

# Open Data and Open Libraries:

## Supporting the Open Data Project in Library Services

Jason A. Clark  
Associate Professor  
Head of Library Informatics & Computing  
Montana State University (MSU) Library

Florida Library Webinars  
July 22, 2015

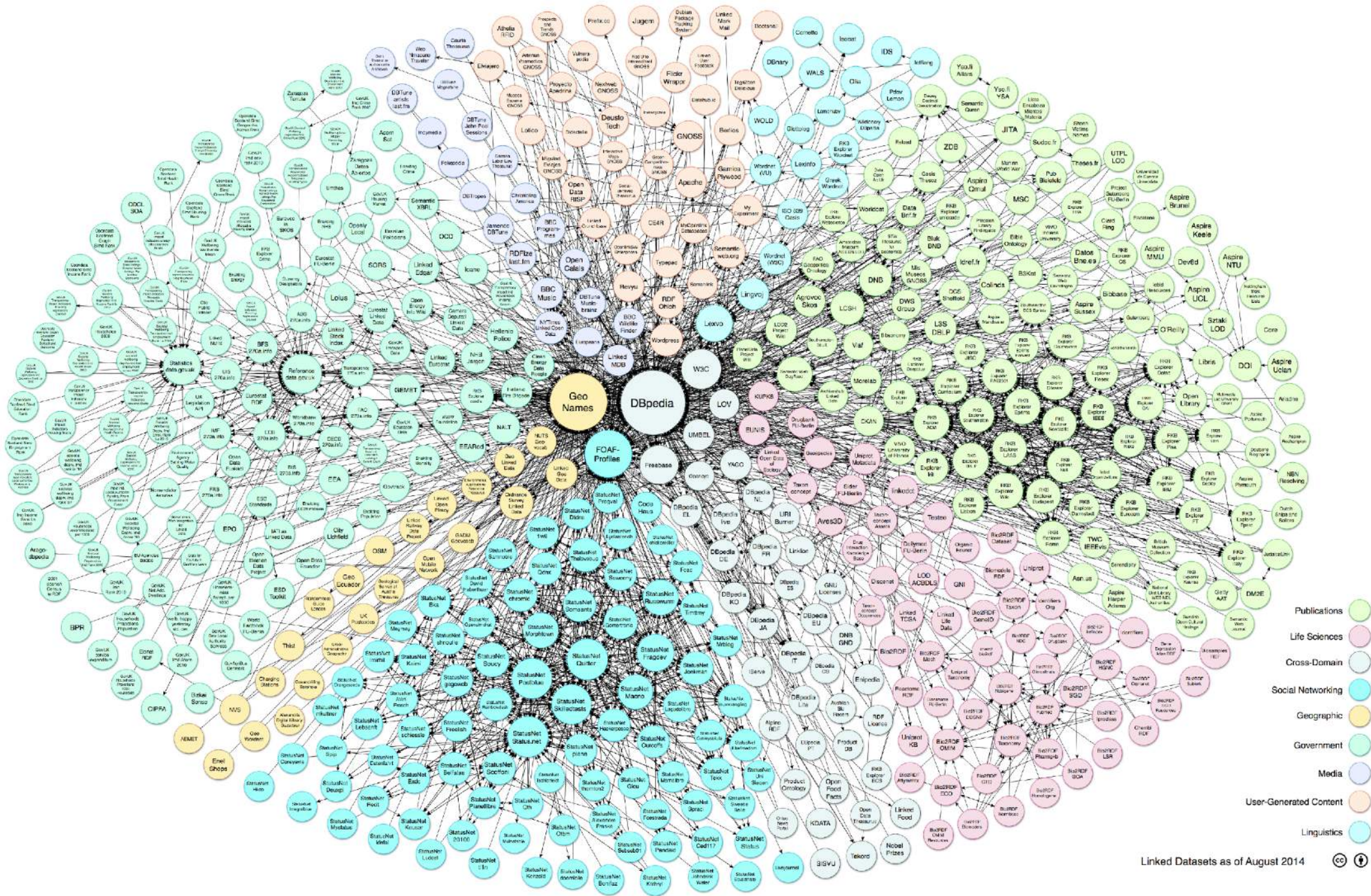
pinboard.in tag

pinboard.in/u:jasonclark/t:open/

twitter as channel (#hashtag)

