

Working With Open Data

Vancouver Public Library



Learning Opportunities @ VPL!

- Digital Essentials: vpl.ca/program/digital-essentials

A certificate of completion template with a decorative header featuring three horizontal lines (yellow, red, blue) and three overlapping circles (blue, red, yellow). The text reads: "CERTIFICATE OF COMPLETION", "This certificate is presented to", "Patron Name", "for their completion of", "Series Name", and "Month XX, YYYY". The VPL logo and "Vancouver Public Library" are at the bottom left.

CERTIFICATE OF COMPLETION

This certificate is presented to

Patron Name

for their completion of

Series Name

VPL
Vancouver Public Library

Month XX, YYYY



Vancouver Public Library

Learning Opportunities @ VPL!



Lynda.com: vpl.ca/lynda



Learning Express Library:

vpl.ca/digital-library/learning-express-library



VPL Research Guides: guides.vpl.ca





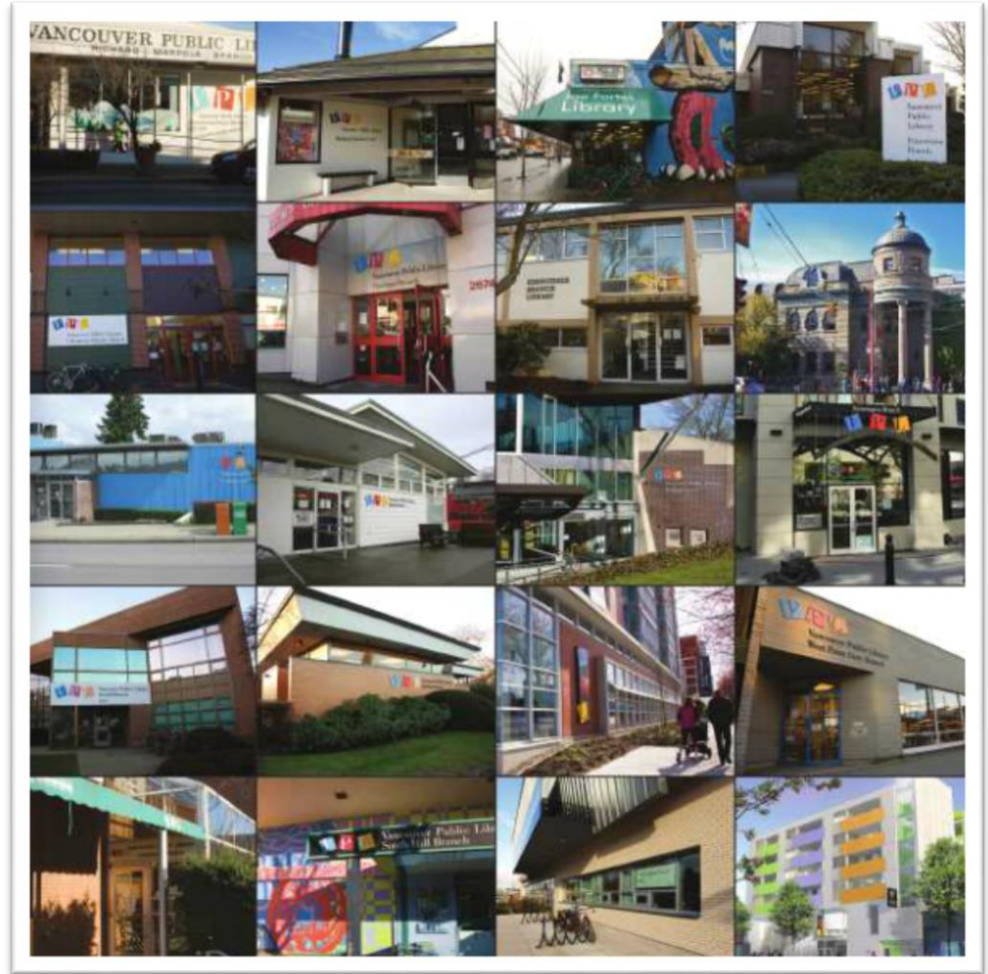
Vancouver Public Library

Learning Opportunities @ VPL!

We are here
to help!

604-331-3603

info@vpl.ca



Today We Will Learn How To . . .

