# Research on US Federal Government Handling of Data

Dr. Brand Niemann

Director and Senior Enterprise Architect – Data Scientist

Semantic Community

http://semanticommunity.info/

**AOL Government Blogger** 

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

January 21, 2013

http://semanticommunity.info/A Japan METI Open Data Dashboard

http://semanticommunity.info/A Japan METI Open Data Dashboard/CKAN Japan Open Data

http://semanticommunity.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:

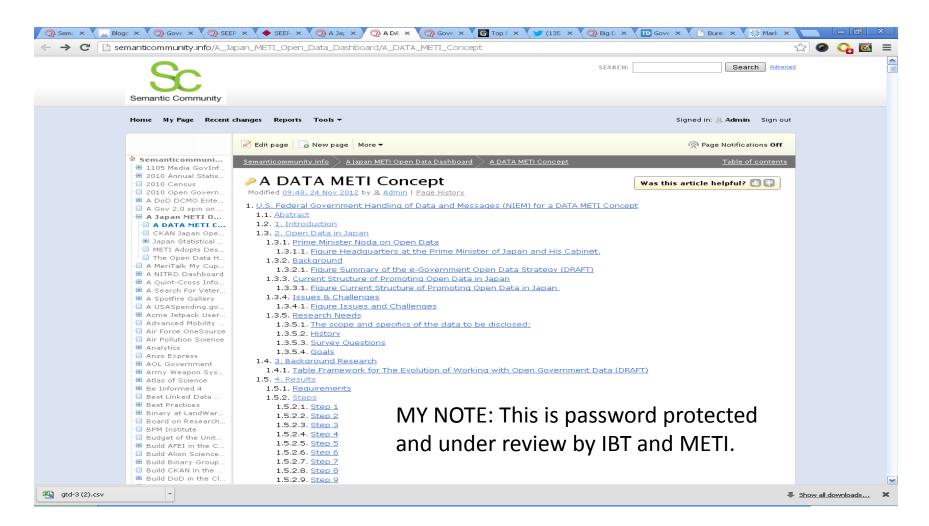
- Japan Statistical Yearbook 2012
- CKAN Japan Open Government Data
- The US District of Columbia Data Catalog and 311
   Message Service
- Future (proposed):
  - Traffic Monitoring with GPS on Cabs in Tokyo
  - MetaTags for Classified Data Before Encryption
    - Add facets to metadata that can be searched

- 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://semanticommunity.info/Information Sharing Environment#2012 Annual Report and
       http://semanticommunity.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: <a href="http://www.data.gov/data-policy">http://www.data.gov/privacy-policy</a>

- 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://semanticommunity.info/National Information Exchange Model/Assess ment Report
- http://semanticommunity.info/Information\_Sharing\_Environment#2012\_Annual\_Report
- http://semanticommunity.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!



#### 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

<sup>\*</sup> Examples of tools used.

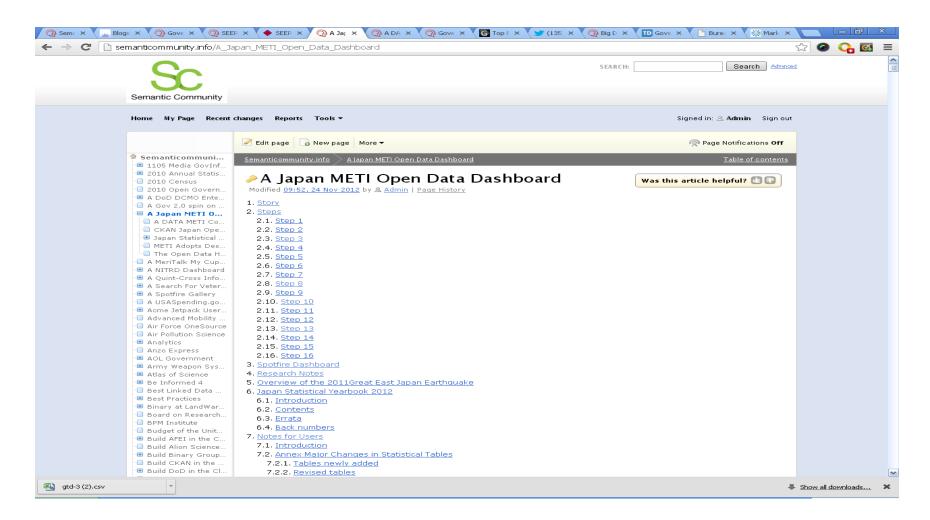
## To Get to 5-Stars With Open Data

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)	<u>Spreadsheet</u> / 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