@jaclark #opendata



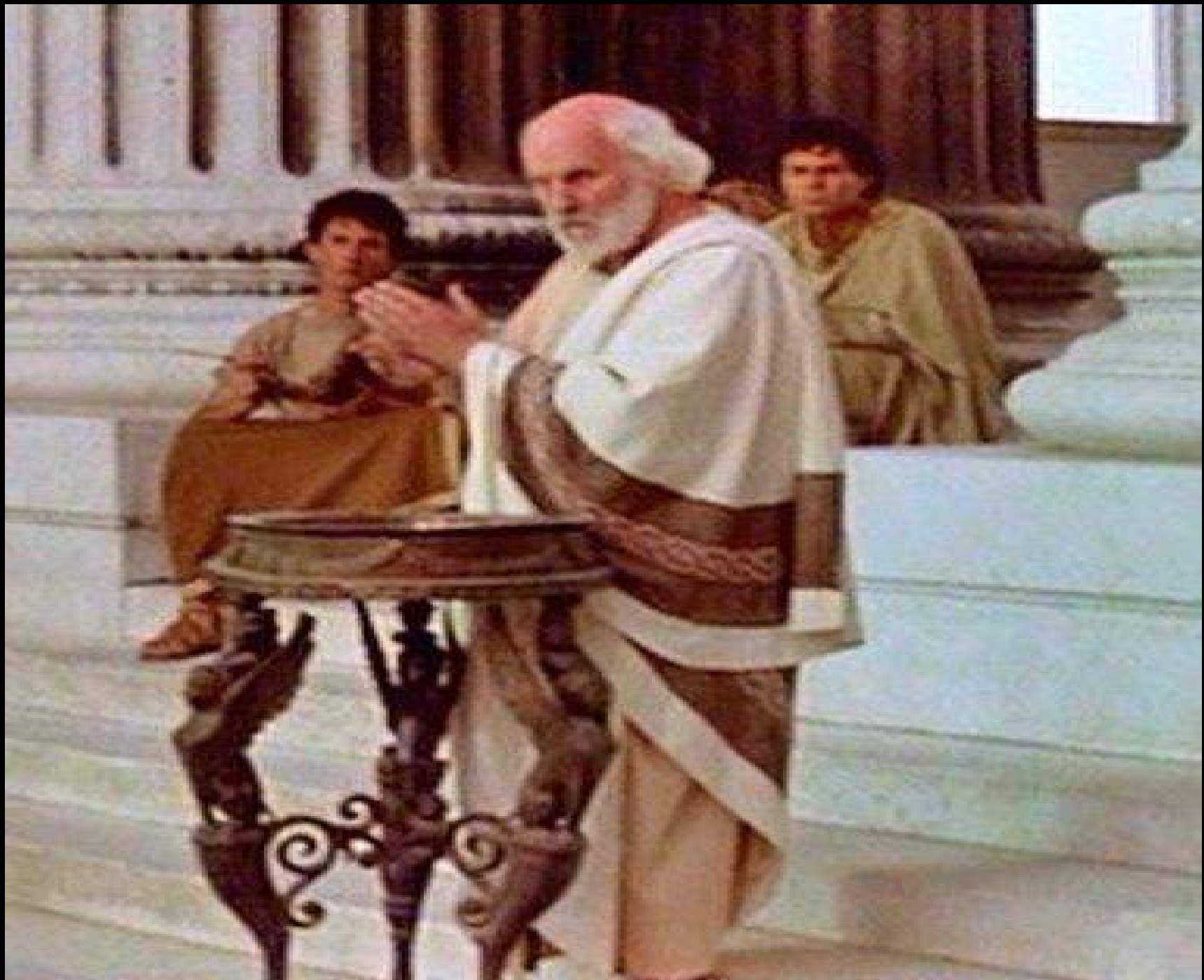


Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>



# Overview

- Why Open Data?
- Open Data in Practice
- Open Data + Libraries
- Open Data Processes + Tools
- Questions



Defining Terms and Truths

# What do you mean by “open”?

- An open work must satisfy requirements around open licenses, open access, and open formats.

