

Research on US Federal Government Handling of Data

Dr. Brand Niemann

Director and Senior Enterprise Architect – Data Scientist

Semantic Community

<http://semanticcommunity.info/>

AOL Government Blogger

<http://gov.aol.com/bloggers/brand-niemann/>

January 21, 2013

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard/CKAN_Japan_Open_Data

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard/A_DATA_METI_Concept (Password Protected)

Research Specification

- Analysis of current open government initiatives and programs in the U.S.A./research on rules governing data and information disclosure:
 - The research should cover the latest trends in regulations and operational rules governing the disclosure of federal agency and other government-held data and information, including the handling of highly sensitive data related to national security, in light of the Obama administration's open government policy, and the latest information available on the operation of the federal cloud.
 - Rules governing public access to data via DATA.gov.
- NIEM content and implementation:
 - Investigate the content and implementation of NIEM.
 - Investigate and analyze the realities of current usage by U.S. government departments and agencies and overseas:
 - What are the actual content and usage by U.S. federal institutions, state and local governments, foreign governments, and private-sector corporations in the case of DHS and NIEM?
 - In particular, conduct an in-depth survey on the current usages and its issues/challenges by New York City Government, Pennsylvania State Government, New York State Government. It will be appreciated if you can add more good example of any State Government.

Big Picture

- We all have:
 - Content that we now call “Big Data”
 - Technology that we now call “BYOX” (bring your own device, data, etc.)
- Governments have data:
 - Statistical (collected to answer questions)
 - Open (just happens)
 - Classified (national security and proprietary work)
- Governments have concerns:
 - Privacy (personal information)
 - Security (need to know)
 - “Mosaic effect” (aggregation can reveal a privacy or security breach)

Big Solution

- Governments need:
 - Chief Data Officers
 - Teams of Data Scientists and Statisticians
 - Return on Investment
- Governments should:
 - Work with the right data
 - Work with the right people
 - Work on the right projects

My Examples

- Current:
 - Japan Statistical Yearbook 2012
 - CKAN Japan Open Government Data
 - The US District of Columbia Data Catalog and 311 Message Service
- Future (proposed):
 - Traffic Monitoring with GPS on Cabs in Tokyo
 - MetaTags for Classified Data Before Encryption
 - Add facets to metadata that can be searched

Research Specification Response 1

- Analysis of current open government initiatives and programs in the U.S.A./research on rules governing data and information disclosure:
 - The research should cover the latest trends in regulations and operational rules governing the disclosure of federal agency and other government-held data and information, including the handling of highly sensitive data related to national security, in light of the Obama administration's open government policy, and the latest information available on the operation of the federal cloud.
 - GENERAL RESPONSE: http://semanticcommunity.info/Information_Sharing_Environment#2012_Annual_Report and http://semanticcommunity.info/Emerging_Technology_SIG_Big_Data_Committee/Cloud_Computing_AND_Big_Data_Forum_and_Workshop_January_15-17_2013
 - Rules governing public access to data via DATA.gov.
 - GENERAL RESPONSE: <http://www.data.gov/data-policy> and <http://www.data.gov/privacy-policy>

Research Specification Response 2

- NIEM content and implementation:
 - Investigate the content and implementation of NIEM.
 - Investigate and analyze the realities of current usage by U.S. government departments and agencies and overseas:
 - What are the actual content and usage by U.S. federal institutions, state and local governments, foreign governments, and private-sector corporations in the case of DHS and NIEM?
 - In particular, conduct an in-depth survey on the current usages and its issues/challenges by New York City Government, Pennsylvania State Government, New York State Government. It will be appreciated if you can add more good example of any State Government.
- GENERAL RESPONSE:
 - [http://semanticcommunity.info/National Information Exchange Model/Assessment Report](http://semanticcommunity.info/National_Information_Exchange_Model/Assessment_Report)
 - [http://semanticcommunity.info/Information Sharing Environment#2012 Annual Report](http://semanticcommunity.info/Information_Sharing_Environment#2012_Annual_Report)
 - [http://semanticcommunity.info/Collaboration and Transformation SIG#Responsible Information Sharing: Engaging Industry to Improve Standards-Based Acquisition .26 Interoperability](http://semanticcommunity.info/Collaboration_and_Transformation_SIG#Responsible_Information_Sharing:_Engaging_Industry_to_Improve_Standards-Based_Acquisition_.26_Interoperability)
 - MY NOTE: Do not know or more States than New York and Pennsylvania. Oracle may be a larger and better implementer of NIEM than “NIEM” itself!

