **AND**
Organization *<hasPurpose>* Africa

Would find::

- the WBG unit(s) dealing with Africa
- other units in the UN family
- the US Agency for International Development unit(s) dealing with Africa
- government units in other countries
- non-governmental organizations

3.3 System implementation

- **Unified system design treats authority and classification data for all kinds of entities following the same abstract scheme** — could be subjects, places, times, events, people, organizations and documents
 - **One system module displays any hierarchical structure** and handles all user interaction including type-ahead search. Inputs:
 - (1) a reference to the set of XML objects that represent all the entity instances to be included and their relationships and
 - (2) a list of relationship types that are considered hierarchical.
 - **One system module handles query expansion** – hierarchic and synonym – for any entity type.
- **Unified approach simplifies system development and gives a consistent user experience.**

4 Populating an Organization Authority Database

- **External sources**, such as
 - DBpedia <http://wiki.dbpedia.org/>
 - Library of Congress Name Authority File <http://id.loc.gov/authorities/names.html>.
 - Dun & Bradstreet
- **Internal sources**
- **User input**
- Existing sources require **mapping of relationship types**
- Merging from multiple sources requires **name matching** and **disambiguation**

DBpedia property (relationship type)	OAD schema relationship type
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<isa>
<http://dbpedia.org/ontology/type>	<isa>
<http://dbpedia.org/property/type>	<isa>
<http://dbpedia.org/property/companyName>	<hasName> NameStatus: LegalName
<http://www.w3.org/2000/01/rdf-schema#label>	<hasName> NameStatus: LegalName
<http://xmlns.com/foaf/0.1/name>.	<hasName> NameStatus: LegalName
<http://dbpedia.org/ontology/parentOrganisation>	<hasBroader>
<http://purl.org/dc/terms/subject>	<isAbout>
<http://dbpedia.org/property/purpose>	<hasPurpose>
<http://www.w3.org/2000/01/rdf-schema#comment>	<hasShortDescription>
<http://dbpedia.org/ontology/abstract>	<hasLongDescription>
<http://xmlns.com/foaf/0.1/homepage>	<hasWebAddress>
<http://dbpedia.org/ontology/owner>	<owns> REVERSE

Figure 5. Correspondence DBpedia and OAD schema. Some examples

5 Conclusions

- An enterprise wants to perform **powerful data analytics considering the complex interactions among many variables** to develop successful strategies and prevent costly operational mistakes.
- **Requires linking data across the many applications** in the entire enterprise and many external sources.
- **In turn requires consistent identifiers for core entity types:**
subjects/topics, diseases, procedures, organisms, chemical substances, products, types of costs/expenses, places, times/historical periods, events, people, organizations and documents
- **Solution: The unified approach to handling all kinds of authority data**, focusing on the common problems of
 - multiple names for the same thing and of
 - interacting with hierarchical structures.

5 Conclusions 2

- **Use general definitions of entity types** (classes) **and relationship types** (properties) with useful abstraction to **capture structural elements that are common to multiple domains.**
- This **logical analysis lays the foundation for**
 - **general software modules**, saving development effort
 - **a unified user experience.**
- **We applied these principles in a pilot system** to demonstrate their usefulness in a large organization with highly varied information requirements such as the World Bank Group.
- **So can you.**

Thank you

Questions?

dsoergel@buffalo.edu

www.dsoergel.com