

Evangelizing Metadata

... and other prerequisites for usable open data

Seattle Open Data Champions

March 10, 2017

What makes open data exciting?



DATA



KNOWLEDGE



ACTION

What makes open data usable?

FINDABILITY

- Persistent identifier
- Registered in searchable repository
- Rich metadata that includes identifier

ACCESSIBILITY

- Retrievable via a standardized, free, universal communications protocol that, when necessary, allows for authentication and authorization
- Metadata still accessible when data is not

INTEROPERABILITY

- Meta/data should be machine-readable
- Meta/data uses standardized (or at minimum, published) vocabularies

RE-USABILITY

- Meta/data has many accurate and relevant attributes
- Accessible and clear license
- Disclose provenance
- Meta/data meets domain-relevant standards

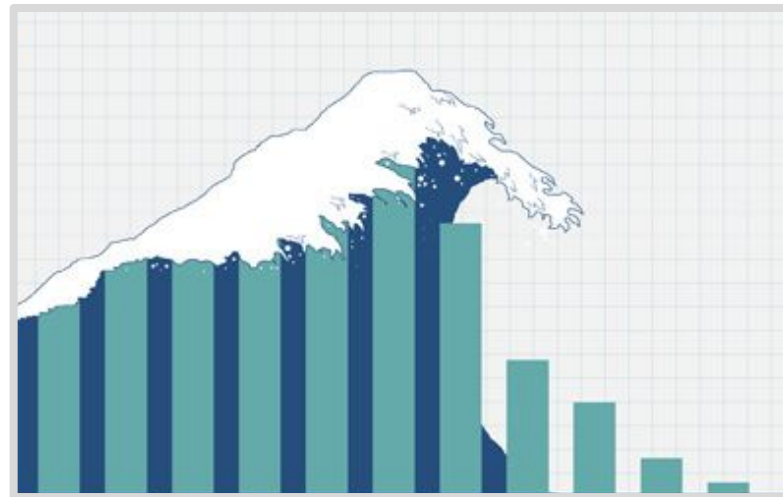
What makes open data usable?

8 PRINCIPLES for government data:

- As **complete** as possible, while protecting privacy and security
- As **granular** as possible, while protecting privacy
- As **timely** as possible
- Made **accessible**:
 - to **machines**;
 - to **people**, without any **discrimination**;
 - in **non-proprietary** formats;
 - without **legal barriers** (copyright, patent, trademark, license)

Data selection

- **Examine available data**
- **Examine peer portals:**
 - <https://data.cityofchicago.org/>
 - <https://data.cityofboston.gov/>
 - <http://data.opencolorado.org/>
 - <https://data.chattlibrary.org/>
- **Identify user communities:**
 - Public records requests
 - Web traffic on existing portal
 - Interviews, focus groups, surveys
- **Find applicable data standards:**
 - <https://project-open-data.cio.gov/open-standards/>
 - <http://archive.codeforamerica.org/our-work/data-formats/>
 - <http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html>
 - <https://developers.google.com/transit/gtfs/>



What is metadata?

TYPES

- Descriptive
- Structural (data dictionary)
- Administrative/provenance

LOCATIONS

- Catalog listing
- Web page
- README text file
- Codebook
- File header
- Graphical, e.g. FreeMind
- Downloadable JSON, XML, etc.

Field	Value
Author	City and County of Denver, Denver Police Department / Data Analysis Unit
Maintainer	City and County of Denver, Technology Services / DenverGIS Data
Version	1.0.873
Last Updated	March 10, 2017, 1:29 PM (UTC-08:00)
Created	May 30, 2013, 10:19 AM (UTC-07:00)

Field Name	Data Type	Data Format	Field Size	Description	Example
License ID	Integer	NNNNNN	6	Unique number ID for all drivers	12345
Surname	Text		20	Surname for Driver	Jones
First Name	Text		20	First Name for Driver	Arnold
Address	Text		50	First Name for Driver	11 Rocky st Como 2233
Phone No.	Text		10	License holders contact number	0400111222
D.O.B	Date / Time	DD/MM/YYYY	10	Drivers Date of Birth	08/05/1956

What is GOOD metadata?