It can be summed up in the statement that:

*“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).”*

Put most succinctly:

*“Open data and content can be freely used, modified, and shared by anyone for any purpose”*

[Read the full Open Definition »](#)

<http://opendefinition.org/od/>



“A piece of data or content is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share-alike.”

<http://opendefinition.org/>

# Why Open Data?

## Part of a Library Continuum

Sharing



Search Twitter





**Open City**  
@opencityapps  
Group of volunteers that create web apps with open data to improve transparency, efficiency, and decision-making for governments and our democracy.  
Chicago  
[opencityapps.org](http://opencityapps.org)  
Joined January 2012

Tweet to Open City

TWEETS 1,209 FOLLOWING 293 FOLLOWERS 2,446 FAVORITES 32

Tweets Tweets & replies Photos & videos

Open City retweeted

**DataMade** @DataMadeCo · Jul 20  
Thank you to @CJPJustProj for fighting to make public safety data public. We are using data they liberated in [chicagosmilliondollarblocks.com](http://chicagosmilliondollarblocks.com)  
4 7 View summary

Open City retweeted

**Cathy Deng** @cthydng · Jul 20  
there are better ways to spend public dollars. [chicagosmilliondollarblocks.com](http://chicagosmilliondollarblocks.com)

<https://twitter.com/opencityapps>

# Attribution





## Cite your data

Why is it so important to cite data? Books and journal articles have long benefited from an infrastructure that makes them easy to cite, a key element in the process of research and academic discourse. We believe that you should cite data in just the same way that you can cite other sources of information, such as articles and books. Data citation can help by:

- enabling easy reuse and verification of data
- allowing the impact of data to be tracked
- creating a scholarly structure that recognises and rewards data producers

### Examples of data citation

We recognise that the challenges associated with data publication vary across disciplines, and we encourage research communities to develop citation systems that work well for them. Our recommended format for data citation is as follows:



## Events

Keep an eye out for the upcoming future events

## News

### Announcing the THOR Project

We are pleased to announce DataCite's participation in THOR (Technical and Human...

### DataCite Event -- Persistent Identifiers: Enabling Services for Data Intensive Research

Save the date and join...