- Identify and locate open data sets on the internet.
- Work with a single data set to follow the steps to use open data to create something new.
- Distinguish between information license types and determine their impact on information access.



What is Open Data? Review



Steps to Using Open Data

1. Acquiring (Finding)
2. Extracting (Downloading)
3. Cleaning & Processing
4. Publishing
5. Maintaining



ACQUIRING: Where to find Data Sets

- **Government Data** (Federal, Provincial, Municipal)
- **Research Data** (Research Data Canada, Academic Institutions)
- **Social Data** (Facebook, Instagram)
- **Weather Data** (OpenWeatherMap, Environment & Climate Change Canada)
- **News Data** (The Guardian, The New York Times)
- **Sports Data** (Sports Reference)
- **Data Hubs** (Qlik, Infochimps, CKAN, Google Public Data)





ACQUIRING: Searching for Data on the Web

Vancouver Public Library



Advanced Search

Find pages with...

all these words:

Vancouver Open Data

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from:

to

To do this in the search box.

Type the important words: tri-colour rat terrier

Put exact words in quotes: "rat terrier"

Type OR between all the words you want: miniature OR standard

Put a minus sign just before words that you don't want:
-rodent, -"Jack Russell"

Put two full stops between the numbers and add a unit of measurement:
10..35 kg, £300..£500, 2010..2011

Then narrow your results by...

language:

any language

Find pages in the language that you select.

region:

any region

Find pages published in a particular region.

last update:

anytime

Find pages updated within the time that you specify.

site or domain:

.ca

Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov

terms appearing:

anywhere in the page

Search for terms in the whole page, page title or web address, or links to the page you're looking for.

SafeSearch:

Show most relevant results

Tell SafeSearch whether to filter sexually explicit content.

file type:

Microsoft Excel (.xls)

Find pages in the format that you prefer.

usage rights:

not filtered by licence

Find pages that you are free to use yourself.

Advanced Search



Vancouver Public Library

ACQUIRING: Searching for Hidden Data

Finding hidden data



00:00



ACQUIRING: Reminder

Don't forget to . . .

- Check the license
- Evaluate the data

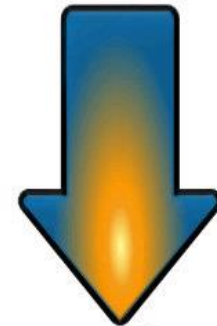


Any Questions about Acquiring (Finding) Data?



EXTRACTING: Tips & Tricks

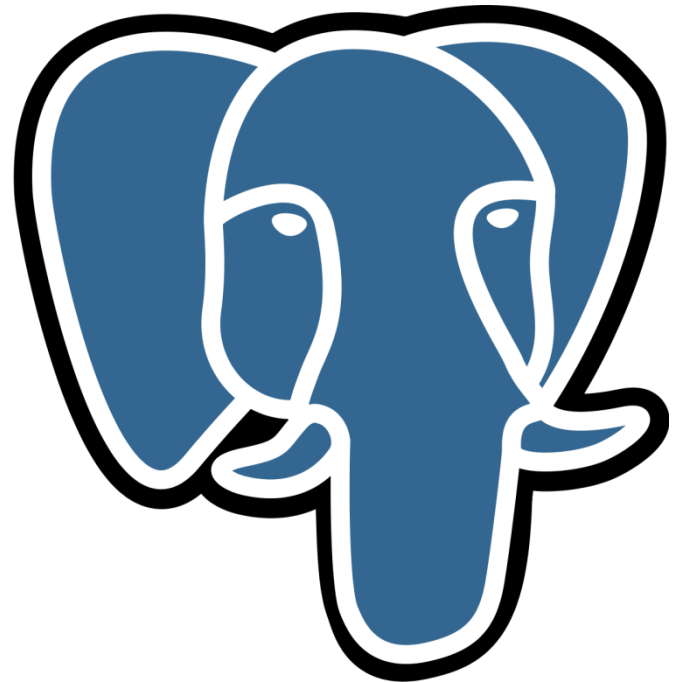
- Look For Download Options
- Download the Entire Database
- Automate Downloads of Multiple Data Files



down THEM all!



EXTRACTING: Data Management



Any Questions about Extracting (Downloading) Data?

