

UBC Social Ecological Economic Development Studies (SEEDS) Sustainability Program

Student Research Report

**Environmental Scan of University Open Data Strategies**

**Yang (Julia) Zhu**

**University of British Columbia**

**LIBR 597**

**Themes: Community**

**July 29, 2019**

*Disclaimer: "UBC SEEDS Sustainability Program provides students with the opportunity to share the findings of their studies, as well as their opinions, conclusions and recommendations with the UBC community. The reader should bear in mind that this is a student research project/report and is not an official document of UBC. Furthermore, readers should bear in mind that these reports may not reflect the current status of activities at UBC. We urge you to contact the research persons mentioned in a report or the SEEDS Sustainability Program representative about the current status of the subject matter of a project/report".*

# Environmental Scan of University Open Data Strategies

Yang (Julia) Zhu

July 29, 2019

Supervisor: Dr. Luanne Freund

Supervisory Committee: David Gill

Luanne Freund

Lea Starr

Rohan Nuttall

Jonathan Kew

## Table of Contents

1. <i>Executive Summary</i> .....	3
2. <i>Introduction</i> .....	4
2.1. What is Open Data? .....	4
2.2. Motivations and Barriers.....	4
2.3. Open Data in Higher Education .....	5
2.4. Linked Open Data Initiatives .....	6
3. <i>University Open Data Initiative Overview</i> .....	8
3.1. Methods of Data Collection .....	8
3.2. University Open Data Initiatives: Case Studies .....	8
3.2.1 Harvard University .....	9
3.2.2 Cornell University .....	10
3.2.3 University of Waterloo .....	11
3.2.4 Concordia University.....	11
3.2.5 Southampton University .....	11
3.2.6 Oxford University.....	12
4. <i>Appendix A</i> .....	13
5. <i>Works Cited</i> .....	21

## 1. Executive Summary

The availability of open data has grown considerably, as public organizations responding to external pressure and internal motivation to release administrative data in formats and under conditions that enable reuse. Open data is an international phenomenon that is strongly associated with the open government movement and motivated by the desire for increased transparency and social and economic benefits made possible by data reuse. In the government realm, open data initiatives include the growing 2003 Public Sector Information (PSI) Directive in Europe, the Open Government Partnership established in 2011, and the expanding G8 Open Data Charter developed in 2013. In March 2011, the government of Canada launched its first-generation Open Data Portal and Canada is now a leader in the open data movement, currently chairing an international Open Data Working Group through its involvement in the Open Government Partnerships.

Given this context, we became interested in the impact of the open data movement on other, non-governmental public institutions, and specifically in higher education, where open data could be thought to bring the same or similar benefits. As a working group at UBC has begun to investigate the feasibility of establishing an open data initiative, I set out to determine: do university-based open data initiatives exist, what forms do they take, and what challenges do they face? I conducted an environmental scan of open data initiatives and data sharing practices within major Canadian and international universities and present the results in this report.

Based on the six case studies presented, the overarching goal of open data movement in higher education seems to focus on making data more easily accessible to members of the campus community and to improve overall decision making. At the same time, while some universities have chosen to shape their open data initiatives towards a more economic direction (targeted to app developers), other universities are more focused on data transparency.

Additionally, there appears to be little order or consistency in the data offered. Some portals have data and services; some contain only the data. What is rather surprising is that most appear almost random in terms of the datasets included – this speaks to the challenge of pulling together data from many masters. Some datasets are also not actually “open” in the sense that their PDF format is not machine-readable.

As open data is becoming increasingly utilized in governments, it is still a relatively new concept in higher education. Overall, open data initiatives in post-secondary institutions seem to be an area where little systematic work has been done. Some potential reasons for this could be due to the lack of a uniform or systematic policy or license agreement, and ambiguous guidelines on the classification of private or insensitive information.