Research Specification Response 3

The screenshot shows a web browser window with the Semantic Community website. The page title is "A DATA METI Concept". The content includes a table of contents with sections like "U.S. Federal Government Handling of Data and Messages (NIEM) for a DATA METI Concept", "Background Research", and "Results". A note is overlaid on the page: "MY NOTE: This is password protected and under review by IBT and METI."

A DATA METI Concept
Modified 09:48, 24 Nov 2012 by Admin | Page History

1. [U.S. Federal Government Handling of Data and Messages \(NIEM\) for a DATA METI Concept](#)

- 1.1. [Abstract](#)
- 1.2. [1. Introduction](#)
- 1.3. [2. Open Data in Japan](#)
 - 1.3.1. [Prime Minister Noda on Open Data](#)
 - 1.3.1.1. [Figure Headquarters at the Prime Minister of Japan and His Cabinet.](#)
 - 1.3.2. [Background](#)
 - 1.3.2.1. [Figure Summary of the e-Government Open Data Strategy \(DRAFT\)](#)
 - 1.3.3. [Current Structure of Promoting Open Data in Japan](#)
 - 1.3.3.1. [Figure Current Structure of Promoting Open Data in Japan](#)
 - 1.3.4. [Issues & Challenges](#)
 - 1.3.4.1. [Figure Issues and Challenges](#)
 - 1.3.5. [Research Needs](#)
 - 1.3.5.1. [The scope and specifics of the data to be disclosed.](#)
 - 1.3.5.2. [History](#)
 - 1.3.5.3. [Survey Questions](#)
 - 1.3.5.4. [Goals](#)
- 1.4. [3. Background Research](#)
 - 1.4.1. [Table Framework for The Evolution of Working with Open Government Data \(DRAFT\)](#)
- 1.5. [4. Results](#)
 - 1.5.1. [Requirements](#)
 - 1.5.2. [Steps](#)
 - 1.5.2.1. [Step 1](#)
 - 1.5.2.2. [Step 2](#)
 - 1.5.2.3. [Step 3](#)
 - 1.5.2.4. [Step 4](#)
 - 1.5.2.5. [Step 5](#)
 - 1.5.2.6. [Step 6](#)
 - 1.5.2.7. [Step 7](#)
 - 1.5.2.8. [Step 8](#)
 - 1.5.2.9. [Step 9](#)

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard/A_DATA_METI_Concept

Research Specification Response 4






- Governments need:
 - Chief Data Officers (IBT and METI)
 - Teams of Data Scientists and Statisticians (Me to Start)
 - Return on Investment (My Consulting Fee)
- Governments should:
 - Work with the right data (Statistical and Open)
 - Work with the right people (Data Scientists)
 - Work on the right projects (Pilot Dashboards)

My 5-Step Method

- So what I like to do to illustrate (data science) and explain (data journalism) is the following (like a recipe):
 - Put the Best Content into a Knowledge Base (e.g. MindTouch*)
 - The Japan Statistical Yearbook 2012
 - Put the Knowledge Base into a Spreadsheet (Excel*)
 - Linked Data to Subparts of the Knowledge Base
 - Put the Spreadsheet into a Dashboard (Spotfire*)
 - Data Integration and Interoperability Interface
 - Put the Dashboard into a Semantic Model (Excel*)
 - Data Dictionaries and Models
 - Put the Semantic Model into Dynamic Case Management (Be Informed*)
 - Structured Process for Updating Data in the Dashboard

* Examples of tools used.

To Get to 5-Stars With Open Data

Star	Definition	Example / Tool*
	Make your stuff available on the Web (whatever format) under an open license	This Story / MindTouch
	Make it available as structured data (e.g., Excel instead of image scan of a table)	Spreadsheet / Excel
	Use non-proprietary formats (e.g., CSV instead of Excel)	Table / MindTouch and Spotfire
	Use URIs to identify things, so that people can point at your stuff	Table of Contents / MindTouch and Spotfire
	Link your data to other data to provide context	Table / MindTouch and Spotfire

* Examples of tools used.