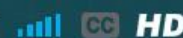


CLEANING & PROCESSING

Turning data into diamonds



00:01



CLEANING & PROCESSING: Tools



3) Cleaning & Processing: **Common Errors**

- Wrong date formats
- Multiple representations
- Duplicate records
- Redundant data
- Mixed numerical scales
- Mixed ranges
- Spelling errors



Any Questions about Cleaning or Processing Data?





Vancouver Public Library

Let's Practice!



1. Find/Acquire a simple dataset
2. Download that dataset and convert it into a CSV file
3. Clean and Process the dataset (locate fix errors)
4. Use the dataset to create an amazing visualization (chart or graph)





Vancouver Public Library



Saving an Excel Document as a CSV File

- CSV ‘Comma-separated values’, is a standard format for spreadsheet data.
- Data is represented in a plain text file, with each data row on a new line and commas separating the values on each row.
- As a very simple open format it is easy to consume and is widely used for publishing open data.



CSV Example

A table in Excel:

Year	Make	Model	Description	Price
1997	Ford	E350	ac, abs, moon	3000.00
1999	Chevy	Venture "Extended Edition"		4900.00
1999	Chevy	Venture "Extended Edition, Very Large"		5000.00
1996	Jeep	Grand Cherokee	MUST SELL! air, moon roof, loaded	4799.00

The above table of data represented in CSV format :

```
Year,Make,Model,Description,Price
1997,Ford,E350,"ac, abs, moon",3000.00
1999,Chevy,"Venture ""Extended Edition""",,,4900.00
1999,Chevy,"Venture ""Extended Edition, Very Large""",,,5000.00
1996,Jeep,Grand Cherokee,"MUST SELL! air, moon roof, loaded",4799.00
```



Using Excel to Create a CSV File

The screenshot shows an Excel spreadsheet with the following data:

listing_type	profession	company	website	email	phone_no	address1	city	state_cod	zip_code
Company	Dentist	Dental Ins	http://brl-usersema	(720) 111- 120 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 222- 121 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 333- 122 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 444- 123 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 555- 124 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 666- 125 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 777- 126 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 888- 127 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 999- 128 W 111	Denver	CO	80204		
Company	Dentist	Dental Ins	http://brl-usersema	(720) 900- 129 W 111	Denver	CO	80204		



