Working With Open Data

Vancouver Public Library
Learning Opportunities @ VPL!

• Digital Essentials: vpl.ca/program/digital-essentials
Learning Opportunities @ VPL!

Lynda.com: vpl.ca/lynda

Learning Express Library: vpl.ca/digital-library/learning-express-library

VPL Research Guides: guides.vpl.ca
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
ACQUERING: 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)
• Sports Data (Sports Reference)
• Data Hubs (Qlik, Infochimps, CKAN, Google Public Data)
ACQUIRING: Searching for Data on the Web

Find pages with...
- **all these words**: Vancouver Open Data
- **site or domain**: .ca
- **file type**: Microsoft Excel (.xls)

To do this in the search box:
- Type the important words: "Vancouver Open Data"
- Put exact words in quotes: "Vancouver Open Data"
- Put a minus sign just before words you don't want: -site: .ca
- Put two full stops between the numbers and add a unit of measurement: 10.38 kg, £800–£800, 2010–2011

Then narrow your results by...
- **language**: any language
- **region**: any region
- **last update**: anytime

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

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

Show most relevant results

Find pages in the format that you prefer.

Find pages that are free to use yourself.
ACQUIRING: Searching for Hidden Data

Finding hidden data
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
Any Questions about Extracting (Downloading) Data?
CLEANING & PROCESSING

Turning data into diamonds
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?
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)
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:

<table>
<thead>
<tr>
<th>Year</th>
<th>Make</th>
<th>Model</th>
<th>Description</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Ford</td>
<td>E350</td>
<td>ac, abs, moon</td>
<td>3000.00</td>
</tr>
<tr>
<td>1999</td>
<td>Chevy</td>
<td>&quot;Venture &quot;Extended Edition&quot;</td>
<td></td>
<td>4900.00</td>
</tr>
<tr>
<td>1999</td>
<td>Chevy</td>
<td>&quot;Venture &quot;Extended Edition, Very Large&quot;</td>
<td></td>
<td>5000.00</td>
</tr>
<tr>
<td>1996</td>
<td>Jeep</td>
<td>Grand Cherokee</td>
<td>MUST SELL! air, moon roof, loaded</td>
<td>4799.00</td>
</tr>
</tbody>
</table>

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
1996,Jeep,Grand Cherokee,"MUST SELL! air, moon roof, loaded",4799.00
```
Using Excel to Create a CSV File

<table>
<thead>
<tr>
<th>Listing_type</th>
<th>Company</th>
<th>Website</th>
<th>Phone</th>
<th>Address</th>
<th>City</th>
<th>State/Province Code</th>
<th>Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental</td>
<td>Dentist</td>
<td><a href="http://www.userma.com">http://www.userma.com</a></td>
<td>(720) 333-1234</td>
<td>123 Main St, Denver, CO</td>
<td>CO</td>
<td>80204</td>
<td></td>
</tr>
</tbody>
</table>
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
- Creating an App
What is an API?
PUBLISHING: App Development

Welcome
From: Learning Android App Development

Learning Android App Development
with Doug Winnie
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
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.

<table>
<thead>
<tr>
<th>Field</th>
<th>Label</th>
<th>Definition</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Title</td>
<td>Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.</td>
<td>Always</td>
</tr>
<tr>
<td>description</td>
<td>Description</td>
<td>Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.</td>
<td>Always</td>
</tr>
</tbody>
</table>
PUBLISHING: Metadata Training
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?
• Maintain data and metadata regularly
• Check links
• Check user feedback
• 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?
Online Courses

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

Essentials
Data Fundamentals
Data Cleaning
Exploring Data
Extracting Data
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!