

Progress Review on the Ontology Based Standards for
Sample Projects Identified at the 2009 Ontology
Summit: Ontologies for Data Standards in Support of e-
Gov Applications (e.g. Data.gov)

Brand Niemann, US EPA

Invited Expert to the W3C eGovernment Interest Group

October 8, 2009

Note: First part of this was used October 6th for the W3C/XBRL [Workshop](#) on
Improving Access to Financial Data on the Web

Abstract

- Data.gov version 1.0 has adopted the taxonomy from the Annual Statistical Abstract and Data.gov version 2.0 will probably adopt RDF as a data sharing - networking format. The work of [David Hay](#) to build an ontology for government data provides the basis for this work and our progress will be reported.
 - Reference http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2009_Communique#nid1WN3

Abstract

- The ISO 11179 standard is being evolved to include an ontology:
 - Kevin Keck will report on this in several months:
 - Reference: http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2009_Communique#nid1WN1

Data.gov Concepts of Operations

- Key Excerpt (September 4th DRAFT):
 - As semantically enabled data concepts are matured and pilots are successfully executed, the Data.gov team will provide specific guidance to help agencies implement semantic markup within their datasets. Specifically, the Data.gov team will not only leverage semantically enabled techniques within the Data.gov site, but will help agencies implement semantically enabled data within their datasets so that the datasets can be better leveraged not only by Data.gov but also by other end users of the data.
- What is your reaction to this?

Open Linked Data

- Key Concepts of RDF and OWL:
 - Works for structured, semi-structured, and unstructured information.
 - Metadata and data travel together.
 - Machine - processible.
 - Supports relational – like joins over the network.
 - Inferencing / reasoning (still in development for scalability).

Federal Segment Architecture Methodology Step 4: Principle 5*

Data (markup)	Business	Presentation
None	Silos	“One-Offs”
XML / XQuery	Collaboration	“Author Once-Use Many”
RDF / SPARQL	Relational-like Joins	Federation
RIF / SPIN	Rules	Logic
OWL	Reasoning	Inferencing

* [FSAM](#). Evolution based on September 2009 DRAFT Data.gov CONCOPs.

A Little History of How We Got Here

- [SOA CoP](#) and [SICOP](#) Conferences and Pilots: SOA, Semantic Interoperability, and XBRL , NIEM, etc.
 - Pilots for the 9th SOA for eGov Conference, April 5th, 2010
- [ISE EA Statement on RDF/OWL](#): Jeremy Warren, DoJ CTO, at [EA2009](#) agreed still the goal!
- [NSF Accounting Ontology Workshop 2008](#): Conceptual Diagram – Semantic Expressiveness Versus Benefit.
- [SemTech 2009](#): Diane and Dave Conversion of XBRL to RDF.
- [W3C eGov SIG](#): Series of Meetings and Notes.
- [SICoP Special Conference February 17th](#): Netherlands Demo and TQ - [SPIN](#) Demo.
- [DRM CoP Work on DRM 3.0 and Pilots](#) (see next slides):
 - [Google: Federal data population data](#) - two hits illustrate DRM 3.0 concepts - difference from Census hit - real data and reuseable (Web 2.0/3.0). Data Model Ontology for the Government (Hay).
- [Semantic Web Meetups](#):
 - [May 13th](#) in Washington, DC
 - 1000s working around the world on Open Linked Data

Data.gov Version 1 Taxonomy: Topics

Statistical Abstract of the United States: 2009 - Federal Data Web 2.0 Wiki Pilot - Windows Internet Explorer provided by EPA

http://federaldata.wik.is/Statistical_Abstract_of_the_United_States%3a_2009

Ids employment services

File Edit View Favorites Tools Help

Statistical Abstract of th... Federal SOA - Federal SOA Twitter / People who follow ... Cox High Speed Internet En... Network Centric Industry O...