## 2. Introduction

The UBC Data Governance Program identifies institutional data as a strategic asset for the university. One of the Program's principle guidelines is to ensure that such data is easily accessible to members of the UBC community. The university produces a number of datasets – if they are structured for usability, assessed for openness, and licensed appropriately, they could provide the foundation for a culture of sharing data on campus. One method of achieving this is through developing an open data framework. As a result, the UBC Open Data Working Group has been created to investigate the feasibility of establishing an open data portal at UBC. This report serves to inform the working group by offering an environmental scan of university open data strategies around the world, and by providing synthesis of key findings. The report's findings could enable students, faculty, researchers, and the wider UBC community to access diverse datasets and improve decision making.

The availability of open data has grown considerably, with public organizations responding to external pressure and internal motivation to release their raw data. In recent years, there have been increasing numbers of ongoing international open data programs with transparency and reuse of data being two of the main goals (Attard, Orlandi, Scerri, & Auer, 2015). Such initiatives include the growing 2003 Public Sector Information (PSI) Directive in Europe, the Open Government Partnership established in 2011, and the expanding G8 Open Data Charter developed in 2013 (Attard et al., 2015). In March 2011, the government of Canada launched its first-generation Open Data Portal in machine-readable formats, and in June 2013, the second-generation Open Government Portal was launched with additional functionality (Open Data 101, 2017). Canada has now become one of the leaders in international open data, currently chairing an international Open Data Working Group through its involvement in the Open Government Partnerships (Open Data 101, 2017).

### 2.1. What is Open Data?

To be considered “open”, the data must be online, machine-readable, open-licensed, available for bulk download and redistribution, and free of charge (Open Data Movement Reaches Turning Point, 2016). Governments, businesses, and individuals can use open data to bring about social, economic, and environmental benefits. Indeed, open data can include a broad range of information including health data, traffic data, real estate listings, geographical maps and meteorological data, and many more (Open Data, 2016).

### 2.2. Motivations and Barriers

One of the main issues that encouraged the founding of open government data initiatives such as the ones mentioned above, was corruption. Corruption is an international problem that threatens economic growth and undermines peoples' faith in the government, both of which adversely affect our society in general (Attard et al., 2015). It is believed that with the transparency involved in open data, open data initiatives could be a promising step towards minimalizing corruption. At the same time, open data would also be a vital tool for developing public policy and delivering valuable service to users.

Together, there are three main reasons for open government data:

- 1) *Transparency* – a complete disclosure of information, rules, plans, processes and actions of companies, organizations and government;
- 2) *Releasing social and commercial value* – data about the city, province, or nation have both social and commercial value which could be used for policy development or new service implementation; and
- 3) *Participatory Governance* – as government data is made public, citizens can feel more involved with government processes such as decision- and policy-making as opposed to passively voting in elections (Attard et al., 2015).

Collectively, improving government transparency through open data could empower citizens as they could use readily-available data to aid their decision-making processes.

Although open data initiatives have become increasingly adopted, the movement also has a number of barriers that prevent it from reaching its full potential. Attard et al (2015) identify the following barriers:

- 1) *Nature of data* – the heterogenous nature of data formats, ranging from PDF and CSV files to higher structured XML files and database records, make any comparison and analysis of data almost impossible;
- 2) *Diversity of tools* – the different possibility of tools to search, download and visualize government data means there is no uniform way of obtaining and understanding the data;
- 3) *Cultural differences* – as different regions of the world have their own cultures, some governments may be unwilling to publish their data because they may understand that as requiring extra resources, time, and effort;
- 4) *Low consumer data literacy* – low data literacy amongst consumers could contribute to low exploitation of government open data portals, and if the projected consumers of the dataset do not use it, then the potential of open government initiatives would not be fulfilled;
- 5) *Data discoverability* – though published as “open data”, this data may not always be discoverable because the metadata describing the data may be inaccurate or complete;
- 6) *Publishing open government data* – there is currently a lack of mutually agreed upon standards for internet data publishing;
- 7) *Data quality* – data quality is cross-disciplinary, subjective, and has not yet reached an agreed-upon definition;
- 8) *Budget provision* – as open data is still a relatively new concept, there may be limited allocated budget for open data government efforts;
- 9) *Competition* – open data can be seen as unfair competition for the private sector as it can allow competitors to invest minimal efforts to obtain the freely available open data; and
- 10) *Liability* – public entities fear being held responsible for any damage related to the use of provided public data because of possible data inaccuracies or inaccurate interpretation.