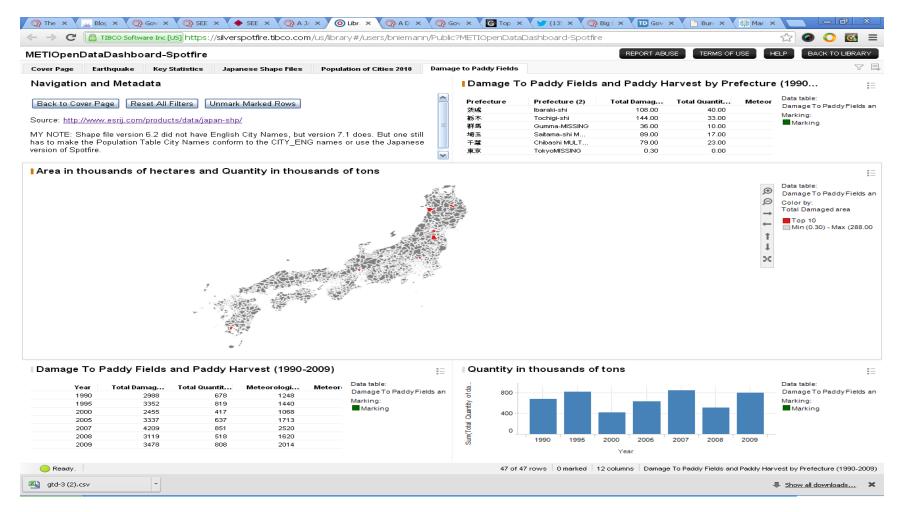
Source of Star and Definition: <a href="http://www.w3.org/DesignIssues/LinkedData.html">http://www.w3.org/DesignIssues/LinkedData.html</a>

<sup>\*</sup> Examples of tools used.

## Japan Statistical Yearbook 2012

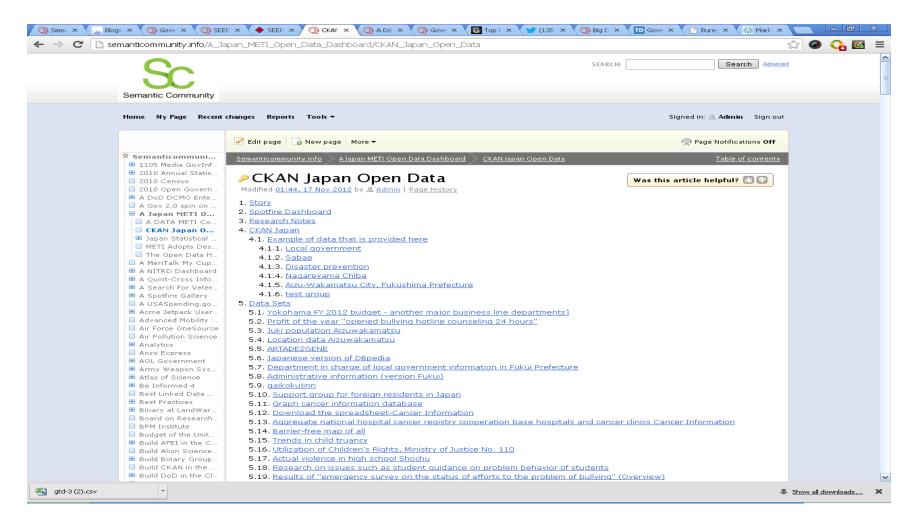


#### METI Open Data Dashboard-Spotfire

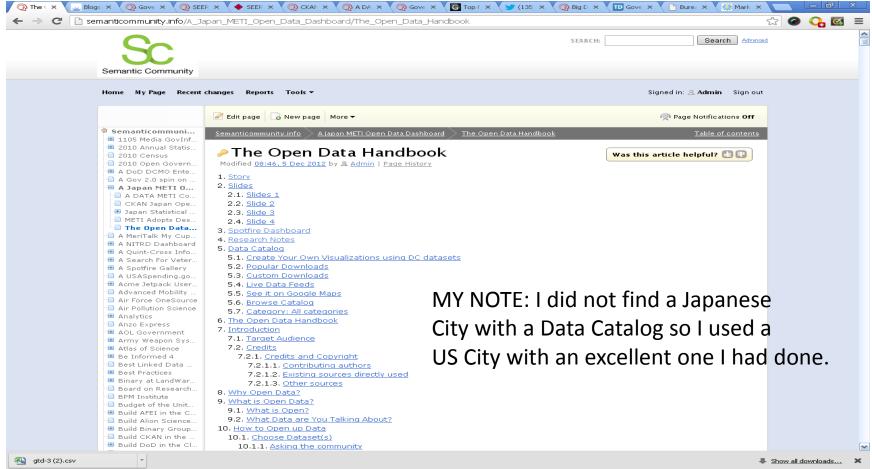


https://silverspotfire.tibco.com/us/library#/users/bniemann/Public?METIOpenDataDashboard-Spotfire

#### CKAN Japan Open Government Data



## The US District of Columbia Data Catalog and 311 Message Service



http://semanticommunity.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.