Week 9

Open Data

Outline

- Compelling Reasons for Open Data
- Issues & challenges in adoption
- Guest speaker: Mr. Lo Yoong Khong (Cluster Director, Government Infocomm Governance Division, and Justice & Law Cluster, IDA)

Definition of Open Data

- Broadly defined as making data & information produced or commissioned by government freely usable, reusable & redistributable by anyone

Compelling Reasons for Open Data

Desire to Open Up More Govt Data

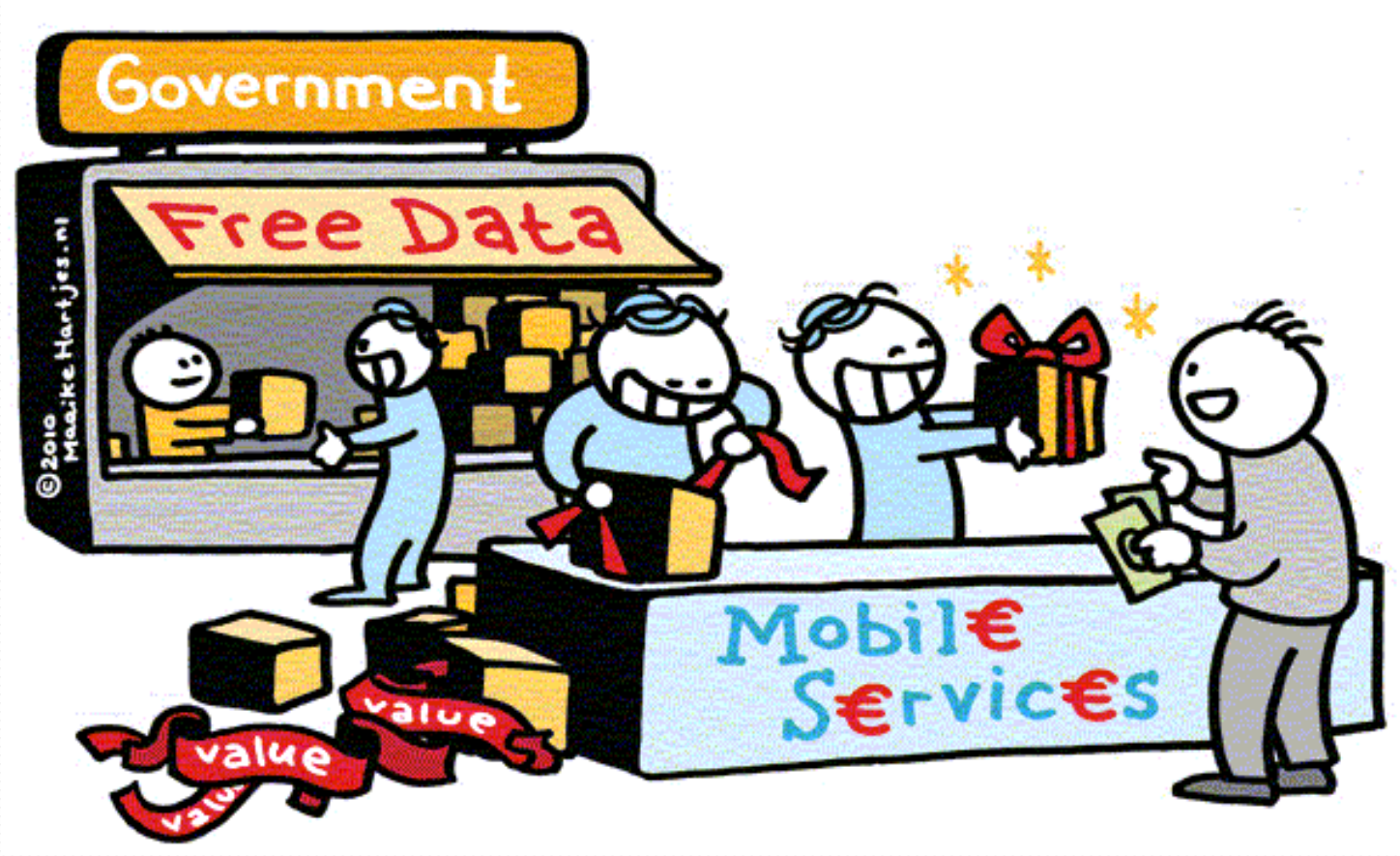
"We recognize the power of tapping into the ingenuity of the American people *and recognize that* government doesn't have a monopoly on the best ideas *or always have the best idea on finding an innovative path to solving the toughest problems the country faces. By* democratizing data and making it available *to the public and private sectors ... we can tap into that ingenuity."*

Vivek Kundra
US Federal CIO
20 May 2009 in interview by Nextgov.com

Vivek Kundra
(former) US Federal CIO
on data.gov

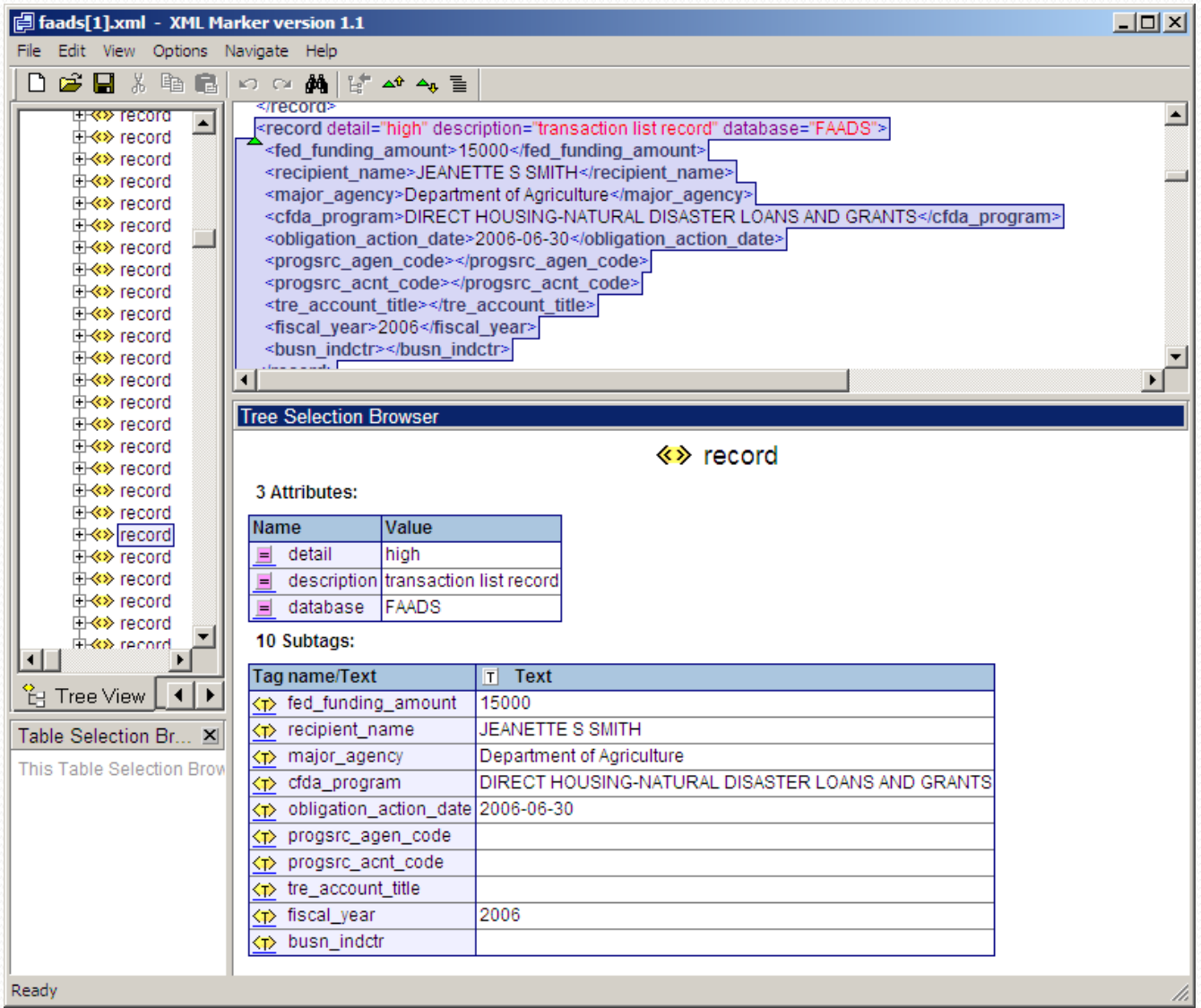
Information is power. By "democratizing data," ordinary citizens have the ability to shift the balance of power in positive ways that can encourage innovative ideas to be developed into practical goods and services. Washington, DC does not have a monopoly on the best ideas. The public has the ability to innovate.