8. Telephone and Internet Contacts

- Section 1. Population
- Section 2. Births, Deaths, Marriages, and Divorces
- Section 3. Health and Nutrition
- Section 4. Education
- Section 5. Law Enforcement, Courts, and Prisons
- Section 6. Geography and Environment
- Section 7. Elections
- Section 8. State and Local Government Finances and Employment
- Section 9. Federal Government Finances and Employment
- Section 10. National Security and Veterans Affairs
- Section 11. Social Insurance and Human Services
- Section 12. Labor Force, Employment, and Earnings
- Section 13. Income, Expenditures, Poverty, and Wealth
- Section 14. Prices
- Section 15. Business Enterprise
- Section 16. Science and Technology
- Section 17. Agriculture
- Section 18. Natural Resources
- Section 19. Energy and Utilities
- Section 20. Construction and Housing
- Section 21. Manufactures
- Section 22. Wholesale and Retail Trade
- Section 23. Transportation
- Section 24. Information and Communications
- Section 25. Banking, Finance, and Insurance
- Section 26. Arts, Recreation, and Travel
- Section 27. Accommodation, Food Services, and Other Services
- Section 28. Foreign Commerce and Aid
- Section 29. Puerto Rico and the Island Areas
- Section 30. International Statistics
- Section 31. Appendices

Appendix Ia. Guide to Sources of Statistics

Done Internet

http://federaldata.wik.is/Statistical Abstract of the United States%3a_2009

Data.gov Version 1 Taxonomy: Subtopics

The screenshot shows a web browser window with the URL [http://federaldata.wik.is/Statistical Abstract of the United States%3a_2009/Section 1. Population](http://federaldata.wik.is/Statistical_Abstract_of_the_United_States%3a_2009/Section_1._Population). The page title is "Section 1. Population - Federal Data Web 2.0 Wiki Pilot". The browser's address bar shows the URL and search engines like "ids employment services". The page content is organized into several sections:

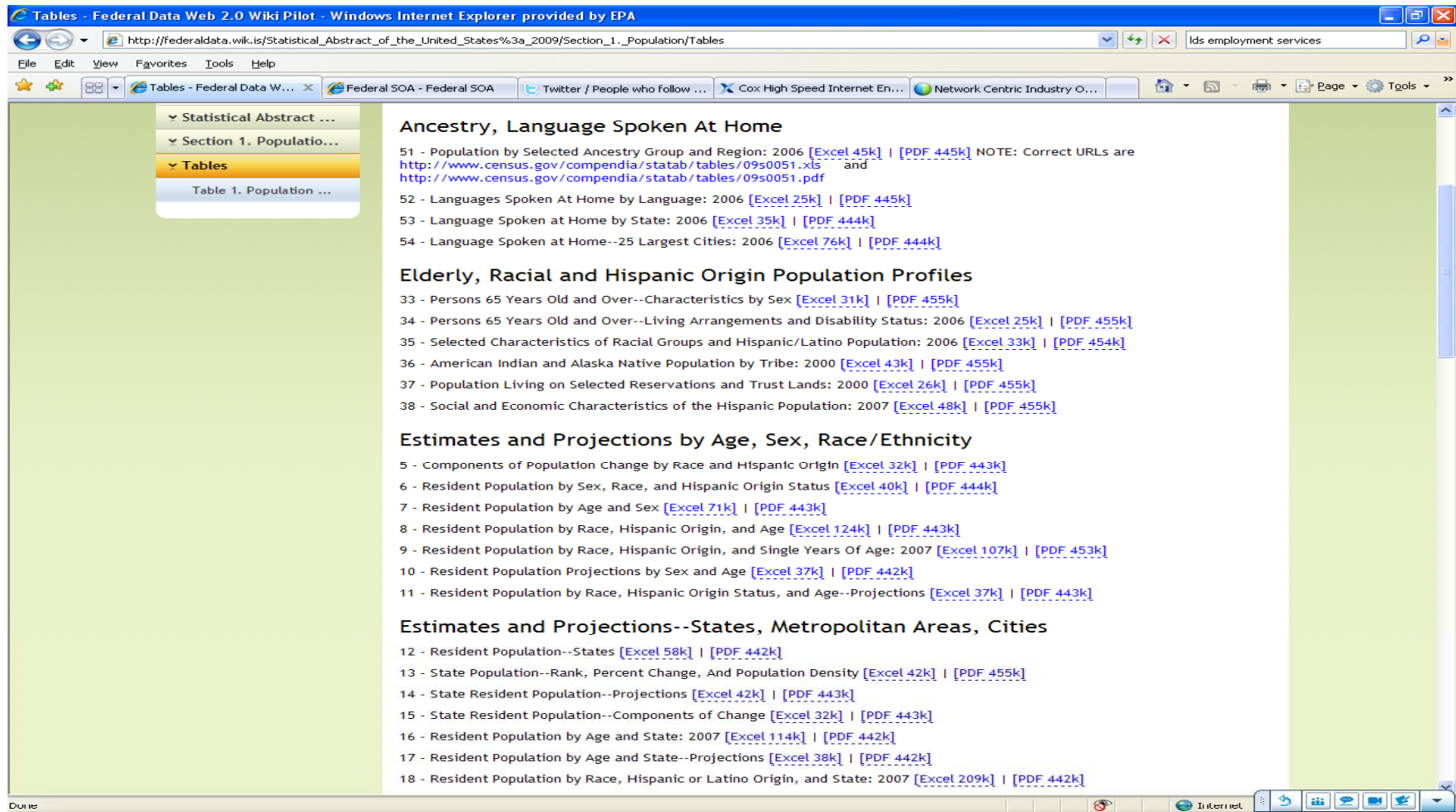
- Introduction**: This section presents statistics on the growth, distribution, and source of these data is the U.S. Census Bureau, which conducts a survey, a program of population estimates and projections, and...
- Decennial censuses**: The U.S. Constitution provides for a census of the population members of the House of Representatives among the states. Since the 1790 census, the census organization was a temporary one. Since the 1940 census, in addition to the complete count information, the Census Bureau has been compiling statistics on other population and housing characteristics.
- Current Population Survey (CPS)**: This is a monthly nationwide survey of a scientifically selected sample of the civilian population. The sample is located in 824 areas with coverage in every state. At the present time, about 60,000 occupied households are eligible for interview. Various reasons, unavailable for interview.