<https://www.datacite.org/services/cite-your-data.html>

Access

# The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.

## GET STARTED

SEARCH OVER 157,813 DATASETS



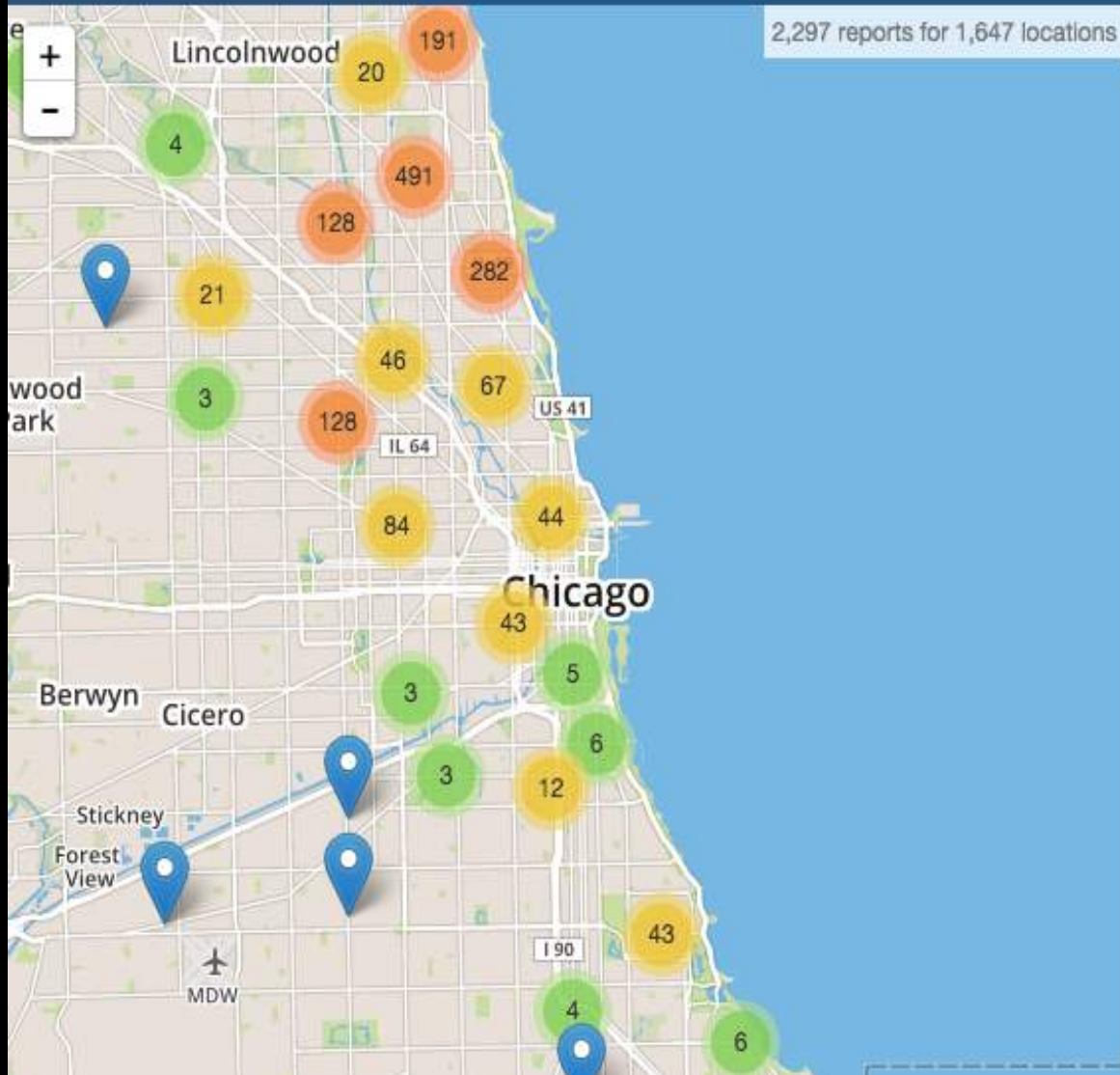
## BROWSE TOPICS



<http://www.data.gov/>

Reuse and Repurpose

# MY BUILDING DOESN'T RECYCLE!

[Report Your Building](#)[Get Involved](#)[About](#)[Contact](#)

## Hey Chicagoans!

### No recycling in your building?

#### Tell us about it!

[Chicago law](#) requires buildings with 5 or more residential units to provide adequate recycling services to occupants.

However, the City of Chicago does not enforce this law—and this leads many landlords NOT to provide recycling services to tenants.

Add your report to our registry to show that your building does not recycle.

Your report shows that many buildings are NOT offering recycling and that Chicagoans support stronger recycling laws.

**Enter your Chicago street address:**

<http://mybuildingdoesntrecycle.com/>



Archive



## Browse by Topic

[Education](#) / [Arts](#) / [Humanities](#) / [Social Sciences](#) / [Natural Sciences](#) / [Applied Sciences & Technology](#) / [Mathematics & Statistics](#)



New on OER  
Commons



Social /  
Emotional  
Learning



Common Core



Primary  
Sources



Teachers as  
Makers



Flexible  
Learning



Game-based  
Learning

<https://www.oercommons.org/>

# Open Data in Practice

So what?

# EmbedThis (Oembed)

## Pass a URL & Get Embed Code

Get Code

Key: [Home](#) [What?](#) [Code](#) [Credit](#) [@jaclark](#)

<http://www.lib.montana.edu/~jason/files/oembed-this/>

# EmbedThis (Oembed)

## Pass a URL & Get Embed Code

Hmmm... Thanks for trying out the URL at <https://www.youtube.com/watch?v=H7pkpSVbyc8>

Here's what I found:

MSU Library - Everywhere You Go



embed source code:

```
<iframe width="320" height="180" src="https://www.youtube.com/embed/H7pkpSVbyc8?feature=oembed" frameborder="0" allowfullscreen></iframe>
```