Free Data for e-Government

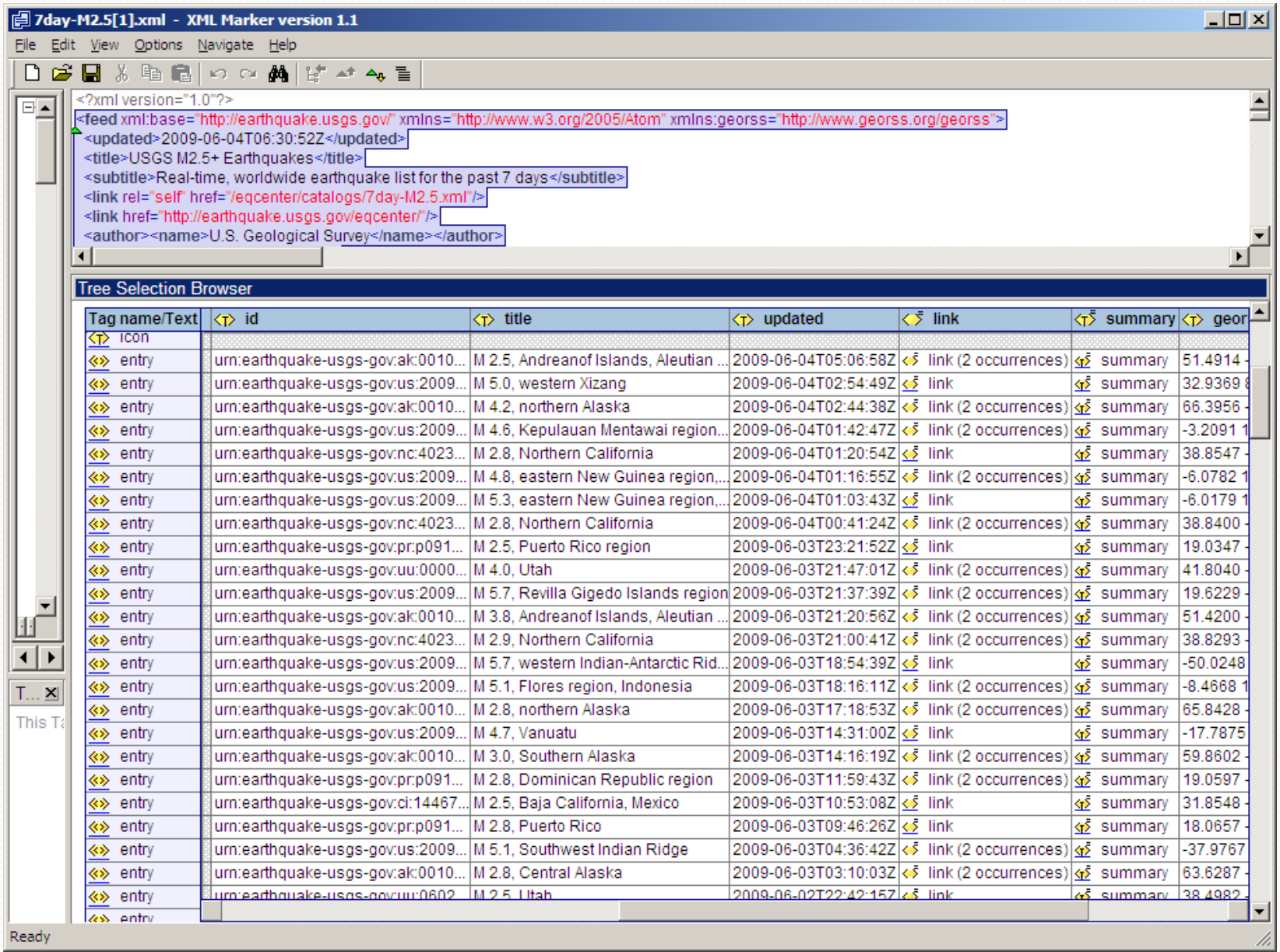


twitpic

USASpending.gov



US Data.gov


7day-M2.5[1].xml - XML Marker version 1.1
File Edit View Options Navigate Help
<?xml version="1.0"?>
<feed xml:base="http://earthquake.usgs.gov/" xmlns="http://www.w3.org/2005/Atom" xmlns:georss="http://www.georss.org/georss">
<updated>2009-06-04T06:30:52Z</updated>
<title>USGS M2.5+ Earthquakes</title>
<subtitle>Real-time, worldwide earthquake list for the past 7 days</subtitle>
<link rel="self" href="/eqcenter/catalogs/7day-M2.5.xml"/>
<link href="http://earthquake.usgs.gov/eqcenter/"/>
<author><name>U.S. Geological Survey</name></author>
Tree Selection Browser
Tag name/Text | id | title | updated | link | summary | geor
entry | urn:earthquake-usgs-gov:ak:0010... | M 2.5, Andreanof Islands, Aleutian ... | 2009-06-04T05:06:58Z | link (2 occurrences) | summary | 51.4914
entry | urn:earthquake-usgs-gov:us:2009... | M 5.0, western Xizang | 2009-06-04T02:54:49Z | link | summary | 32.9369
entry | urn:earthquake-usgs-gov:ak:0010... | M 4.2, northern Alaska | 2009-06-04T02:44:38Z | link (2 occurrences) | summary | 66.3956
entry | urn:earthquake-usgs-gov:us:2009... | M 4.6, Kepulauan Mentawai region... | 2009-06-04T01:42:47Z | link (2 occurrences) | summary | -3.2091
entry | urn:earthquake-usgs-gov:nc:4023... | M 2.8, Northern California | 2009-06-04T01:20:54Z | link | summary | 38.8547
entry | urn:earthquake-usgs-gov:us:2009... | M 4.8, eastern New Guinea region,... | 2009-06-04T01:16:55Z | link (2 occurrences) | summary | -6.0782
entry | urn:earthquake-usgs-gov:us:2009... | M 5.3, eastern New Guinea region,... | 2009-06-04T01:03:43Z | link | summary | -6.0179
entry | urn:earthquake-usgs-gov:nc:4023... | M 2.8, Northern California | 2009-06-04T00:41:24Z | link (2 occurrences) | summary | 38.8400
entry | urn:earthquake-usgs-gov:pr:p091... | M 2.5, Puerto Rico region | 2009-06-03T23:21:52Z | link | summary | 19.0347
entry | urn:earthquake-usgs-gov:uu:0000... | M 4.0, Utah | 2009-06-03T21:47:01Z | link (2 occurrences) | summary | 41.8040
entry | urn:earthquake-usgs-gov:us:2009... | M 5.7, Revilla Gigedo Islands region | 2009-06-03T21:37:39Z | link (2 occurrences) | summary | 19.6229
entry | urn:earthquake-usgs-gov:ak:0010... | M 3.8, Andreanof Islands, Aleutian ... | 2009-06-03T21:20:56Z | link (2 occurrences) | summary | 51.4200
entry | urn:earthquake-usgs-gov:nc:4023... | M 2.9, Northern California | 2009-06-03T21:00:41Z | link (2 occurrences) | summary | 38.8293