BEST PRACTICES (<https://www.w3.org/TR/dwbp/#metadata>)

- “Provide metadata for both human users and computer applications.”
- “Provide metadata that describes the overall features of datasets and distributions.”
- “Provide metadata that describes the schema and internal structure of a distribution.”

STANDARDS

- Metadata standards directory: <https://rd-alliance.github.io/metadata-directory/standards/>
- Open metadata standards: <http://knowhow.opendatamonitor.eu/odresearch/metadata-standards/>
- U.S. government open data standard: <https://project-open-data.cio.gov/metadata-resources/>
- Crosswalks: <http://www.ands.org.au/online-services/rif-cs-schema/crosswalks-transform-your-metadata>

WRITING METADATA (<https://marinemetadata.org/guides/mdataintro/writegoodmdata>)

- Describe data (where, who, when, what, why) and its limitations
 - Choose words carefully: simple, unambiguous, controlled vocabulary
 - Get feedback from an outside audience
- 

Supporting data deposits

- **Interview**

- ["Conducting a data interview"](#)
- <http://datacurationprofiles.org/>

- **Provide standards and guidelines**

- Excel template
- Online form

- **Motivate**

- <http://theodi.org/the-value-of-open-data>
- <http://opencityapps.org/> (Chicago)
- <http://civicapps.org/> (Portland, OR)

more Open Data can help make better decisions

- 7,000 lives saved due to quicker response
- 5.5% less road fatalities
- 2,549 hours wasted finding parking
- 629 million hours saved is equivalent to € 27.9 bn
- 16% less energy used

CHICAGO COUNCILMATIC

Search Chicago Legislation

police, zoning, O2015-7825, etc.

Order by: Date Title Relevance

3308 legislation results

Report 2017-16 Active

Affordable Housing Plan 2014-2018 Quarterly Progress Report (2016 Q4)

3/29/2017 - Placed on file

Non-Routine City Matters Affordable Housing

Ordinance 2017-1752 Active

Vacation of public alley(s) in area bounded by N Milwaukee Ave, W Huron St and N Morgan St

3/23/2017 - Direct Introduction

Burnett, Jr., Walter

Non-Routine City Matters City Business Vacation of Public Street

Ordinance 2017-1801 Active

Amendment of Municipal Code Section 7-38-117 concerning establishment of mobile food vehicle stand measuring maximum of 40 ft. in length

Resources

Principles for usable data

- FAIR principles: <https://www.force11.org/group/fairgroup/fairprinciples>
- Open government data principles: https://public.resource.org/8_principles.html

Data standards

- Federal/international standards: <https://project-open-data.cio.gov/open-standards/>
- Open Referral, Open Trail System, LIVES, Housefacts: <http://archive.codeforamerica.org/our-work/data-formats/>
- SPOTCrime: <http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html>
- GTFS for transit: <https://developers.google.com/transit/gtfs/>

Metadata and metadata standards

- Open Data Metadata Guide (basic introduction): <https://centerforgov.gitbooks.io/open-data-metadata-guide/content/>
- Best practices for metadata: <https://www.w3.org/TR/dwbp/#metadata>
- How you can publish your metadata: <https://marinemetadata.org/guides/mdataintro/gettingstarted>
- Guidelines for writing metadata: <https://marinemetadata.org/guides/mdataintro/writegoodmdata>
- Metadata standards directory: <https://rd-alliance.github.io/metadata-directory/standards/>
- Open data metadata standards directory: <http://knowhow.opendatamonitor.eu/odresearch/metadata-standards/>
- Project Open Data: <https://project-open-data.cio.gov/metadata-resources/>

Supporting data deposits

- Data interview guide: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1092&context=lib_research
- PwC report on value of open data: <http://theodi.org/the-value-of-open-data>
- Chicago civic apps: <http://opencityapps.org/>
- Portland, OR civic apps: <http://civicapps.org/>