---

Key: Home What? Code Credit @jaclark

<https://github.com/jasonclark/oembed-this/>



# Application Programming Interfaces (API)

```

<?xml version="1.0"?>
<items>
  <item>
    <id>1</id>
    <name>Angel Food Cake</name>
    <creator>Laura Cameron</creator>
    <url>http://arc.lib.montana.edu/book/home-cooking-history-409/article-laura-cameron.html</url>
    <description>Angel food cake recipe written by Laura Cameron for Montana State University's upper division history c
    <about>baking cake</about>
    <image>http://upload.wikimedia.org/wikipedia/commons/thumb/6/64/AngelFoodCake.jpg/200px-AngelFoodCake.jpg</image>
    <cookingMethod>Baking</cookingMethod>
    <ingredients>18 egg whites, 2 t. cream of tartar, 1 pinch salt, 1 1/2 cups white sugar, 1 cup cake flour, 1/2 cup co
    <recipeInstructions>&lt;li&gt;Preheat oven to 350 degrees&lt;/li&gt;
    &lt;li&gt;Sift cake flour and confectioners sugar together 5 times. Set aside.&lt;/li&gt;
    &lt;li&gt;In a large bowl whip egg whites with salt until foamy. Add cream of tartar and continue to beat until soft
    &lt;li&gt;Quickly fold in flour mixutre.&lt;/li&gt;
    &lt;li&gt;Pour into a 10 inch tube pan.&lt;/li&gt;
    &lt;li&gt;Bake for 45 minutes.&lt;/li&gt;</recipeInstructions>
    <cookTime>PT45M</cookTime>
    <prepTime>PT15M</prepTime>
    <totalTime>PT60M</totalTime>
    <recipeCategory>dessert</recipeCategory>
    <recipeCuisine>American</recipeCuisine>
    <recipeYield>8 to 10 servings</recipeYield>
    <sameAs>http://dbpedia.org/page/Angel_food_cake</sameAs>
    <keywords>cream of tartar, confectioners sugar, vanilla extract, egg whites, cup cake flour, cake flour</keywords>
    <publisher>Montana State University Library</publisher>
    <dateCreated>2012-12-01T18:34Z</dateCreated>
    <dateModified>2012-12-01T18:34Z</dateModified>
    <datePublished>2012-12-01T18:34Z</datePublished>
    <inLanguage>en</inLanguage>
  </item>
</items>

```

...api?id=1&format=xml

...api?type=search&q=poetry&format=json

# Open Data + Libraries

Where does it fit?

# Teaching Data Science or Data Literacy Skills

# Archiving and Packaging Data for Discovery and Reuse



# Sponsoring Open Data Creation Moments

INTERNATIONAL  
**OPEN DATA**  
HACKATHON

[What is this?](#)

[Who is the hackfest for?](#)

[Hackathon details](#)

[Let's talk](#)

[Media](#)

[Map of Events](#)

[Wiki](#)

[English version](#)

**You're Invited. Again.**  
**Open Data Day will take place on**  
**Saturday, February 21, 2015**

Open Data + You + Your Community + Your Timezone = Open Data  
Around the World

Stay tuned for more Updates via the [Community mailing list](#), [website](#) and [Community wiki](#) updates.

<http://opendataday.org/>

# Open Data Processes + Tools

## Getting Started

# Structured Data

	A	B	C	D	E	F
1	Title	Author	Publisher	OCLC #	ISBN	Notes
2	This house of sky : landscapes of a Western mind	Ivan Doig	New York: Harcourt Brace Jovanovich, 1992.	25629631	0151900558	
3	To skin a cat : stories	Thomas McGuane	New York: Dutton/S. Lawrence, 1987.	15549047	0394755219	
4	Young men & fire	Norman Maclean	Chicago : University of Chicago Press, 1992.	283736539	0226500624	
5	A river runs through it, and other stories	Norman Maclean	Chicago: University of Chicago Press, 1976.	442266245	0671776975	Also a film version from Robert Redford <a href="http://en.wikipedia.org">http://en.wikipedia.org</a>
6	The horse whisperer	Nicholas Evans	New York: Delacorte Press, 1995.	35722048	0440222656	
7	Hiking Yellowstone National Park	Bill Schneider	Helena, Mont: Falcon Press, 2003.	52500539	0762725397	2nd edition (Hiking Guide Series)