entry | urn:earthquake-usgs-gov:us:2009... | M 5.7, western Indian-Antarctic Rid... | 2009-06-03T18:54:39Z | link | summary | -50.0248
entry | urn:earthquake-usgs-gov:us:2009... | M 5.1, Flores region, Indonesia | 2009-06-03T18:16:11Z | link (2 occurrences) | summary | -8.4668
entry | urn:earthquake-usgs-gov:ak:0010... | M 2.8, northern Alaska | 2009-06-03T17:18:53Z | link (2 occurrences) | summary | 65.8428
entry | urn:earthquake-usgs-gov:us:2009... | M 4.7, Vanuatu | 2009-06-03T14:31:00Z | link | summary | -17.7875
entry | urn:earthquake-usgs-gov:ak:0010... | M 3.0, Southern Alaska | 2009-06-03T14:16:19Z | link (2 occurrences) | summary | 59.8602
entry | urn:earthquake-usgs-gov:pr:p091... | M 2.8, Dominican Republic region | 2009-06-03T11:59:43Z | link (2 occurrences) | summary | 19.0597
entry | urn:earthquake-usgs-gov:ci:14467... | M 2.5, Baja California, Mexico | 2009-06-03T10:53:08Z | link | summary | 31.8548
entry | urn:earthquake-usgs-gov:pr:p091... | M 2.8, Puerto Rico | 2009-06-03T09:46:26Z | link | summary | 18.0657
entry | urn:earthquake-usgs-gov:us:2009... | M 5.1, Southwest Indian Ridge | 2009-06-03T04:36:42Z | link (2 occurrences) | summary | -37.9767
entry | urn:earthquake-usgs-gov:ak:0010... | M 2.8, Central Alaska | 2009-06-03T03:10:03Z | link (2 occurrences) | summary | 63.6287
Ready

Neighbourhood Statistics

Neighbourhood Statistics

The data map displays the data for your selected table

Should you have any problems accessing the map, please contact info@statistics.gov.uk

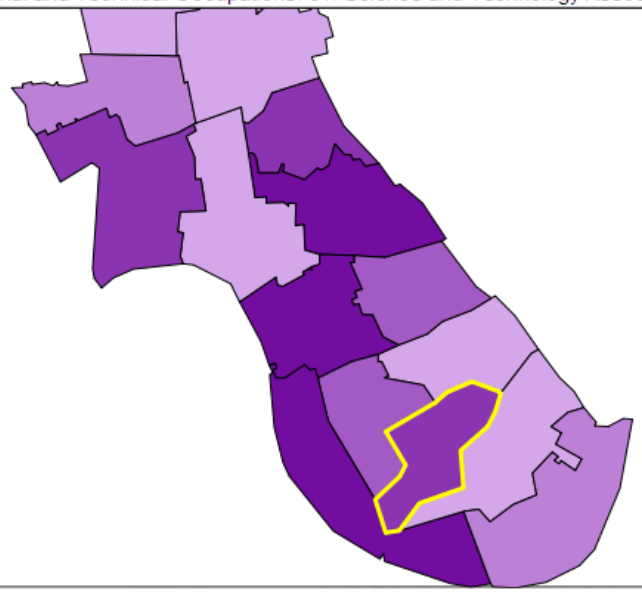
Close Window

Map | Advanced Tabs >>

Print view

2. Professional Occupations: 22. Health Professionals (Count, Persons, Apr01)

All People (Count, Persons, Apr01)
1. Managers and Senior Officials (Count, Persons, Apr01)
1. Managers and Senior Officials: 11. Corporate Managers (Count, Persons, Apr01)
1. Managers and Senior Officials: 12. Managers and Proprietors in Agriculture and Services (Count, Persons, Apr01)
2. Professional Occupations (Count, Persons, Apr01)
2. Professional Occupations: 21. Science and Technology Professionals (Count, Persons, Apr01)
2. Professional Occupations: 23. Teaching and Research Professionals (Count, Persons, Apr01)
2. Professional Occupations: 24. Business and Public Service Professionals (Count, Persons, Apr01)
3. Associate Professional and Technical Occupations (Count, Persons, Apr01)
3. Associate Professional and Technical Occupations: 31. Science and Technology Associate Professionals (Count, Persons, Apr01)



