

DATA 301
Introduction to Data Analytics
Open Data

Dr. Ramon Lawrence
University of British Columbia Okanagan
ramon.lawrence@ubc.ca



What is Open Data?

Open Data is the movement to make data freely available to all with no restrictions on use or copyright.

Governments have been major supporters and providers of open data as data collected by governments is primarily done to benefit its citizens.

Corporations and other organizations are both producers and consumers of open data.

Open Data in Canada

Federal, provincial, and local governments have all been involved in the open data movement.

Canadian Federal government: <http://open.canada.ca/en>

- How to use: <http://open.canada.ca/en/working-data>
- Statistics Canada: <http://www.statcan.gc.ca/eng/rdc/data>

British Columbia government: <http://www.data.gov.bc.ca/>

City of Kelowna:

<https://www.kelowna.ca/city-services/city-maps-open-data/open-data-catalogue>

Open Data in Canada



Government of Canada

Gouvernement du Canada

[Français](#)

Search Canada.ca Q

Jobs ▾
Immigration ▾
Travel ▾
Business ▾
Benefits ▾
Health ▾
Taxes ▾
More services ▾

[Home](#) → Open Government

Open Government

Welcome to Canada's new Open Government portal. Here you can explore how the Government of Canada is working with the national and international open government community to create greater transparency and accountability, increase citizen engagement, and drive innovation and economic opportunities through Open Data, Open Information, and Open Dialogue.

Follow:



Search

Open Data
Open Government
Open Information

Search through our Open Data Portal.

Q

About Open Government

Explore Canada's Open Government commitments, discover Open Government programs across Canada, and learn more about the

Open Data

Search Government of Canada data, learn how to work with datasets, see what people have done with our data.

Open Information

Search Government of Canada digital records, completed Access to Information request summaries, and contracts over \$10,000.

Open Dialogue

Interact with Open Government experts and enthusiasts and share your ideas for new innovative approaches to Open Government.

Winners Showcase





CODE
CANADIAN OPEN DATA EXPERIENCE

And the #CODE2015 winners are...

◀ Item 5 of 5 ▶
⏸ Pause

Latest blogs

- [Library and Archives Canada: Opening Government Records](#)
- [Open Data Across Canada – A Snapshot](#)
- [Open Government, and the 4th Industrial Revolution](#)
- [Teaching Open Data](#)
- [View all blogs](#)

New and updated datasets

- [Financial Performance Data 2014](#)
- [Canadian Occupational Projection System \(...\)](#)
- [2016 Monthly pipeline occurrence statistics](#)
- [2016 Monthly aviation occurrence statistics](#)
- [2016 Monthly railway occurrence statistics](#)
- [2016 Monthly marine occurrence statistics](#)

Open Data in BC

Search the BC Data Catalogue 🔍

Applications & Web Services

The Province has a variety of Applications and Web Services that help you find and use information. Click the links below to see the applications and web services in BC Data Catalogue.

- 👁 View the Applications
- 👁 View the Web Services

Map of British Columbia showing various locations. A callout window displays a webcam view of a road through a forest, dated Sep 11, 2009 12:54 PM. Below the image, it says 'View 99.1 km south of Squamish, looking south' and 'Get Directions To here - From here'.

- What is DataBC?
- Open Data Initiative
- Geographic Services
- Datasets
- Applications & Web Services**

Featured

- [Fire Locations - Current](#)
(Geospatial)
- [MSP Blue Book](#)
(Non-Geospatial)
- [BC Personal Information Directory \(PID\)](#)
(Non-Geospatial)
- [FYE10 - Employees with a gross salary of \\$75...](#)
(Non-Geospatial)
- [FYE10 - BC Government Transfers](#)
(Non-Geospatial)

Fire Locations - Current

March 09, 2011

Wildfire location points for the current fire season. This includes both active and inactive fires. Data is supplied through various sources and is updated from the operational systems every two hours. [Learn more](#)

Latest Additions

DataBC Blog

[DriveBC Webcam Open Data Released, BCDevExchange Discovery Day March 16](#)
March 10, 2016

By Gord Smith, Web Services Manager, Business Management Services Branch, Ministry of Transportation and Infrastructure

A picture is worth a thousand words, especially if it helps plan a safe journey. DriveBC, one of government's most popular and valued services, lets commuters, truckers and

Open Data in Kelowna




[City Hall](#)
[Residents](#)
[Business](#)
[Visitors](#)
[Search](#)

[Home Page](#) > [City Hall](#) > [City Maps](#) > [Open Data Catalogue](#)

Open Data Catalogue

The City of Kelowna is recognized as one of the most open and accessible local governments in Canada. Open Data gives those interested in certain aspects of municipal services or infrastructure easy access to data.