Source of Star and Definition: <http://www.w3.org/DesignIssues/LinkedData.html>

Japan Statistical Yearbook 2012

The screenshot shows a web browser window displaying the Semantic Community website. The address bar shows the URL: semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard. The page features a search bar at the top right and a navigation menu at the top left. The main content area displays the title "A Japan METI Open Data Dashboard" with a "Was this article helpful?" button. Below the title is a list of links for various statistical data sets and reports, including "1. Story", "2. Steps", "3. Spotfire Dashboard", "4. Research Notes", "5. Overview of the 2011 Great East Japan Earthquake", "6. Japan Statistical Yearbook 2012", and "7. Notes for Users".

Semantic Community

Home My Page Recent changes Reports Tools

Signed in: Admin Sign out

Edit page New page More

Page Notifications Off

Semanticcommunity.info > A Japan METI Open Data Dashboard Table of contents

A Japan METI Open Data Dashboard

Modified 09:52, 24 Nov 2012 by Admin | Page History

Was this article helpful?

1. [Story](#)
2. [Steps](#)
 - 2.1. [Step 1](#)
 - 2.2. [Step 2](#)
 - 2.3. [Step 3](#)
 - 2.4. [Step 4](#)
 - 2.5. [Step 5](#)
 - 2.6. [Step 6](#)
 - 2.7. [Step 7](#)
 - 2.8. [Step 8](#)
 - 2.9. [Step 9](#)
 - 2.10. [Step 10](#)
 - 2.11. [Step 11](#)
 - 2.12. [Step 12](#)
 - 2.13. [Step 13](#)
 - 2.14. [Step 14](#)
 - 2.15. [Step 15](#)
 - 2.16. [Step 16](#)
3. [Spotfire Dashboard](#)
4. [Research Notes](#)
5. [Overview of the 2011 Great East Japan Earthquake](#)
6. [Japan Statistical Yearbook 2012](#)
 - 6.1. [Introduction](#)
 - 6.2. [Contents](#)
 - 6.3. [Errata](#)
 - 6.4. [Back numbers](#)
7. [Notes for Users](#)
 - 7.1. [Introduction](#)
 - 7.2. [Annex Major Changes in Statistical Tables](#)
 - 7.2.1. [Tables newly added](#)
 - 7.2.2. [Revised tables](#)

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard

METI Open Data Dashboard-Spotfire

Navigation and Metadata

Source: <http://www.esri.com/products/data/japan-shp/>

MY NOTE: Shape file version 6.2 did not have English City Names, but version 7.1 does. But one still has to make the Population Table City Names conform to the CITY_ENG names or use the Japanese version of Spotfire.

Damage To Paddy Fields and Paddy Harvest by Prefecture (1990-2009)

Prefecture	Prefecture (2)	Total Damag...	Total Quantit...	Meteor
茨城	Ibaraki-shi	108.00	40.00	
栃木	Tochigi-shi	144.00	33.00	
群馬	Gumma-MISSING	36.00	10.00	
埼玉	Saitama-shi M...	89.00	17.00	
千葉	Chibashi MULT...	79.00	23.00	
東京	TokyoMISSING	0.30	0.00	

Area in thousands of hectares and Quantity in thousands of tons

Damage To Paddy Fields and Paddy Harvest (1990-2009)

Year	Total Damag...	Total Quantit...	Meteorologi...	Meteor.
1990	2988	678	1248	
1995	3352	819	1440	
2000	2455	417	1068	
2005	3337	637	1713	
2007	4209	851	2520	
2008	3119	518	1620	
2009	3478	808	2014	

Quantity in thousands of tons

47 of 47 rows | 0 marked | 12 columns | Damage To Paddy Fields and Paddy Harvest by Prefecture (1990-2009)

Show all downloads...

<https://silverspotfire.tibco.com/us/library#/users/bniemann/Public?METIOpenDataDashboard-Spotfire>

CKAN Japan Open Government Data

The screenshot shows a web browser window displaying the CKAN Japan Open Government Data page on the Semantic Community website. The browser's address bar shows the URL: semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard/CKAN_Japan_Open_Data. The page features a search bar at the top right and a navigation menu with options like Home, My Page, Recent changes, Reports, and Tools. The main content area is titled "CKAN Japan Open Data" and includes a "Was this article helpful?" feedback button. A list of 19 numbered links is provided, covering various data sets and reports such as "Yokohama FY 2012 budget", "Profit of the year", "Tuki population Aizuwakamatsu", and "Results of 'emergency survey on the status of efforts to the problem of bullying'".