# 5 Star Open Data



make your stuff available on the Web (whatever format) under an open license<sup>1</sup>



make it available as structured data (e.g., Excel instead of image scan of a table)<sup>2</sup>



use non-proprietary formats (e.g., CSV instead of Excel)<sup>3</sup>



use URIs to denote things, so that people can point at your stuff<sup>4</sup>



link your data to other data to provide context<sup>5</sup>

<http://5stardata.info/>



```

<?xml version="1.0"?>
<items>
  <item>
    <id>1</id>
    <name>Angel Food Cake</name>
    <creator>Laura Cameron</creator>
    <url>http://arc.lib.montana.edu/book/home-cooking-history-409/article-laura-cameron.html</url>
    <description>Angel food cake recipe written by Laura Cameron for Montana State University's upper division history c
    <about>baking cake</about>
    <image>http://upload.wikimedia.org/wikipedia/commons/thumb/6/64/AngelFoodCake.jpg/200px-AngelFoodCake.jpg</image>
    <cookingMethod>Baking</cookingMethod>
    <ingredients>18 egg whites, 2 t. cream of tartar, 1 pinch salt, 1 1/2 cups white sugar, 1 cup cake flour, 1/2 cup co
    <recipeInstructions>&lt;li&gt;Preheat oven to 350 degrees&lt;/li&gt;
    &lt;li&gt;Sift cake flour and confectioners sugar together 5 times. Set aside.&lt;/li&gt;
    &lt;li&gt;In a large bowl whip egg whites with salt until foamy. Add cream of tartar and continue to beat until soft
    &lt;li&gt;Quickly fold in flour mixutre.&lt;/li&gt;
    &lt;li&gt;Pour into a 10 inch tube pan.&lt;/li&gt;
    &lt;li&gt;Bake for 45 minutes.&lt;/li&gt;</recipeInstructions>
    <cookTime>PT45M</cookTime>
    <prepTime>PT15M</prepTime>
    <totalTime>PT60M</totalTime>
    <recipeCategory>dessert</recipeCategory>
    <recipeCuisine>American</recipeCuisine>
    <recipeYield>8 to 10 servings</recipeYield>
    <sameAs>http://dbpedia.org/page/Angel_food_cake</sameAs>
    <keywords>cream of tartar, confectioners sugar, vanilla extract, egg whites, cup cake flour, cake flour</keywords>
    <publisher>Montana State University Library</publisher>
    <dateCreated>2012-12-01T18:34Z</dateCreated>
    <dateModified>2012-12-01T18:34Z</dateModified>
    <datePublished>2012-12-01T18:34Z</datePublished>
    <inLanguage>en</inLanguage>
  </item>
</items>

```

...api?id=1&format=xml

...api?type=search&q=poetry&format=json

# Minimum Viable Metadata

Search the menus (Option+/) 100% Viewing 1 of 5

Label	Definition	Input Rules	Local Repository Fields	Crosswalk
Descriptive Metadata				
Title	"The name given to the resource. Typically, a Title will be a name by which the resource is formally known." (DCMI)	Dataset: Data from [study] of [entity] regarding [subject]  Newsletter: [title of newsletter], [issue date]  Book: title from the catalog record  Other formats: Devise a brief descriptive title that uniquely identifies the resource (e.g. names of persons, corporate bodies, objects, activities, and events).	DC Map: dc.title  Data Type: Text  Search: Yes  Hide: No  Required: Yes  Vocab: No  Required in ScholarWorks Field  Template:Yes	DC: <a href="#">Title</a>  Schema.org: <a href="#">name</a>  MARC: <a href="#">245 \$a</a> <a href="#">245 \$b</a>

Data Dictionary for Datasets - Montana State University (MSU) Library

Licensing



# Licenses

- [Public Domain Dedication and License \(PDDL\)](#) — “Public Domain for data/databases”
- [Attribution License \(ODC-By\)](#) — “Attribution for data/databases”
- [Open Database License \(ODC-ODbL\)](#) — “Attribution Share-Alike for data/databases”

## Applying a License

For instructions on how to apply the licenses to your material please see each license’s home page.