This Open Data option provides direct, free access to download the data in a variety of common, popular [data formats](#). These datasets are made available under the Open Data Commons [Public Domain Dedication and License](#).

The City of Kelowna does NOT warranty the accuracy or completeness of this information and no representations are being made by providing this data. Any reliance on this information will be solely at YOUR OWN RISK and not that of the City.

Google  or Bing  links will result in a display of the dataset in those mapping environments. The City takes no responsibility for the accuracy of Google Maps and Bing Maps.

Contact: opendata@kelowna.ca

[A](#) [B](#) [C](#) [D](#) [E](#) [E](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

Dataset	FGDB	SHP	DWG	KML	Other Formats	Google Maps	Bing Maps
B							
Bicycle Network	FGDB	SHP	DWG	KML		G	
Business Licenses					CSV		
C							
City Boundary	FGDB	SHP	DWG	KML		G	B
Contours 1m 2015	FGDB	SHP	DWG				
Council Contact Information					CSV		

Open Data in United States

United States government: <https://www.data.gov/>

Individual states have their own open data sites as well.

- Example: Washington state: <https://data.wa.gov/>

United States: Data.gov

The screenshot shows the Data.gov homepage. At the top left is the Data.gov logo with the American flag. To the right is a navigation menu with links for DATA, TOPICS, IMPACT, APPLICATIONS, DEVELOPERS, and CONTACT. Below the navigation is a blue banner with the text "The home of the U.S. Government's open data" and a subtext "Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more](#)." Below the banner is a "GET STARTED" button with the text "SEARCH OVER 194,266 DATASETS" and a downward arrow. Below the button is a search bar containing the text "Credit Card Complaints" and a search icon. Below the search bar is a "BROWSE TOPICS" section with a grid of 14 topic icons and labels: Agriculture, Business, Climate, Consumer, Ecosystems, Education, Energy, Finance, Health, Local Government, Manufacturing, Ocean, Public Safety, and Science & Research.

DATA.GOV DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT

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 194,266 DATASETS

Credit Card Complaints

BROWSE TOPICS

Agriculture	Business	Climate	Consumer	Ecosystems	Education	Energy
Finance	Health	Local Government	Manufacturing	Ocean	Public Safety	Science & Research

Open Data Worldwide

UK: <http://data.gov.uk>

The World Bank: <http://data.worldbank.org/>

- Financial information and statistics

United Nations: <http://data.un.org/>

OECD: <https://data.oecd.org/>

Open Data Aggregators

There are many sites that aggregate open data sets (and some data sets for a cost). A Canadian based site is Quandl (<http://www.quandl.com>).

Kaggle provides many data sets and competitions and techniques for data analytics and machine learning.

<https://www.kaggle.com/datasets>

Open Data from Companies

Many companies either have public data or application programming interfaces (APIs) that allow people to use their data.

- Google: <https://www.google.com/publicdata/directory> (public data explorer) and <https://developers.google.com/maps/> (Google Maps API)
- Facebook: <https://developers.facebook.com/> (API)
- reddit: <https://www.reddit.com/dev/api> (API)
- Twitter: <https://dev.twitter.com/rest/public> (API)
- Amazon: <https://aws.amazon.com/public-data-sets/> (public data sets) and <https://developer.amazon.com/> (API for developers)
- Best Buy: <https://developer.bestbuy.com/> (API)

Try it: Open Data

Explore the federal, provincial, and City of Kelowna data sets to discover "something interesting". Report to your neighbors and to the class.

From any Canadian government open data site, retrieve a data set and analyze and visualize it using one of our tools: Excel, R, Python, Tableau.

Open Data for Researchers

Increasingly publicly funded researchers are responsible for making their data sets, procedures, and results available to the public (and other researchers).

- Canadian researchers funded by NSERC, SSHRC, CIHR must make their publications freely available within 12 months of publication.
- Researchers in bioinformatics and other fields must make their data sets publically available in a database or repository.

Researchers benefit by having access to public data sets and data sets of other researchers, but there is also a challenge as producing data sets (and perhaps commercializing results) may restrict open access.

Open Data Biology/Bioinformatics

Huge number of databases with most related to NCBI
(<http://www.ncbi.nlm.nih.gov/>) but distributed world-wide.