Colour

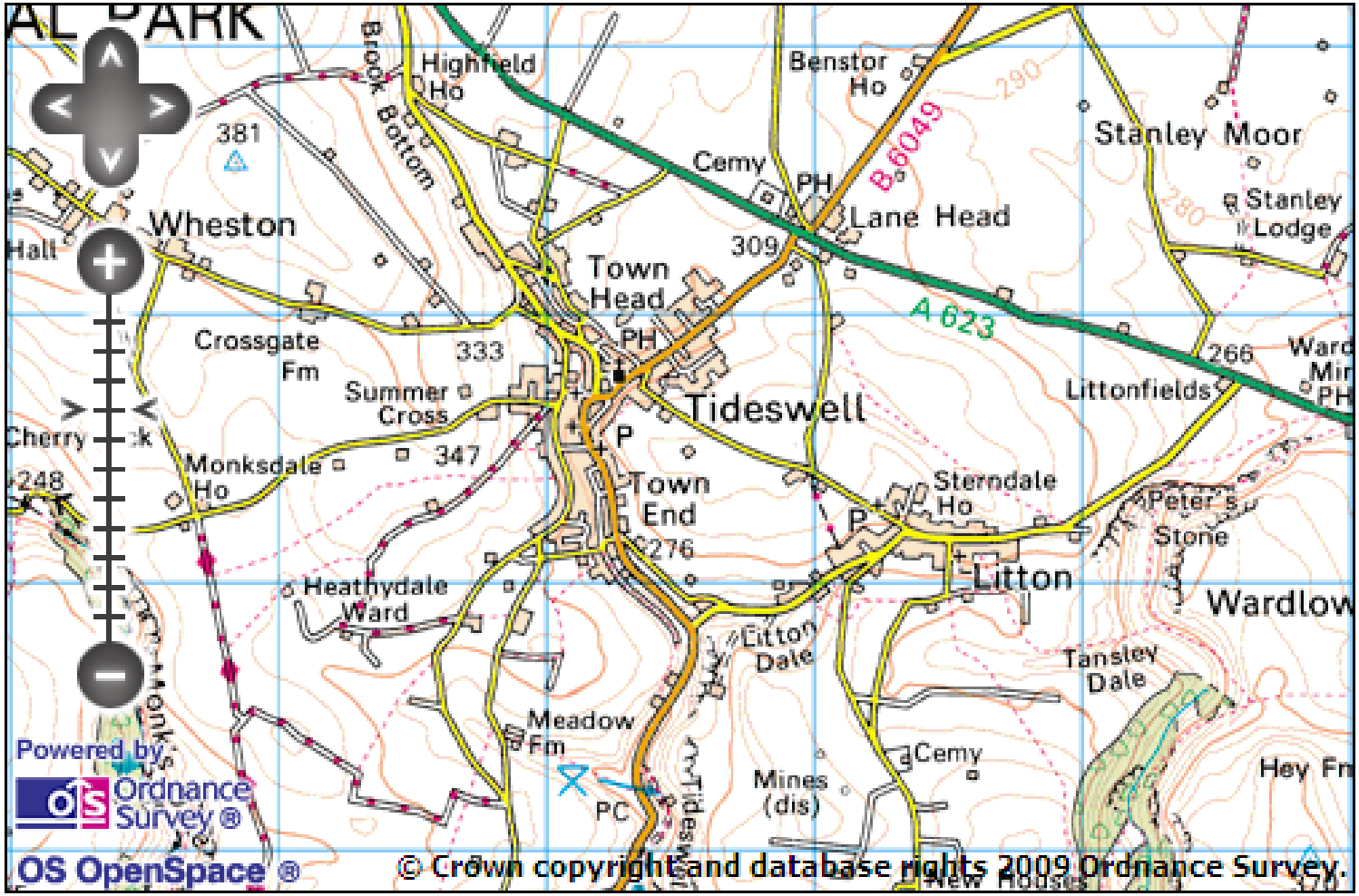
Hammersmith and Fulham (London Borough) 1,214

Caution! Check the legend to understand the shading. Ideally use rates and percentages instead of counts.
More advice