### 2.3. Open Data in Higher Education

In addition to the open data movement in some governments, several universities have also begun to participate in making their data public. Data produced by educational institutions are generally categorized into research data or “grey data” (Borgman, 2018). Grey data is an umbrella term that includes the huge array of data outside the research field which

universities accumulate (Borgman, 2018). Typically, grey data refers to reference data such as the location of academic institutions; internal data such as staff names, personnel data, or identity data; curriculum data such as learning objects or course data; or user-generated data like learning analytics, performance data, or job placements (Krumova, 2017). Arguably, one of the most valuable grey data universities collect is that of their students. By collecting grey data about their students, some universities create unique learning paths for each student or profile students according to their choice of majors, financial aid status, scholarships, and other matters (Borgman, 2018).

With all these information produced and collected, universities are stewards of immense amounts of data that can provide countless opportunities for local and global partnerships in teaching, research, and strategic planning (Borgman, 2018). Indeed, as public institutions, universities should commit to an openness of information via open access. Universities need to be open to support collaborative knowledge exchange to maintain public trust and confidence (Krumova, 2017). However, because these data can also have many origins and take various forms, exact data ownership may be ambiguous. Specifically for the case of grey data, it is difficult to allocate responsibility for these types of information as grey data is highly diffuse in the university environment – those who collect the grey data may become the stewards or they may pass the information along to other stewards (Borgman, 2018). Possible grey data stewards include libraries, instructional development, food services and student stores, to list a few (Borgman, 2018). Furthermore, the methods and financial resources required for effective data stewardship in universities are still currently poorly understood (Borgman, 2018). Still, data management is relatively difficult compared to governmental data management (Borgman, 2018).

The university community's stakeholders – students, faculty, and staff – expect the institution to achieve a balance between academic and intellectual freedom while maintaining a reasonable level of confidentiality, security, and responsibility towards their data (Borgman, 2018). At the same time, audiences extending beyond the university community also expect the university to be an educational institution that is impartial, transparent, and reliable with respect to its resources (Borgman, 2018). As universities become increasingly engaged in intra – and – inter institutional collaborations, good stewardship requires having frameworks to guide the release of certain kinds of data while preventing the release of others (Borgman, 2018).

#### 2.4. Linked Open Data Initiatives

Despite these challenges, many universities continue to build their datasets with the purpose of eventually connecting them to various existing Semantic Web data to create a pool of related data. This would not only share, but it would also connect pieces of data and knowledge on the Semantic Web. This process is known as linking open data. Traditionally, academics and students go through the laborious task of researching information and mentally linking conceptual data across all the resources they used to generate their own ideas about the topic (Zablith, Fernandez, & Rowe, 2012). If related data was virtually connected to each other, this linked open data system could generate the necessary structures and connections to make the process of information research much more efficient.

Indeed, there has been a substantial increase in the number of UK universities adopting linked open data initiatives, under the leadership of the national Joint Information Systems Council (JISC). For instance, the Department of Computer Science at the University of Sheffield provides a Linked Data Service that semantically links together all the university's research groups and their publications (Zablith et al., 2012). Similarly, the University of Southampton has a Linked Open Data portal which stores the university's open data (Zablith et al., 2012).

Universities in the UK are not the only institutions with a developing linked open data system. Tsinghua University in China, for example, has built five datasets based upon data crawled from the public university website (Ma, Xu, Bai, & Li, 2012). Their Linked Data project aims to cooperate with other participating universities to build a linked open data system where academic institutions share their datasets via existing Semantic Web data (Ma et al., 2012). This ultimately would create easily-accessible and diversified datasets from a variety of universities around the world, which could in turn inspire other educational institutions to share their data as well, further enhancing the linked data system (Ma et al., 2012).