RECAP: Steps to Using Open Data

1. Acquiring (Finding)
2. Extracting (Downloading)
3. Cleaning & Processing
- 4. Publishing**
5. Maintaining



PUBLISHING: Platforms

- Publishing as files on a website



- Uploading to a portal



- Publishing through API



```
{"logo": "API Evangelist"}
```

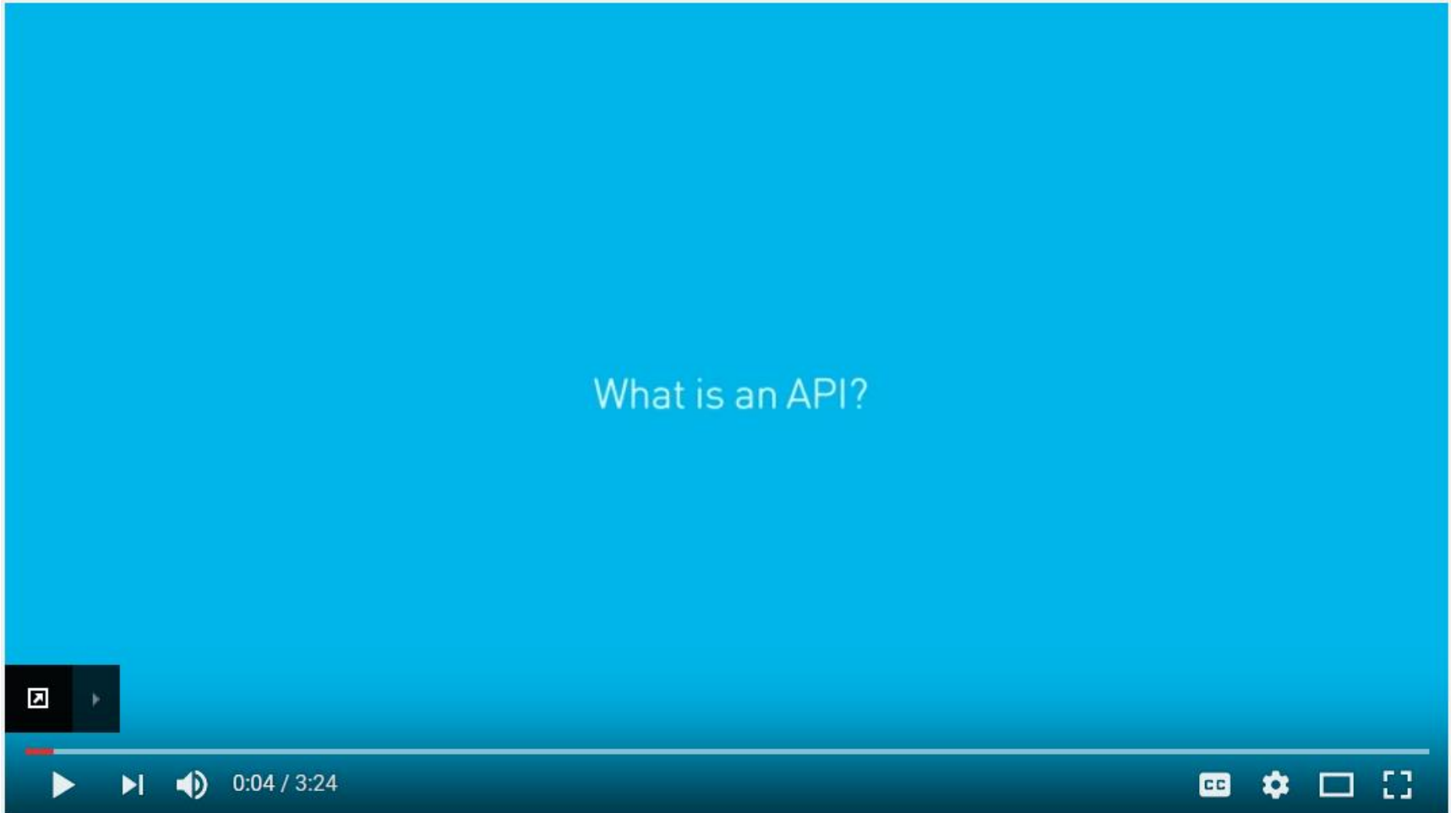
- Creating an App





Vancouver Public Library

PUBLISHING: API





Vancouver Public Library

PUBLISHING: App Development

Lynda.com FROM LINKEDIN LIBRARY Hi Library

Developer > Development Tools

Welcome

From: Learning Android App Development

Layout Add to Playlist Share

Contents Notebook

Search This Course

- Introduction
 - Welcome 55s
 - What you should know 59s
 - Exercise files 1m 34s
- 1. Android Basics
 - Android app basics 1m 41s
 - App Development overview 2m 57s
 - Tools for Android app development 1m 55s
- 2. Setup
 - Android development requirements 1m 46s
 - Download and install Android



PUBLISHING: Adding Metadata

- Metadata is:
 - Data that describes and gives information about other data (data about data)
- Metadata helps people:
 - Discover data
 - Understand the nature and structure of the data
 - Find the origin of the data and under what terms it may be used





PUBLISHING: Standard Dataset Metadata

Type to search

Open Data - Metadata Guide

Introduction

Categories

Dataset Metadata

Column Metadata

Additional Resources

[Appendix A: Sample Dataset Metadata](#)

Appendix B: Sample Column Metadata

Published with GitBook



Appendix A: Sample Dataset Metadata

Standard Dataset Fields

The U.S. federal government has created the Project Open Data metadata schema standard to implement the federal open data policy. The [Project Open Data schema](#) is based on the international DCAT metadata schema used by open data programs around the world and [has been mapped to many standards](#). The Project Open Data schema must be presented as a JSON file to be ingested by Data.gov. This schema is natively available with many open data portal providers including: *Azavea*, *Esri Open Data*, *NuCivic's DKAN*, *OpenGov*, and *Socrata*, and is easily added to *CKAN* sites with an extension or can be generated on an ad hoc basis with these [tools](#).