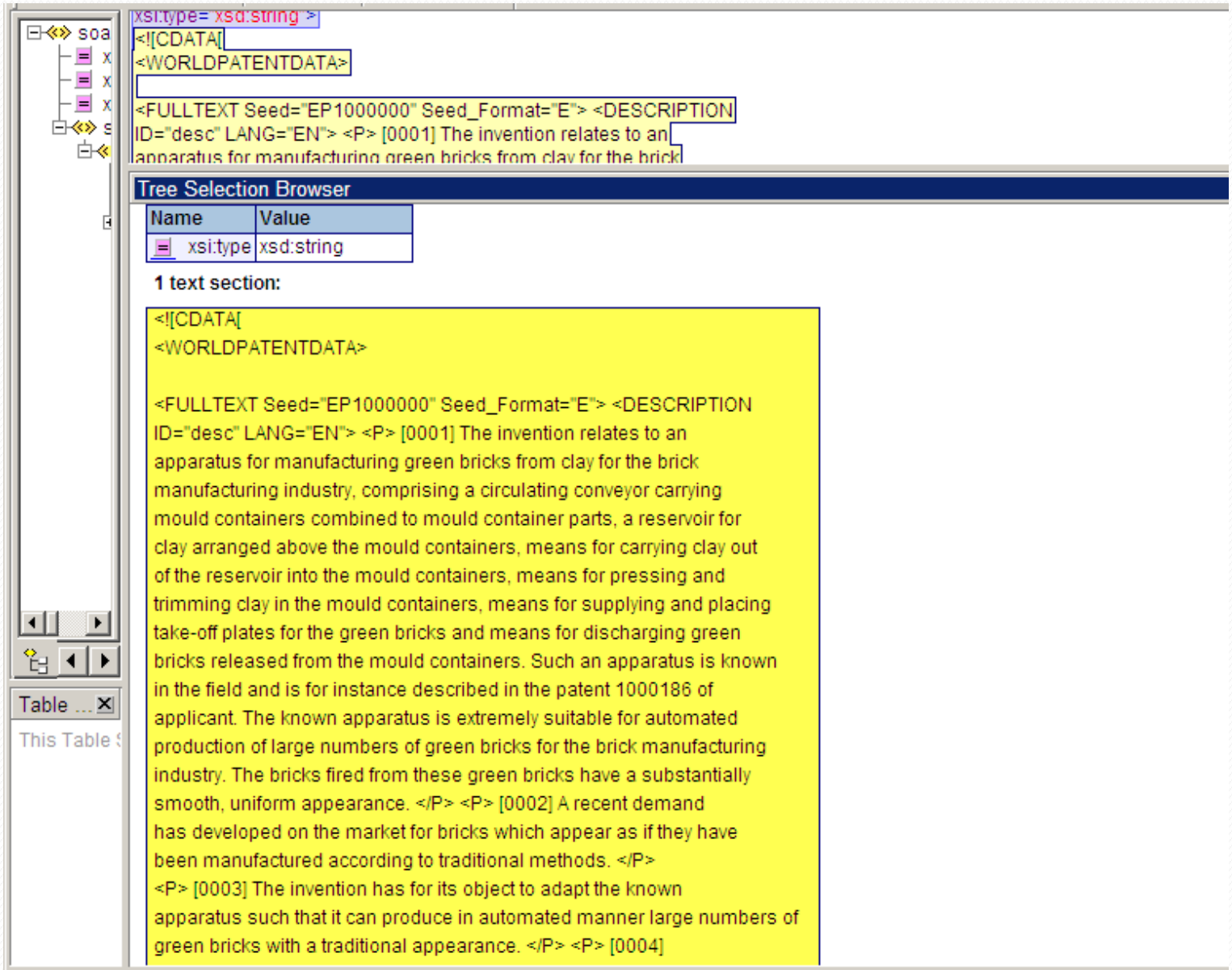
Footnotes

Show background map

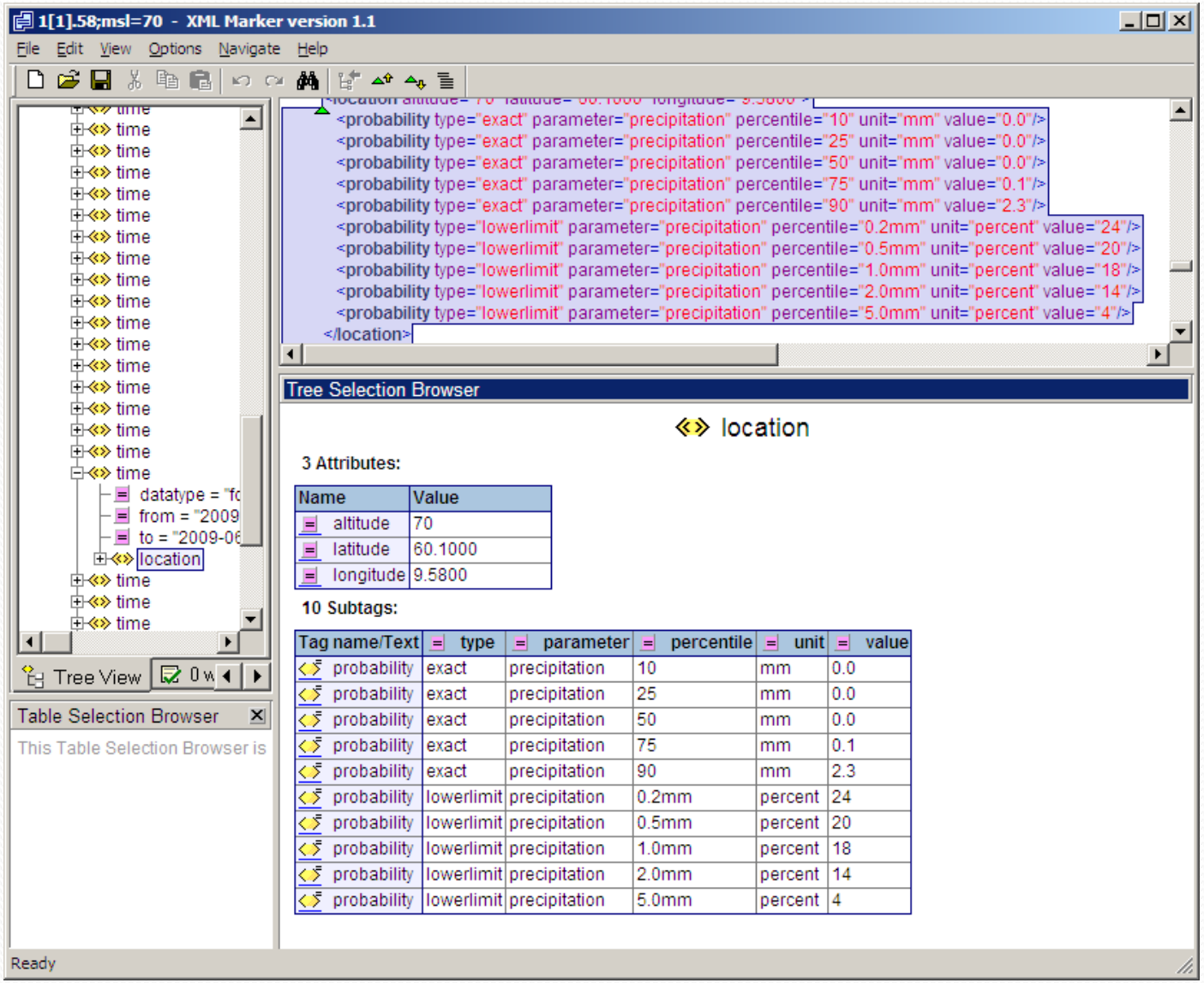
UK Ordnance Survey Maps



European Patent Office



Norway Meteorological Institute



Microsoft OGDI

Interact with the data you selected. Enter filter expression into a text box (text edit area) below and click Run. If you leave the Filter Expression box blank, your query will return all of the data in the selected data set.

Entity Set:	GasStations
Base Query:	http://ogdi.cloudapp.net/v1/dc/GasStations/
Filter Expression: ($filter)	(additional $filter parameters...)