The screenshot shows the NCBI homepage with the following layout:

- Header:** NCBI logo, "Resources" and "How To" dropdowns, "Sign in to NCBI" link, and a search bar.
- Left Navigation Menu:**
 - NCBI Home
 - Resource List (A-Z)
 - All Resources
 - Chemicals & Bioassays
 - Data & Software
 - DNA & RNA
 - Domains & Structures
 - Genes & Expression
 - Genetics & Medicine
 - Genomes & Maps
 - Homology
 - Literature
 - Proteins
 - Sequence Analysis
 - Taxonomy
 - Training & Tutorials
 - Variation
- Main Content Area:**
 - Welcome to NCBI:** The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information. Includes links for "About the NCBI", "Mission", "Organization", "NCBI News", and "Blog".
 - Submit:** Deposit data or manuscripts into NCBI databases. (Icon: Upload arrow)
 - Download:** Transfer NCBI data to your computer. (Icon: Download arrow)
 - Learn:** Find help documents, attend a class or watch a tutorial. (Icon: Book)
 - Develop:** Use NCBI APIs and code libraries to build applications. (Icon: Code blocks)
 - Analyze:** Identify an NCBI tool for your data analysis task. (Icon: Scissors)
 - Research:** Explore NCBI research and collaborative projects. (Icon: Microscope)
- Right Side Panels:**
 - Popular Resources:** PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem.
 - NCBI Announcements:**
 - Register for the April 6th webinar: Using NCBI Databases with Tools that Predict Genomic Variant Effects (24 Mar 2016)
 - In two weeks, NCBI will release...
 - Register for the April 13th webinar, Submitting Data to NCBI and BioSample (23 Mar 2016)
 - In three weeks, NCBI staff will guide you through the process of submitting...
 - New NCBI video on YouTube provides strategies to search ClinVar efficiently (23 Mar 2016)
 - In the newest video on the NCBI YouTube channel, Search ClinVar with Face...

Open Data in Chemistry

ChEMBL (<https://www.ebi.ac.uk/chembl/>) stores structures and properties of pharmacologically active molecules.

- Over 1.5 million compounds.

SureChEMBL (<https://www.surechembl.org>) is a database extracted automatically from patent applications.

- Growing at 80,000 compounds a month and has 16 million compounds from over 13 million annotated patents.

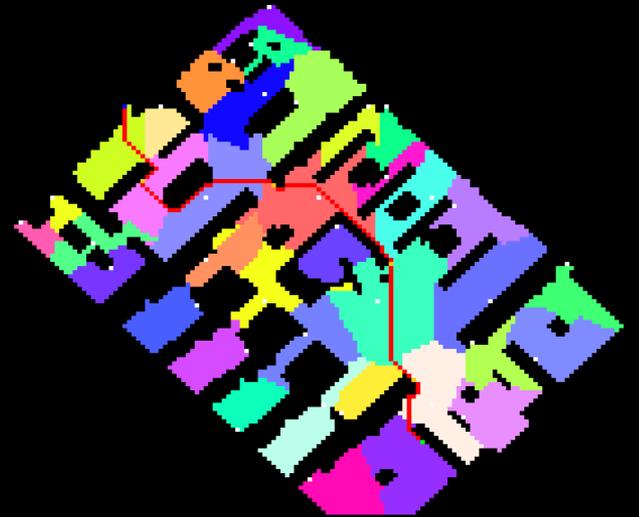
ChemSpider (<http://www.chemspider.com/>) is a free chemical structure database containing over 43 million structures.

- Supported and hosted by Royal Society of Chemistry.

Open Data in Computer Science

Computer scientists in various fields create standardized data sets for experimentation and research.

- Databases: Standard performance benchmarks such as TPC (www.tpc.org).
- Machine learning/data mining: UCI ML repository <http://archive.ics.uci.edu/ml/>
- Game path finding: <http://www.movingai.com/benchmarks/>



Open Data in Earth/Environment Science

- Climate Change Data Portal: <http://sdwebx.worldbank.org/climateportal/>
- National Climatic Data Center: <https://www.ncdc.noaa.gov/cdo-web/>
- National Geographic Data Center: <http://www.nodc.noaa.gov/submit/>
- Polar Data Catalog: <https://www.polardata.ca/>



Open Data in Physics

Modern physics produces a HUGE amount in data in experiments like astronomical observations and the Large Hadron Collider.

- New research systems developed to handle the large amount of data produced.
- CERN open data portal: <http://opendata.cern.ch/>
- Data produced is tens of petabytes/year. Large distributed computing of 170 facilities in 36 countries.

Astronomy:

- Canadian Astronomy Data Centre: <http://www3.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cadc/>
- National Space Science Data Center: <http://nssdc.gsfc.nasa.gov/>

Open Data in Psychology and Social Sciences

Archaeology:

- Archaeology Data Service: <http://archaeologydataservice.ac.uk/>
- Many museums have online exhibits and open data.