## Questions or suggestions? There is:

- A [License FAQ](#) (plus the [General FAQ](#)).
- A public mailing list, wiki page and contact email. See the [contact page](#) for details.

### MORE INFORMATION

- [Introduction to Open Data](#)
- [Open Definition for Data](#)
- [Quick guide to making data open](#)
- [Open Data Handbook](#)

### Meta

An Open Knowledge Foundation project.

Find out more about open data or read the definition of open data on [OpenDefinition.org](#)

Site content licensed under a [Creative Commons Unported v3.0 Attribution license](#).

<http://opendatacommons.org/licenses/>

Tools

# OPEN DATA HANDBOOK

*Guides, case studies and resources for government & civil society  
on the “what, why & how” of open data.*



## Open Data Guide

This guide discusses the legal, social and technical aspects of open data. It can be used by anyone but is especially designed for those seeking to open up data. It discusses why to go open, what open is, and the how to 'open' data.

[Start Reading](#)

## Value Stories

Use cases, stories and case studies highlighting the social and economic value, the impact and the varied applications of open data from cities and countries across the globe.

[Value Stories](#)

## Resource Library

A curated collection of open data resources, including articles, longer publications, how to guides, presentations and videos, produced by the global open data community.

[Open Data Resources](#)



OPEN KNOWLEDGE

SUPPORTED BY:



SHUTTLEWORTH  
FUNDED

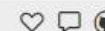


THE  
WORLD  
BANK

[Contribute](#) [Terms of use](#) [Privacy policy](#)

[OPEN CONTENT](#)

Related Projects: [OpenGovernmentData.org](#) | [The DataHub](#) | [DataCatalogs.org](#) | [OpenSpending.org](#) | [SchoolOfData.org](#) | [OpenDefinition.org](#) | [Introducing Open Data](#)



<http://opendatahandbook.org/>



## Tools to Explore, Publish and Share Public Datasets

## Information



<http://opendata-tools.org/en/>



# Refine

OPEN



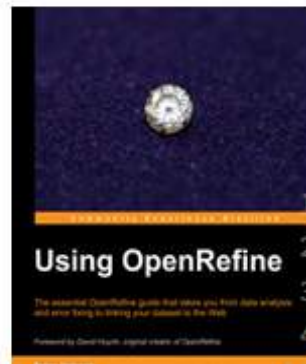
*A free, open source, powerful tool  
for working with messy data*

## Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; extending it with web services; and linking it to databases like [Freebase](#).

Please note that since October 2nd, 2012, Google is not actively supporting this project, which has now been rebranded to OpenRefine. Project development, documentation and promotion is now fully supported by volunteers. Find out more about the [history of OpenRefine](#) and how you can [help the community](#).

## Using OpenRefine - The Book



**Using OpenRefine**, by Ruben Verborgh and Max De Wilde, offers a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds
3. Apply basic and advanced cell transformations
4. Deal with cells that contain multiple values

### [Home](#)

### [Download](#)

### [Documentation](#)

### [Community](#)

### [Post archive](#)

[Mapping OpenRefine  
Ecosystem](#)

[2014 survey results](#)

[A Governance Model for  
OpenRefine](#)

[Using OpenRefine: a](#)

<http://openrefine.org/>

## CKAN, the world's leading open-source data portal platform

CKAN is a powerful data management system that makes data **accessible** – by providing tools to streamline **publishing**, sharing, **finding** and **using** data. CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available.



✓ Take Tour

🛒 Get CKAN

▶ Live Demo

### Feature Overview



Complete catalog system with easy to use web interface and a powerful API



Fine-grained access control



Integrated data storage and full data

### Support and Hosted Solutions

CKAN is open source and can be

<http://ckan.org/>



## store, share, discover research

manage your research in the cloud and control who you share it with  
or make it publicly available and citable

[About figshare](#)[Browse research](#)[See how we support data management for Institutions >](#)[See how we partner with Publishers >](#)

## sign up for free

☐ [Accept Terms & Conditions](#)[Sign up](#)

The background figure: [Iceland scallops photographed with a AUV camera...](#) by Warsha Singh in [Biological Sciences](#)