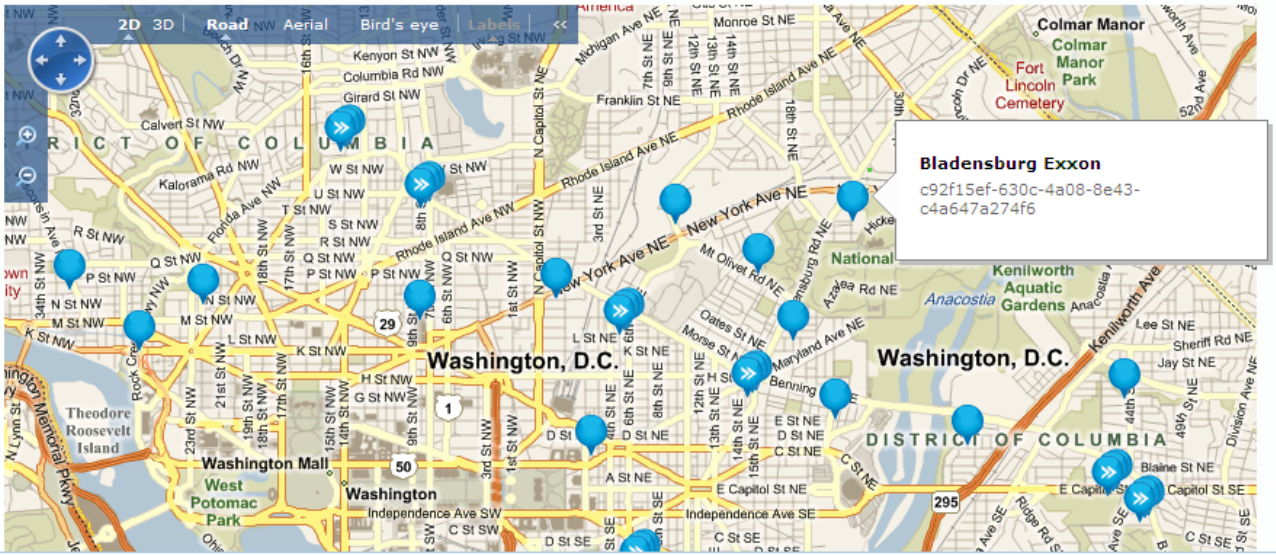
Run

Data View | Map View

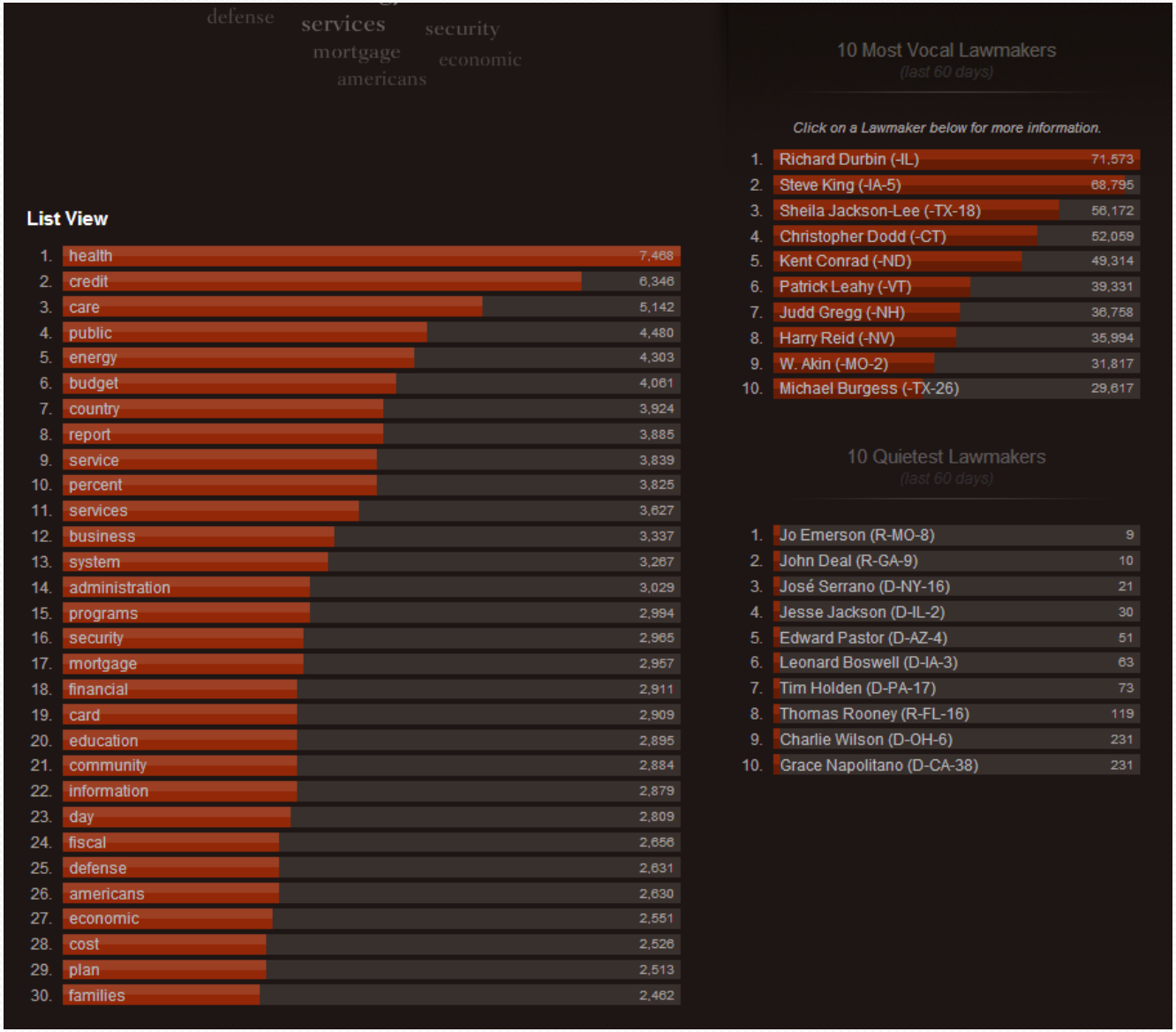
Sample Code

Results

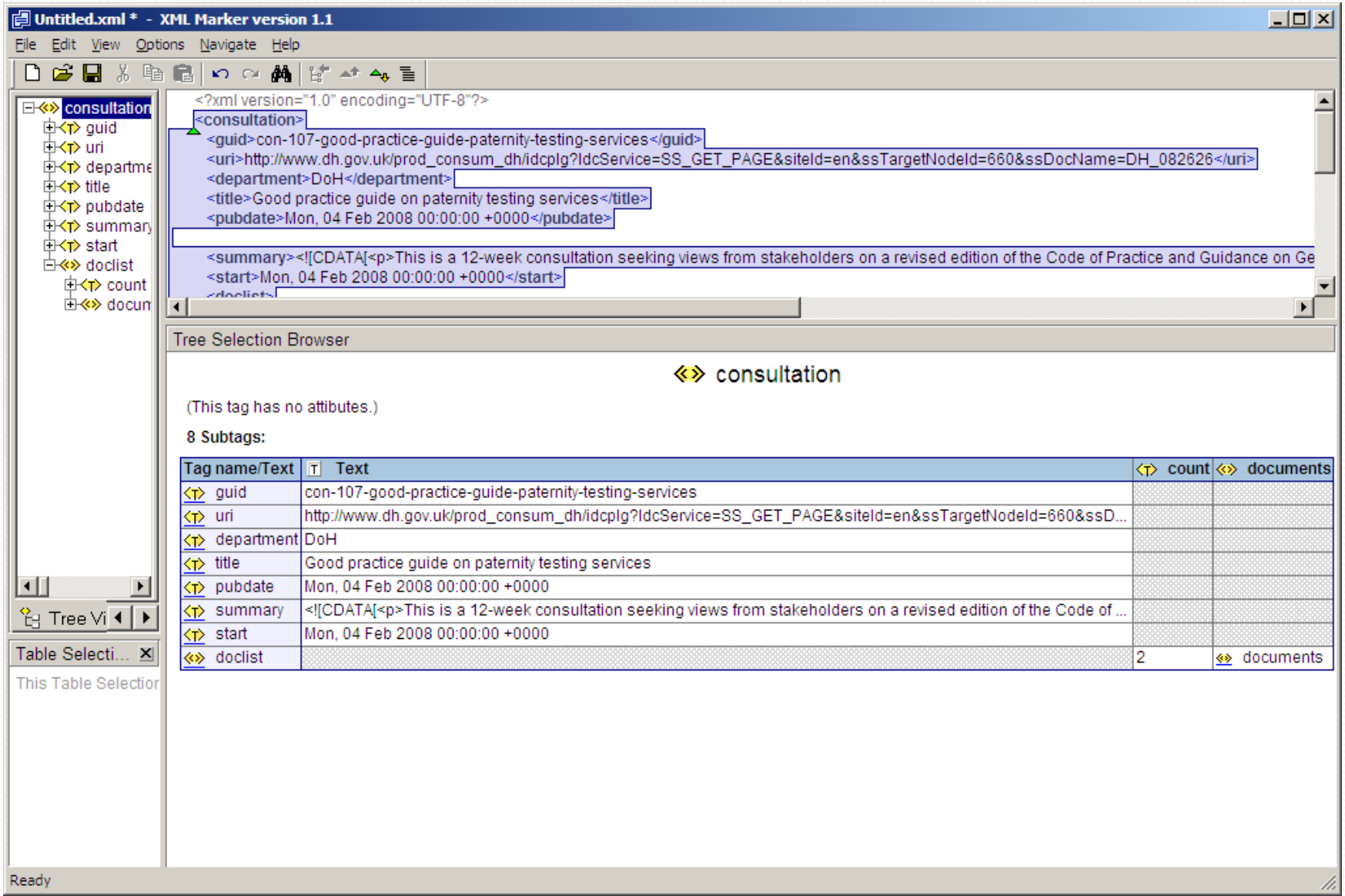
Full Query URL: http://ogdi.cloudapp.net/v1/dc/GasStations/?&format=kml
(Click to view results as XML/KML)