Field	Label	Definition	Required
title	Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always
description	Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Always





Vancouver Public Library

PUBLISHING: Metadata Training

W3C[®]



PUBLISHING: Apply an Open License

Open Data Commons:

- [Public Domain Dedication and License](#)
Puts all your material in the 'Public Domain' (waiving all rights).
- [Open Data Commons/Open Database License](#)
Allows the public to copy, distribute, and adapt your data as long as they attribute , share-alike, and keep it open.
- [Open Data Commons/Attribution License](#)
Allows the public to copy, distribute, and adapt your data as long as they attribute.

Creative Commons:

- Free, easy-to-use copyright licenses to give the public permission to share and use your creative work—on conditions of your choice.



Any Questions about Publishing Data?



MAINTAINING

- Maintain data and metadata regularly
- Check links
- Check user feedback



Brainstorming

- Work in pairs or small groups
 - What sort of open data would you like to have access to?
 - How would you use this open data: what would you create?
 - Do you see any problems or issues that could potentially be solved using open data?



CONTINUED LEARNING

SCHOOL OF DATA

Learn Our Work Fellows & Experts About Join Blog

Online Courses

The School of Data aims to make your learning experience as tailored as possible through independent learning modules. Learning modules are self-paced, stand-alone and can be taken in any order. To make your learning experience easier, we curated modules into a series of courses. Courses range from data basics as well as specific skills. When you identified a topic you are interested in click on "Show Modules" to see all modules you can take.

- Essentials
- Data Fundamentals
- Data Cleaning
- Exploring Data
- Extracting Data

StackExchange sign up log in tour help Search Q&A

Open Data beta

Questions Tags Users Badges Unanswered Ask Question

Open Data Stack Exchange is a question and answer site for developers and researchers interested in open data. Join them; it only takes a minute:

Sign up

Here's how it works:

- Anybody can ask a question
- Anybody can answer
- The best answers are voted up and rise to the top

Explore Our Questions active 1 featured hot week month

data-request usa geospatial data.gov api healthcare government openfda api-request machine-learning

more tags

4 votes	2 answers	45 views	Where can I find data on MARTA?	modified 2 hours ago chicks 111
1 vote	0 answers	1 view	Places to find credit risk data-sets	asked 3 hours ago Ashish Doshi 6
1 vote	2 answers	11 views	Where to find river flow time series for rivers in Europe?	modified 5 hours ago Jan Doggen 582

Site Stats

- 3,343 questions
- 4,427 answers
- 74% answered
- 11,499 users
- 1,856 visitors/day

more site stats on:

- area 51
- stack exchange



Did We Learn How To . . . ?

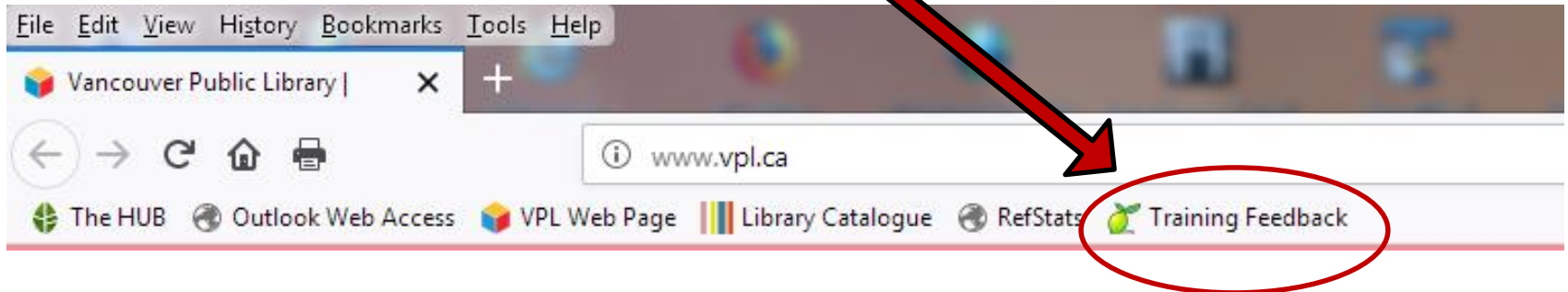
- ✓ Identify and locate open data sets on the internet.
- ✓ Work with a single data set to follow the steps to use open data to create something new.
- ✓ Distinguish between information license types and determine their impact on information access.



We want your feedback!

Please fill out an evaluation form to tell us what you thought of this class and help us improve our programs.

Online: www.vpl.ca/trainingfeedback



Thank you!