The right side of the page features a table of contents with numbered items:

1. Introduction
2. Decennial censuses
3. Current Population Survey (CPS)
4. American Community Survey (ACS)
5. Population estimates and projections
6. Immigration
7. Metropolitan and micropolitan areas
8. Urban and rural
9. Residence
10. Race
11. Hispanic population
12. Foreign-born and native populations
13. Mobility status
14. Living arrangements
15. Householder
16. Family
17. Subfamily
18. Married couple

http://federaldata.wik.is/Statistical Abstract of the United States%3a_2009/Section 1. Population

Data.gov Version 1 Taxonomy: Tables



http://federaldata.wik.is/Statistical_Abstract_of_the_United_States%3a_2009/Section_1._Population/Tables

Data.gov Version 1 Taxonomy: Data Elements

Table 1. Population and Area: 1790 to 2000 - Federal Data Web 2.0 Wiki Pilot - Windows Internet Explorer provided by EPA

http://federaldata.wik.is/Statistical_Abstract_of_the_United_States%3a_2009/Section_1._Population/Tables/Table_1._Population_and_Area%3a_1790_to_2000

1860 (June 1)		31,443,321	10.6	8,251,445	35.6
1870 (June 1)	\2	39,818,449.2	11.2	8,375,128	26.6
1880 (June 1)		50,189,209	14.2	10,370,760	26.0
1890 (June 1)		62,979,766	17.8	12,790,557	25.5
1900 (June 1)		76,212,168	21.5	13,232,402	21.0
1910 (April 15)		92,228,496	26.0	16,016,328	21.0
1920 (January 1)		106,021,537	29.9	13,793,041	15.0
1930 (April 1)		123,202,624	34.7	17,181,087	16.2
1940 (April 1)		132,164,569	37.2	8,961,945	7.3
1950 (April 1)		151,325,798	42.6	19,161,229	14.5
1960 (April 1)		179,323,175	50.6	27,997,377	18.5
1970 (April 1)		203,302,031	57.5	23,978,856	13.4
1980 (April 1)	\3	226,542,199	64.0	23,240,168	11.4
1990 (April 1)	\4	248,718,302	70.3	22,176,103	9.8
2000 (April 1)	\6	281,424,603	79.6	32,706,301	13.1

Symbol:
X Not applicable.

Footnotes:

- 1 Data for 1790 to 1980 cover inland water only. Data for 1990 comprise Great Lakes, inland, and coastal water. Data for 2000 comprise Great Lakes, inland, territorial, and coastal water.
- 2 Revised to include adjustments for underenumeration in southern states; unrevised number is 38,558,371 (10.9 per square mile).
- 3 Total population count has been revised since the 1980 census publications. Numbers by age, race, Hispanic origin, and sex have not been corrected.
- 4 The April 1, 1990, census count includes count question resolution corrections processed through December 1997, and does not include adjustments for census coverage errors.
- 5 Data reflect corrections made after publication of the results.
- 6 Reflects modifications to the Census 2000 population as documented in the Count Question Resolution program.