Psychology:

- Journals increasing requiring open data sets.
- List of open data sites at: <http://guides.library.ucla.edu/c.php?g=180221&p=1188487>

Baptisms administered in the Territory of the
Hudson's-Bay Company, in the Year 1820.

When Baptized	Child's Name	Parents Name	Quality	By whom the Ceremony was performed
1820 Nov. 11 th	Young George	Robertson & his regular Wife	Red-River	John West
1820 Nov. 26 th	John	Robert Campbell	Red-River	John West
1820 Nov. 26 th	John	Magnus Spence & an Indian Woman	Red-River	John West
1820 Nov. 26 th	Mary	Magnus Spence & an Indian Woman	Red-River	John West
1820 Nov. 26 th	Christy	Magnus Spence	Red-River	John West

History

- Digital Archive Database Project (UBC): <http://dadp.ok.ubc.ca>

Google Analytics

Google Analytics is an analysis service for tracking, optimizing, and understanding user interaction with a web site/service.

Using Google analytics is important for all business, but especially web companies, that rely on users interacting with their site to generate revenue and sales.

Google analytics helps identify and improve content to make it more accessible to potential customers.

- **Very important skill set for business owners and managers.**

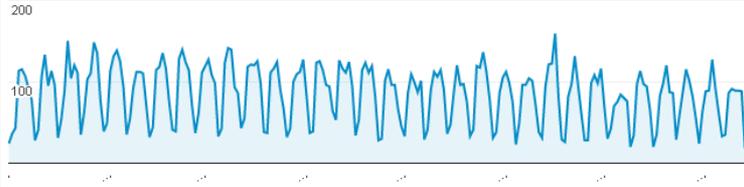
Google Analytics - Audience Overview



Google Analytics - Traffic Dashboard

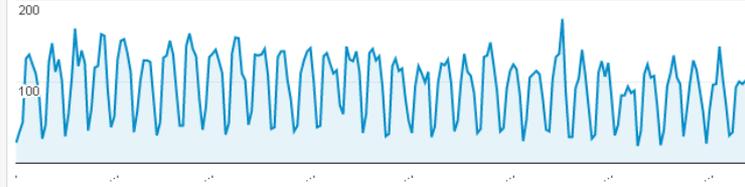
New Users

● New Users

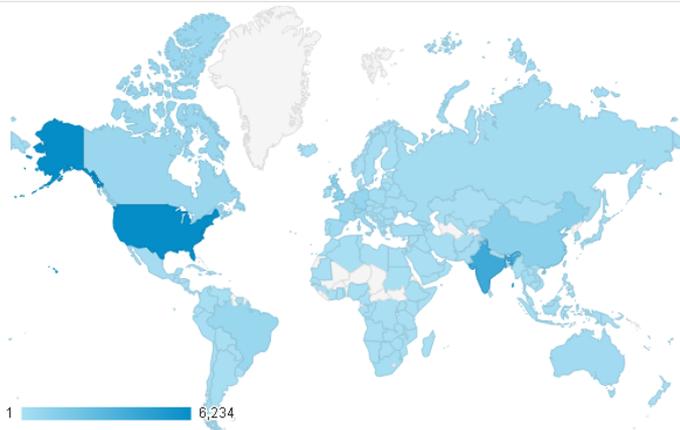


Users

● Users



Sessions



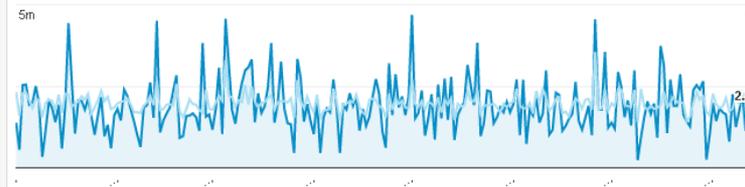
Goal Completions

● Goal Completions

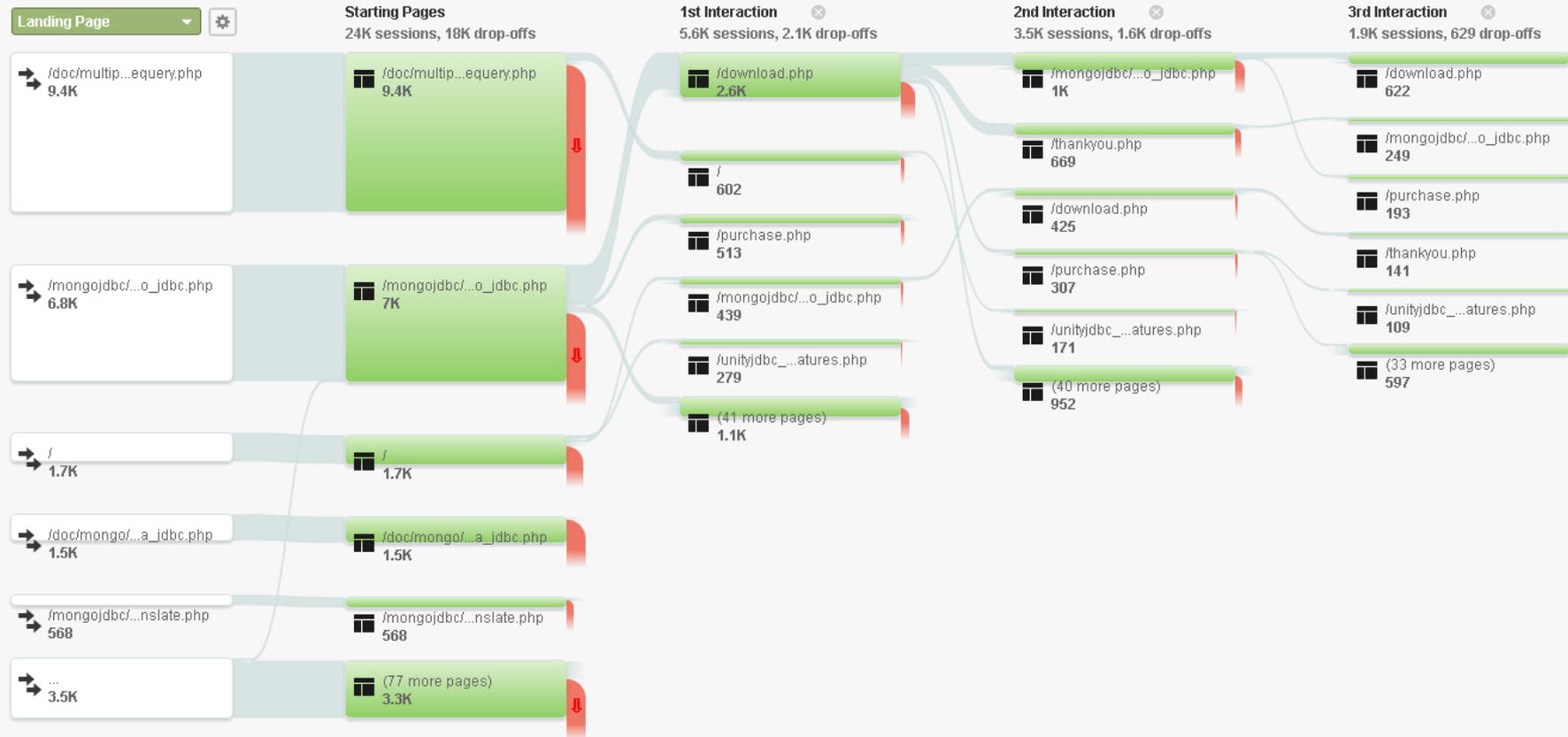


Avg. Session Duration and Pages/Session

● Avg. Session Duration ● Pages/Session



Google Analytics - Behaviour Flow