Not all data have equal priority in linking, however. Ma and his colleagues emphasize that some of the most important university data include basic institutional information such as university departments, university size, and university location; campus information such as the geographical description of campus and university buildings; faculty information such as the basic information about researchers and staff members as well as their contact information; and educational administrative information such as those of students and their courses (Ma et al., 2012).

In addition, the University of Münster in Germany also has a linked open data project that aims to publish select university data as Linked Open Data (Keßler & Kauppinen, 2012). The Linked Open Data of Münster (LODUM) project was established to “improve the transparency and visibility of the university, publishing any non-sensitive data [which includes scientific data and publications, and certain administrative data] online following the Linked Data principles.” (Keßler & Kauppinen, 2012) The LODUM project is designed to be a long-term strategy which will open up and link different data sources across the university's 15 faculties and departments (Keßler & Kauppinen, 2012). As the first German university to implement an institution-wide Linked Open Data program, following the early examples from the UK, this signifies that linked data in science and higher education is gaining momentum (Keßler & Kauppinen, 2012).

Overall, there is not much written specifically on open data movements in higher education, and it is apparent that it is not a widespread phenomenon. Much of what has been published on the topic of open data is closely associated to linked open data, which is not currently an identified priority for the UBC project. Nonetheless, Borgman's 2018 paper provides some valuable insights on the opportunities and challenges of open data in higher education.



### 3. University Open Data Initiative Overview

This section reports on an analysis of select universities to determine how common open data initiatives are, and in cases where they do exist, to describe them in more detail. The results of the analysis show that it is not common among Canadian U15 universities to have open data policies or initiatives, but they do consistently share certain standard datasets in the form of Fact Books. In fact, there is a Fact Book trend among a majority of large Canadian and American universities where they share institutional data most often in PDF format; the Fact Books are not yet commonly found in open format. However, there is some adoption of open data among large US universities, including Harvard, Yale, and Cornell. Further, none of the U15 Canadian universities had explicitly-stated policies or guidelines on how the data was gathered or how the data is made available online: there was no open data policy.

#### 3.1. Methods of Data Collection

Information on Canadian university open data initiatives have been mostly gathered from U15 Canadian Research Universities. I chose to primarily focus on these institutions because the U15 group of Canadian Research Universities is a collective of some of Canada's most research-intensive universities. Hence, they should have a plethora of available data which they may have chosen to make public via an open format. A similar rationale was applied to selecting American and Australian universities for research. The selection criteria for foreign universities also included student enrolment (at least 30,000 total students) and 2018 QS rankings (top ten universities in America and top three universities in Australia).

To find information about open data initiatives in selected universities, I performed a series of Google keyword searches. I approached this firstly by searching for open data initiatives in a particular university. During the research process a challenge I had faced was when I noticed that many Canadian universities often had no information on this topic. Of all the U15 Canadian universities, only the University of Waterloo has an easily findable, public open data initiative. As I searched further, I noticed that almost all U15 Canadian universities have a Fact Book instead – where universities publish institution data such as course data, infrastructure data, student and staff headcounts, and financial data – most often in a PDF format. This inspired me to also search for Fact Books or their equivalents in American and Australian universities.

All foreign universities researched indeed have Fact Books or the equivalent. I also looked for open data initiatives among select foreign institutions; only Harvard University and Cornell University offered such movements.

The results of the analysis for all universities is presented in Appendix A. Those universities that did seem to have an open data initiative were explored in more depth, and these results are presented in the Case Studies section to follow.

#### 3.2. University Open Data Initiatives: Case Studies

Below are six case studies on select Canadian U15 universities, US universities, and UK universities that have an open data initiative. The case studies discuss the kinds of open data made available, the intended audience, the department responsible for managing the open

data initiative, the search interface, and whether there are policies or licenses associated. Those included in the case study are Harvard University, Cornell University, University of Waterloo, Concordia University, Southampton University, and Oxford University.