Search: Search [Advanced](#)

Signed in: [Admin](#) Sign out

[Edit page](#) [New page](#) [More](#) [Page Notifications Off](#)

CKAN Japan Open Data

Modified [01:44, 17 Nov 2012](#) by [Admin](#) | [Page History](#)

[Was this article helpful?](#) [Feedback](#)

- [Story](#)
- [Spotfire Dashboard](#)
- [Research Notes](#)
- [CKAN Japan](#)
 - [Example of data that is provided here](#)
 - [Local government](#)
 - [Sabae](#)
 - [Disaster prevention](#)
 - [Nagareyama Chiba](#)
 - [Aizu-Wakamatsu City, Fukushima Prefecture](#)
 - [test group](#)
- [Data Sets](#)
 - [Yokohama FY 2012 budget - another major business line departments](#)
 - [Profit of the year "opened bullying hotline counseling 24 hours"](#)
 - [Tuki population Aizuwakamatsu](#)
 - [Location data Aizuwakamatsu](#)
 - [ARTADE2GENE](#)
 - [Japanese version of DBpedia](#)
 - [Department in charge of local government information in Fukui Prefecture](#)
 - [Administrative information \(version Fukui\)](#)
 - [gakokujinn](#)
 - [Support group for foreign residents in Japan](#)
 - [Graph cancer information database](#)
 - [Download the spreadsheet-Cancer Information](#)
 - [Aggregate national hospital cancer registry cooperation base hospitals and cancer clinics Cancer Information](#)
 - [Barrier-free map of all](#)
 - [Trends in child truancy](#)
 - [Utilization of Children's Rights, Ministry of Justice No. 110](#)
 - [Actual violence in high school Shocho](#)
 - [Research on issues such as student guidance on problem behavior of students](#)
 - [Results of "emergency survey on the status of efforts to the problem of bullying" \(Overview\)](#)

http://semanticcommunity.info/A_Japan_METI_Open_Data_Dashboard/CKAN_Japan_Open_Data

The US District of Columbia Data Catalog and 311 Message Service

The screenshot shows a web browser window displaying the Semantic Community website. The page title is "The Open Data Handbook" and it was modified on 08:46, 5 Dec 2012 by Admin. The page contains a table of contents with the following items:

- 1. [Story](#)
- 2. [Slides](#)
 - 2.1. [Slides 1](#)
 - 2.2. [Slide 2](#)
 - 2.3. [Slide 3](#)
 - 2.4. [Slide 4](#)
- 3. [Spotfire Dashboard](#)
- 4. [Research Notes](#)
- 5. [Data Catalog](#)
 - 5.1. [Create Your Own Visualizations using DC datasets](#)
 - 5.2. [Popular Downloads](#)
 - 5.3. [Custom Downloads](#)
 - 5.4. [Live Data Feeds](#)
 - 5.5. [See it on Google Maps](#)
 - 5.6. [Browse Catalog](#)
 - 5.7. [Category: All categories](#)
- 6. [The Open Data Handbook](#)
- 7. [Introduction](#)
 - 7.1. [Target Audience](#)
 - 7.2. [Credits](#)
 - 7.2.1. [Credits and Copyright](#)
 - 7.2.1.1. [Contributing authors](#)
 - 7.2.1.2. [Existing sources directly used](#)
 - 7.2.1.3. [Other sources](#)

- 8. [Why Open Data?](#)
- 9. [What is Open Data?](#)
- 9.1. [What is Open?](#)
- 9.2. [What Data are You Talking About?](#)
- 10. [How to Open up Data](#)
- 10.1. [Choose Dataset\(s\)](#)
 - 10.1.1. [Asking the community](#)

MY NOTE: I did not find a Japanese City with a Data Catalog so I used a US City with an excellent one I had done.

<http://semanticcommunity.info/A Japan METI Open Data Dashboard/The Open Data Handbook>

Conclusion

- The Open Government Data/Linked Data Initiatives in the US (Data.gov) and elsewhere started with the wrong data. I recommended that Data.gov start with the best US government data from the Federal Statistical Agencies (e.g. US Census Bureau) and make it the standard practice of high quality data and metadata rendered in the new linked open data way. Experience has shown that to be the case with the statistical community calling Data.gov and similar efforts essentially 'IT projects.' The solution is to first render a countries high quality statistical data in the new way and then try to render the open data the same way as much as possible. I have done this for the US and Europe, and now for Japan.