Google Adwords

Google AdWords is a service to provide advertisements during searches and as display advertisements on web sites and in apps.

- Primary source of revenue for Google. <https://www.google.ca/adwords/>
- Companies bid on *keywords* and display opportunities that are presented by Google and affiliated sites.

Terminology:

- *Ad Impression* - display of an advertisement. Pricing in cost-per-thousand impressions or cost per mille (CPM).
- *Click through* - user clicks on an advertisement (and directly to new location)
- *Click through rate* - fraction of impressions that are clicked on
- *Pay-per-click (PPC)* - companies are billed on each click of an advertisement. The pricing depends on the bid amount and the desirability of the ad location.

Conclusion

Open Data is the movement to make data freely available to all with no restrictions on use or copyright.

Open data has been widely supported by governments and companies wishing to engage users (and developers) with their services.

Data analysts should use open data to help with their analysis whenever available.

Researchers are often responsible for making their publications and data available in an open fashion.

Google provides services for analytics and advertising that are valuable to understand as a business or site looking for user traffic.

Objectives

- Define open data and explain the motivations for making data "open".
- List some of the governments and organizations that provide data in an open fashion.
- Use open data sets when applicable when performing data analysis.
- Explain the role of Google Analytics and Google AdWords. Compare and contrast what these two services provide.