### 3.2.1 Harvard University

Data at Harvard University is shared through the data sharing platform, Dataverse, found at Harvard University's Open Data Page. The Dataverse is managed by Harvard University's Information Technology Department. Harvard's Open Data Page claims that while some of their datasets are open to the world, others are only accessible to members of the university. However, all datasets were able to be downloaded and opened. A majority of files were in PDF format. There was no policy or license agreement listed on the Open Data Page. The intended audience appears to be the public. The open data provided is mostly administrative information.

The Dataverse has 23 datasets:

1. Operational Data Store (ODS) for the my.Harvard Student Information System
2. Identity and Access Management Attributes
3. Harvard Worldwide Activities (offers glimpse at Harvard University's strong connections to students, faculty, and other colleagues and universities around the world)
4. Resources for Harvard Identity and Access Management Data Customers
5. Harvard Catalyst Profiles (a Semantic Web application)
6. GIS Data Layers
7. Harvard Common Data Set (responses to the Common Data Initiative)
8. HUPD Annual Security Report (holds information regarding Harvard University Police Department's crime and reporting statistics)
9. Harvard University Higher Education Opportunity Act Fire Safety Report
10. Canvas LMS API (includes a REST API for accessing and modifying data externally)
11. Harvard Art Museums API (REST-style service designed for developers)
12. Course Enrollment Statistics (enrollment statistics for the most recent semester)
13. Harvard Graduate school tuition
14. Annual Financial Report (university's financial position and results)
15. Group Fitness Classes
16. Harvard College Tuition (tuition broken down into categories for every year since 1985)
17. Public Harvard Directory (a tool to look up Harvard students, faculty, or administrators and access to their phone numbers and emails of each individual)
18. Shuttle Route and Schedule (poster containing a map of shuttle stops and a schedule of all shuttle routes)
19. Harvard Faculty Finder (an institution-wide view of the breadth and depth of Harvard faculty and scholarships)
20. Library Cloud (this is the catalog metadata about Harvard Library's collection)
21. Harvard Library Bibliographic Dataset (contains over 12 million bibliographic records for materials held by the Harvard Library)
22. Digital Access to Scholarship at Harvard (digital repository for scholarly articles, theses, and dissertations)
23. Course Planner API

### 3.2.2 Cornell University

Data at Cornell University is managed through the Cornell Open Data Initiative (CODI). CODI is an initiative that attempts to create an open and easily accessible data repository for the Cornell community. First started in 2016 with the help of students from Cornell AppDev, Cornell Data Science, and Cornell University Sustainable Design, the CODI data and platform shares university data directly through their CODI homepage. The page is currently managed by Cornell AppDev in the university's Information Technology department. The intended audience appears to be the public.

While the CODI team does not explicitly mention a policy or license agreement, it pays consideration to Cornell's IT infrastructure by making minimal direct impact on network resources during their data gather and distribution process. The team further avoids undermining any official or existing resources for gathering such machine-readable data.

CODI has four existing open datasets on the subjects of map data, transit data, calendar data, and dining data; it also has one impending dataset on academic data:

1. Map data includes:
  - Bike racks data
  - Bluelights data
  - Buildings data
  - Campus to campus data
  - Carshare data
  - Diaper changing stations data
  - Info booths data
  - Lactation rooms data
  - Parking data
  - Park mobile data
  - Virtual tour data
2. Transit data includes:
  - a. Route schedule data
  - b. Stop schedule data
  - c. Stop locations data
3. Calendar data includes:
  - a. Cornell events data
4. Dining data includes:
  - a. Cornell dining data
5. Academic data:
  - a. To be added

### 3.2.3 University of Waterloo

The University of Waterloo Open Data API is powered by the university's Open Data Initiative. The initiative allows anyone to build their own programs and applications using data from the University of Waterloo's websites, hinting the intended audience to be public. The open data at Waterloo University is shared through the data sharing platform, Dataverse, found at Waterloo University's Open Data Initiative webpage. The Dataverse is managed by Waterloo University's Information Systems and Technology Department.