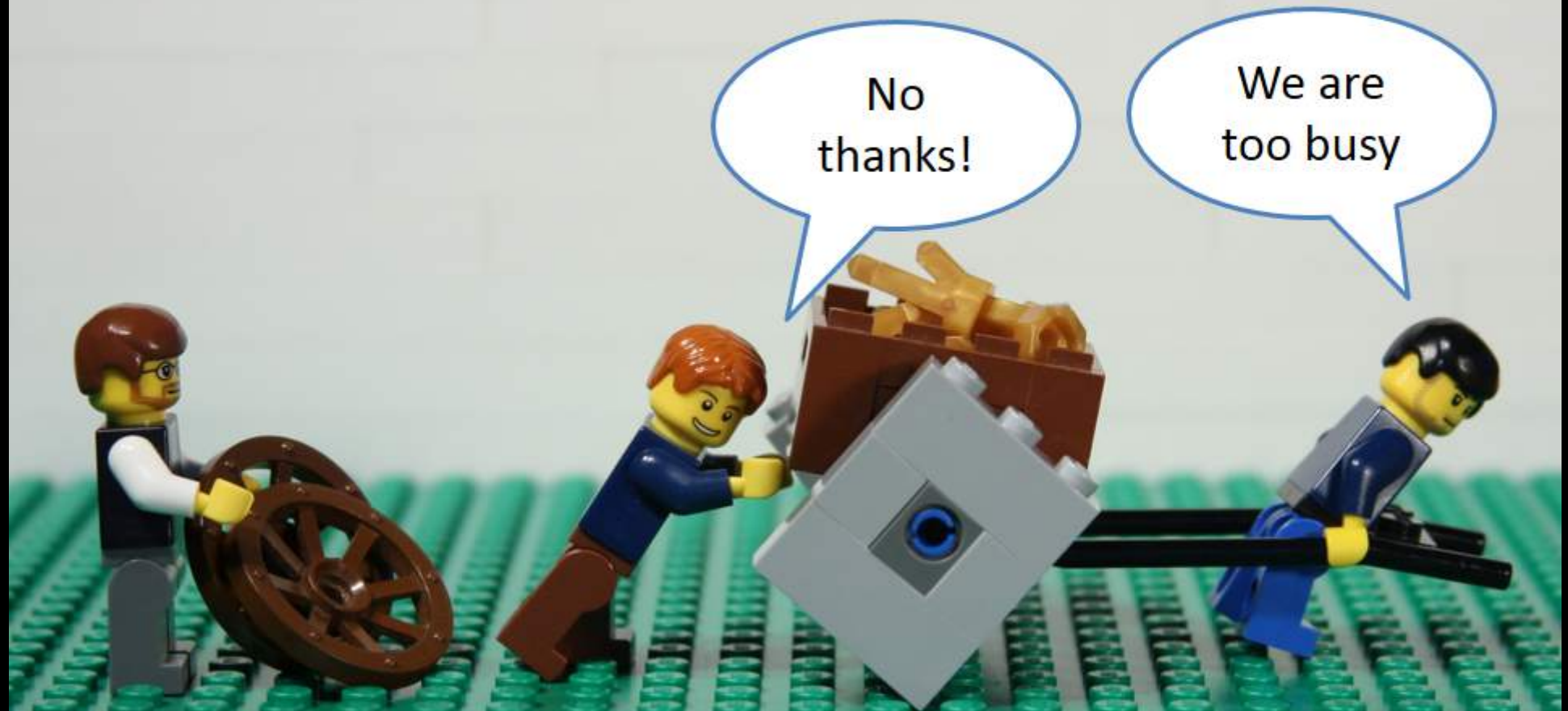
<http://figshare.com/>

# Questions?

[twitter.com/jaclark](https://twitter.com/jaclark)

[www.lib.montana.edu/~jason/talks.php](http://www.lib.montana.edu/~jason/talks.php)

# Are you too busy to improve?



Håkan Forss @hakanforss <http://hakanforss.wordpress.com>

This illustration is inspired by and in part derived from the work by Scott Simmerman, "The Square Wheels Guy" <http://www.performancemanagementcompany.com/>