Source:
U.S. Census Bureau, 2000 Census of Population and Housing, Population and Housing Counts, Series PHC-3-1, United States Summary; Notes and Errata, 2000 SF/01-ER ; <<http://www.census.gov/prod/cen2000/notes/errata.pdf>>; Areas of the United States: 1940; Area data for 1990: unpublished data from TIGER®; and Davis, Warren; personal correspondence; U.S. Census Bureau; 23 June 2006.

For More Information:
<http://www.census.gov/population/www/>

[http://federaldata.wik.is/Statistical Abstract of the United States%3a_2009/Section 1. Population/Tables/Table 1. Population and Area%3a_1790 to 2000](http://federaldata.wik.is/Statistical_Abstract_of_the_United_States%3a_2009/Section_1._Population/Tables/Table_1._Population_and_Area%3a_1790_to_2000)

David Hay: Data Model Patterns

- Presentations to the Library of Congress, March 24, 2009, by:
 - David C. Hay, Capgemini Financial Services, Washington, DC.
 - Describing the World: Data Model Patterns
 - Part One: [The Enterprise Model](#)
 - Part Two: [Metadata](#)
- Presentation of DRM Suggestions on April 10th and Invitation to April 28th Activity:
 - Same as above.

David Hay: Data Model Patterns – A Metadata Map

- Book:
 - Morgan Kaufman / Elsevier, 2006, 406 pages.
 - Also Data Model Patterns: Conventions of Thought, Dorset House, 1996, and [2007 SemTech Conference](#): Data Modeling and OWL: Two Ways to Structure Data (new book in process on ontologies for data modeling) (See next slide).
- Uses an Architecture Framework to organized the broad body of knowledge (chapters in book):
 - Based on John Zachman's 1987 and 1992 Enterprise Architecture Framework (Zachman, 1987; and Sowa and Zachman, 1992).
- The Zachman Framework consist of a matrix in which the rows represent perspectives different people have on a information technology project and the columns represent what they are seeing from each perspective.
 - See next slide.

David Hay: Data Model Patterns – A Metadata Map

- The Rows:
 - Objectives / Scope
 - Planner's View
 - Enterprise Model
 - Business Owner's View
 - Model of Fundamental Concepts
 - Architect's View
 - Technology Model
 - Designer's View
 - Detailed Representation
 - Builder's View
 - Functioning System
 - Inventory View
- The Columns:
 - Data (What)
 - Activities (How)
 - Locations (Where)
 - People (Who)
 - Time (When)
 - Motivation (Why)

Source: John Zachman's 1987 and 1992 Enterprise Architecture Framework (Zachman, 1987; and Sowa and Zachman, 1992).

David Hay: Data Modeling and OWL: Two Ways to Structure Data

- The world of data modeling and database design has developed over nearly 40 years within the information technology industry. The world of semantics and ontologies has developed over about 2500 years within the worlds of philosophy, linguistics, and more recently, artificial intelligence. The time has finally come to bring these worlds together. This presentation is an attempt to take the first steps towards doing that. The presentation will describe data modeling in its various forms and use an example to show the relationships between its concepts and those of the ontology language OWL. Covered will be the relationships between entity classes and OWL classes, attributes and datatype properties, and relationships and object properties. This will include a discussion of taxonomies and alternative classification approaches. The presentation will focus on both the similarities and the fundamental differences between the two approaches.