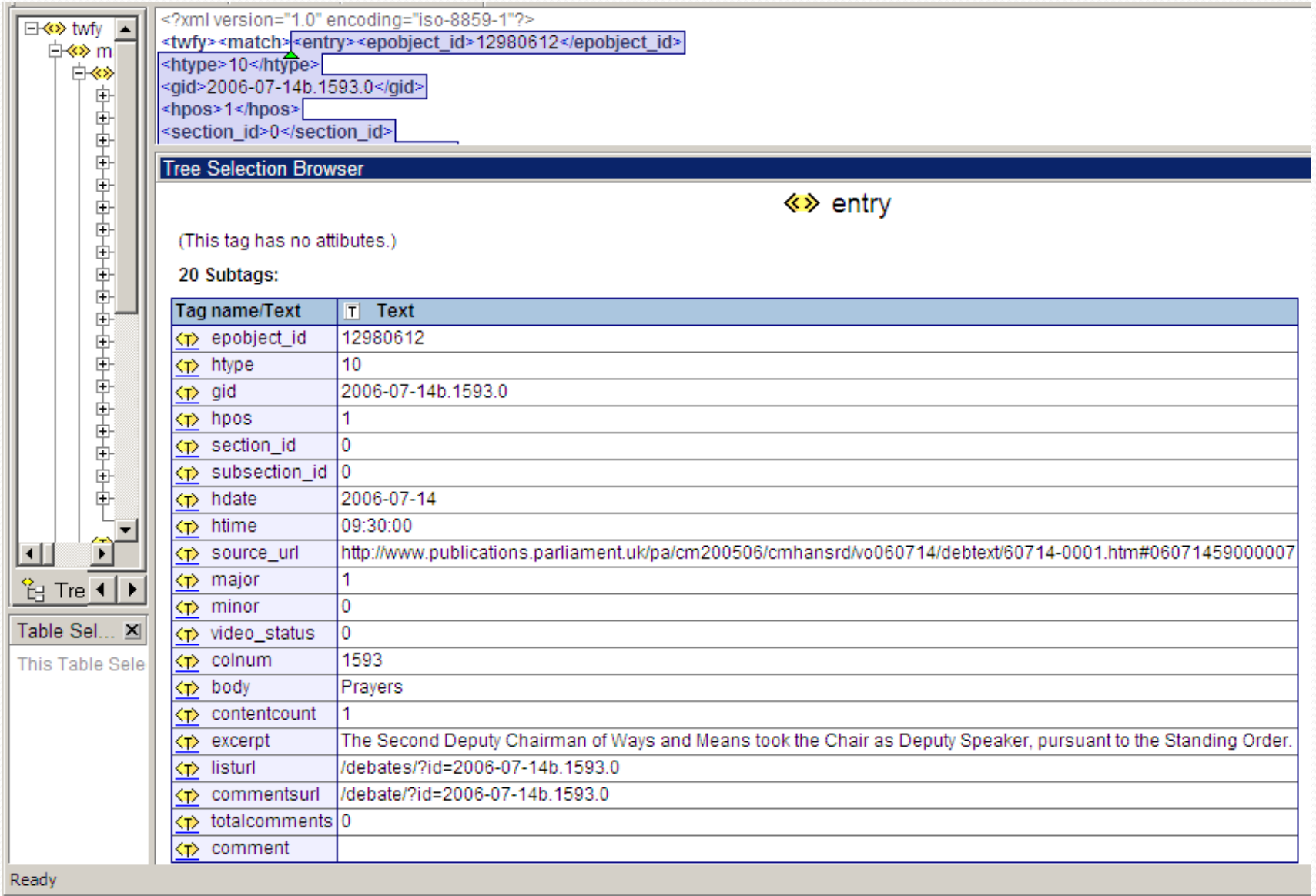
Sunlight Foundation



TellThemWhatYouThink.org



TheyWorkForYou.com



Open Data Everywhere

- US federal govt launched data.gov on 21 May 2009
  - Followed by data.gov.uk
  - … other governments in HK, Thailand, Singapore (20 June 2011)
- Recent developments are driven by govt at city & municipal level
  - Have unique understanding of specific problems which Open Data can help to solve for their communities
  - Are primary data custodian of many useful datasets
    - Real-time public transportation data
    - Environment Conservancy status within city & public facilities info

Key Considerations

- Key to unlocking hidden value of Open Data
  - Apps developers
- Need to recognize developers who contribute innovative apps
  - Many developer centric events in US every month
    - Hackathon
    - Hackday
    - Code-a-thon
    - Apps camp
    - Data camp

Benefits

- New opportunities for private sector to collaborate more closely with government to benefit community
- Co-creation of new economic & social value through collaboration between government & private sector
- Apps developers play a valuable role in getting info to citizens
- Citizens ultimately benefit from govt info

Findings

- City governments are directly engaging app developers thru numerous developer-centric events to promote use of Open Data to create innovative apps & services
- Governments are re-thinking traditional grant mechanism to promote innovation
  - Tapping on contests to motivate & accelerate innovations
- Mash-up of Open Data with private sector data will offer fresh perspectives on problems that are unavailable with govt data alone

Crowd-sourcing & Open Data

- Supporting & incorporating crowd-sourced data into Open Data repository
  - Takes paradigm shift from govt looking to tap into this new source of data

Example: Asthmapolis

- Uses medication sensors that reliably determine time & location when an inhaler is used
- Sensors send that data to Asthmapolis server
- Info allows city government to explore asthma causing hotspots
  - Reveals patterns of inhaler use & trends over time

Issues & Challenges in Adoption

UK Open Data: Concerns

- 2011 Cabinet Office Consultation Report
  - Costs associated with developing systems capable of maintaining large datasets may be prohibitive
  - Public bodies may not have the requisite skills to effectively deliver an enhanced right to data
- Enhanced right to data
  - Needs “change in IT delivery at strategic level”
  - Need to re-examine both tender process & way in which IT contracts are set up
    - Incorporate open data standards into future contracts
    - Government should publish clear guidelines

Inherent Risks in Open Data (1/3)

Ineffective De-identification

- Original data need to be purged of reference or structure that may permit later identification of PII directly or indirectly

Insufficient Quality

- Data must be correct or updated as 3rd parties may use them to make critical decisions or influence behaviors

Inherent Risks in Open Data (2/3)

Incorrect or Malicious Use

- May be used to legitimise fraudulent schemes
  - E.g., mashing up Open Data with biased ratings of services or products

Revelation of Inconvenient Truths

- Data may reveal patterns that were not visible or known within govt
  - May cause embarrassment to govt

Inherent Risks in Open Data (3/3)

Into the Deep End

- As data becomes more readily available & accessible, citizens' interest in what was released will increase
- Operating as status quo may result in Open Data initiatives being regarded as irrelevant
  - May push citizen engagement backwards, if citizens perceive govt efforts as being insincere
- Continuity of Open Data in long term depends on ability to demonstrate value that it can derive