David Hay: Library of Congress Presentations

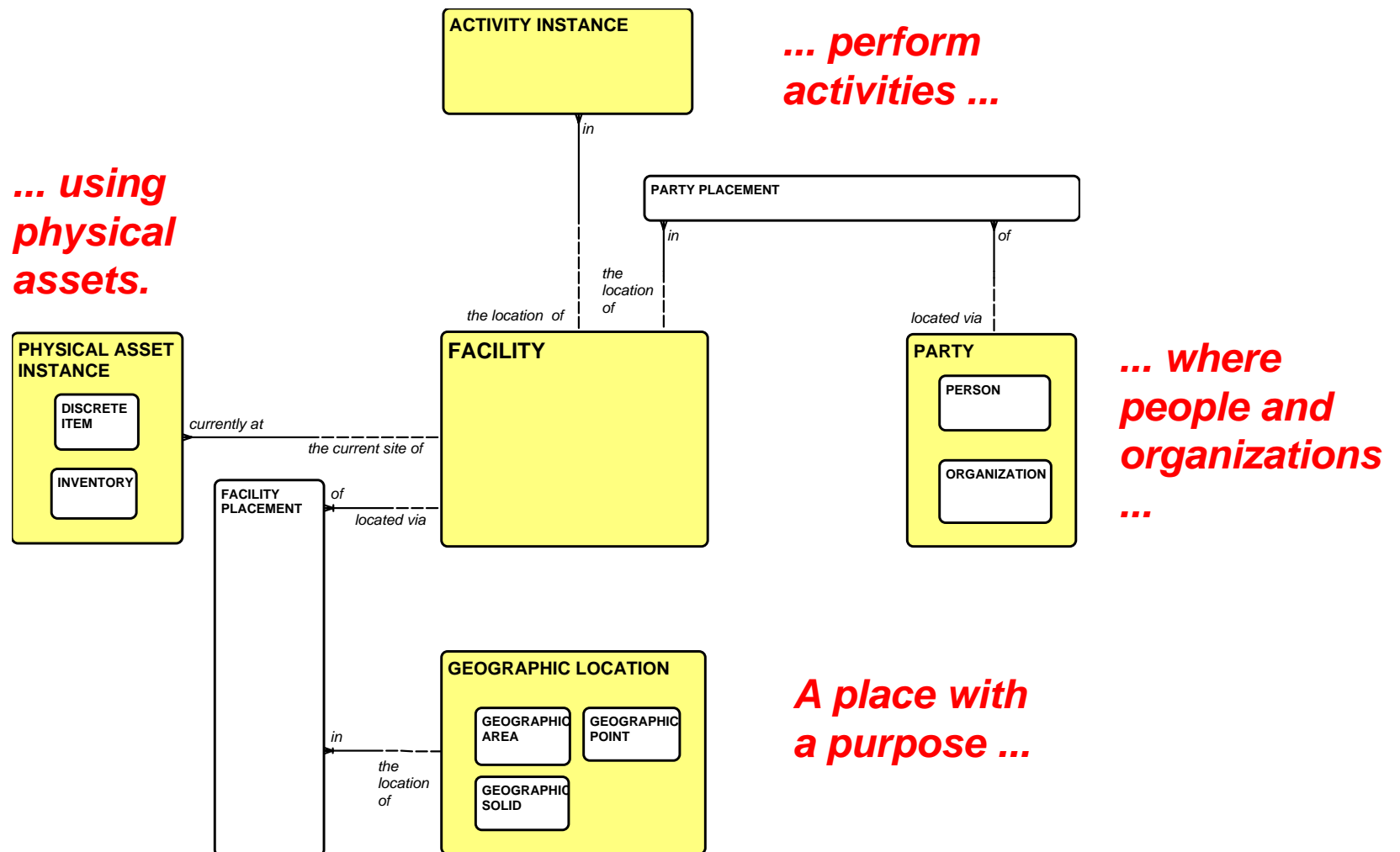
- Part One:
 - Abstraction Level 1: The Generic Enterprise Model
 - People and Organizations (Who?)
 - Geography (Where?)
 - Physical Assets (What?)
 - Activities and Events (How?)
 - Timing Attributes and Entity Classes (When?)
 - Abstraction Level 0: The Template
 - Things
- Part Two:
 - Abstraction Level 1 (continued): Metadata (Documents and Accounting)
 - Composites (Why?)
 - Sites and Facilities – see next slide
 - Contracts

Note: Metadata models are descriptions that can apply to the entire enterprise model or any part of it.

Facilities Overview . . .

- Facility:
 - A place with a purpose.
 - Used to locate resources:
 - People and organizations
 - Activities
 - Materials (Physical Assets)

Facility is in the Middle . . .



Some Next Steps

- Refactor the Statistical Abstract Taxonomy into an Ontology:
 - Topics (30), Sub-topics (about 300), Data Tables (about 1,500), and Data Elements (about 15,000).
- Complete David Hay's Data Model Ontology for the Government:
 - Also see Ralph Hodgson's Ontology for Government Organizations; and
 - OMG's Data Model Ontology for the FSAM.

Contact Information

- niemann.brand@epa.gov
- 202-564-9491
- <http://semanticcommunity.net>
- <http://www.twitter.com/bniemannsr>