Sunlight Foundation

- Non-profit, nonpartisan organization that
  - Uses power of Internet to catalyze greater govt openness & transparency
  - Provides new tools & resources for media & citizens
- Committed to improving access to govt info by
  - Making it available online
  - Creating new tools & websites to enable individuals & communities to better access that info & put it to use
- Catalyze greater govt transparency by engaging individual citizens & communities demanding policies that will enable all of us to hold govt accountable

10 Principles for Opening Up Govt Data

- Pre-2007 era
  - US federal & state govts made some datasets available to public
    - Tend to be inconsistent & incomplete
    - Whetted appetites for more & better data
- Principles, if implemented, would empower public's use of govt-held data
  - Provides a lens to evaluate extent to which govt data is open & accessible to public

Sunlight Principles (1/10)

- Completeness
  - Dataset released should be as complete as possible
    - Must reflect entirety of what is recorded about specific subject
    - But must comply with federal law regarding release of PII
    - Must include metadata that defines & explains raw data, formulae & explanation on how derived data was made
  - Will enable users to understand scope of info available & examine each data item

Sunlight Principles (2/10)

- Primacy
  - Data released should be primary source data
  - Should include
    - Original info collected
    - Details on how data was collected
    - Original source documents recording data collection
  - Allows users to verify that info was collected properly & recorded accurately

Sunlight Principles (3/10)

- Timeliness
  - Datasets released should be available to public in a timely fashion
  - Should be released as quickly as it is gathered & collected
  - Priority should be given where utility is time sensitive
  - Real-time updates can maximise utility public obtains from such info

Sunlight Principles (4/10)

- Ease of Physical & Electronic Access
  - Accessible i.e., info can be obtained easily (whether physical or electronic)
  - Remove barriers like
    - Need to visit specific office in person
    - Requirements to comply with particular procedures (e.g., completing forms or Freedom of Information Act requests)
    - Data access only via submitted forms that require browser-oriented technologies (e., Flash, Javascript, cookies)
    - Preference
      - Bulk access via interface to download all info in dataset
      - Specific calls through API
  - Findability i.e., ability to easily locate & download content

Sunlight Principles (5/10)

- Machine Readability
  - Machines can handle certain kinds of input much better than others
    - Info should be stored in widely-used file formats that are machine processible
  - Avoid
    - Handwritten notes on paper
    - Scanned text via OCR (often with errors)
    - PDF documents are difficult to parse

Sunlight Principles (6/10)

- Non-Discrimination
  - Refers to who can access data & how they must do so
  - Any person can access data at any time without having to identify him/herself or provide justification for doing so
  - Remove barriers to use of data
    - Need for registration
    - Membership requirements
    - Use of "walled garden" where only some apps are allowed access to data

Sunlight Principles (7/10)

- Use of Commonly Owned Standards
  - Removing software license cost makes data available to a wider pool of potential users
  - Strong preference for open standards adoption for data format in which data is stored

Sunlight Principles (8/10)

- Licensing
  - Remove barriers to public use of data
    - Imposition of "Terms of Service" attribution requirements
    - Restrictions on dissemination
  - Maximal openness is desired. Should include
    - Labeling public info as work of govt
    - Available without restrictions on use as part of public domain

Sunlight Principles (9/10)

- Permanence
  - Refers to capability of finding info over time
  - Info released online should be sticky
  - Should be available online in archives in perpetuity
  - Current practice
    - Info is updated, changed or removed without any indication that an alteration is made
    - Streaming data is normally not archived
  - Desired
    - Appropriate version tracking
    - Archiving over time

Sunlight Principles (10/10)

- Usage Costs
  - Cost imposed on public for access should be minimal
  - Current bases for charging public access to their documents
    - Cost of creating info
    - Cost-recovery basis (e.g., cost to produce info divided by expected no. of purchasers)
    - Cost of retrieving info
    - Per page or per inquiry cost
    - Processing cost
    - Cost of duplication
  - Imposing fees for access skews pool of users willing or able to access info
  - May preclude transformative use of data that generates business growth & tax revenues

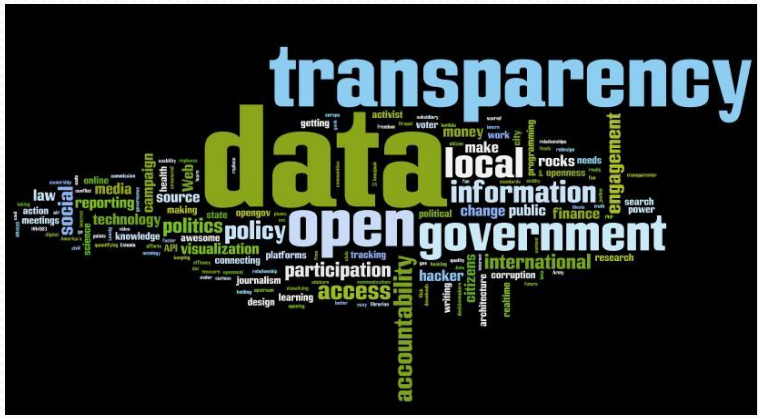

NOISE TO SIGNAL
Rob Cottingham
Suppose information doesn't want to be free?
Suppose what information really wants is to be meted out
in tiny, controlled doses at an outrageously high price?

Open Data vs. Open Government

- Open Govt & Open Data can each exist without the other
  - A govt can be an ‘open government’ (i.e., transparent) without embracing Open Data
  - A govt can provide Open Data on politically neutral topics while it remains deeply opaque & unaccountable

Open Government Projects …

- Policy initiatives that promote or reinforce principles of
  - Transparency
  - Participation
  - Collaboration



- Danger of blurring boundaries between <u>technologies of Open Data</u> & <u>politics of Open Government</u>

Open Data …

- Can help govt transparency

- BUT mundane & practical govt info (e.g., bus schedule, health inspection data that improves quality of life & enhance public service delivery) may not make any difference to transparency or accountability

Conflating

- Conflating release of any old info with releasing specific info journalists & critics ask can set a dangerous precedent

- When Open Data is confused with Open Government, govts may be able to take credit for increased transparency simply by delivering Open Data

What is “open govt data”?

- Deeply ambiguous, with different meanings
  - “open govt” & “data” together means politically important disclosures (whether delivered by computer or not)
  - “open” & “ govt” as separate adjectives modifying “data” means data that is both easily accessed & govt-related, but might or might not be politically important
  - A shortfall reference to mean govt data that is both politically sensitive & computer-provided

Mr. Lo Yoong Khong, Cluster Director, Government Infocomm Governance Division, and Justice & Law Cluster, IDA

Outline

- Will share how Singapore government has harnessed challenges & opportunities of enabling sharing of more than 6,000 datasets, as well as in driving demand for government data by people & private sectors

- Key driving force behind data.gov.sg is to share useful data for citizens to co-create & unearth value from, either by developing apps or being able to make informed decisions

- 1st stop portal for easy access to publicly-available government data for creation of applications, conducting research, or simply looking for interesting applications that have been developed using government data
- Catalyse development of innovative apps & services by private sector