The API offers more than 40 ways of accessing various datasets across the University of Waterloo network. To gain access to the datasets, the user will first need to register for an API key by filling an online registration form with the user's full name and email address. Additionally, users must follow the Open Data License Agreement for the terms of service and licensing information. The Agreement states that the Information Providers grants users an unlimited, royalty-free, perpetual, and non-exclusive license to use the Information, and the Agreement has no effect on any copyright exceptions or limitations available under the Canadian Copyright Act.

A user is able to:

1. copy, publish, distribute, and transmit the Information;
2. adapt, modify, and/or translate the Information; and
3. use the Information commercially

The Open Data License Agreement does not cover the use of:

1. personal information;
2. third-party rights the Information Provider is not authorized to license;
3. information or records that are inaccessible pursuant to the Freedom of Information and Protection of Privacy Act of Ontario; and
4. the Information Provider's existing logos, trademarks, slogans, or crests where they could be included in the Information

### 3.2.4 Concordia University

Concordia University's Open Data API is powered by the university's Webster Library Transformation Project technology program. The initiative's data is machine-readable and open to the public for access. Each dataset submitted undergoes a formal review process, where sensitive, personal and individualized information is omitted prior to being made open. This way the university's privacy and confidentiality guidelines are respected. To access the datasets, the user needs to first register online go obtain a free API developer key. The user can also access the raw data on the Open Data webpage. The datasets are managed by representatives from Instructional and Information Technology Services (IIST), Concordia Library, and Enrolment Services.

### 3.2.5 Southampton University

Data at Southampton University is shared through the University of Southampton's Open Data Service. There was no policy agreement listed on the Open Data Service page; however, their datasets were mostly published under the Open Government License, or other licenses conforming to the Open Definition. The intended audience appears to be the public. The

university's Open Data Service also does not explicitly mention which department manages this initiative.

The university offers seven open datasets:

1. Buildings and Places (a list of buildings ordered by number and name, as well as official university campus IDs)
2. Organization (the structure of the university and its sub-parts)
3. Points of Service (the points of sale or service in and around the university)
4. Research Facilities (a list of the university's facilities and major research equipment)
5. Term Dates (the dates that the university terms and semesters begin and end)
6. Products and Services (useful products and services available in and around campus)
7. Datasets (an index of all the datasets maintained)

### 3.2.6 Oxford University

Data at Oxford University is shared through the University of Oxford's institutional linked open data store. The Open Data program is managed by IT Services at the University of Oxford. Their initiative aims to collect data from around the university and make it available for reuse as linked data under open licenses. There was no policy or license agreement listed on the Open Data Program page. The intended audience appears to be the public. The site states that all data are also available through their SPARQL endpoint. Their Open Data initiative also supports content negotiation, allowing users to request data in a variety of RDF serializations.

The university's Open Data webpage has four open datasets. The datasets have their own publishers, licenses, and descriptions:

1. Dataset catalogue for the University of Oxford is published by the University of Oxford
2. OxPoints is published by IT services and is the university's geospatial, temporal, and organizational RDF store
3. Research equipment and facilities at the University of Oxford is published by Research Services and contains metadata about research equipment and facilities
4. Vacancy data from the University of Oxford is published by the University of Oxford and contains information about vacancy advertisement information from the university's central recruitment site

Based on the five case studies presented above, the overarching goal of the open data movement in higher education seems to focus on making data more easily accessible to members of the campus community and to improve overall decision making. At the same time, while some universities have chosen to shape their open data initiatives towards a more economic direction (targeted to app developers), other universities are more focused on data transparency.

Generally, there appears to be little order or consistency in what data is offered. Some datasets are also not actually "open" in the sense that their PDF format is not machine-readable (note that one of the criteria for data to be considered "open" is that the data must be

machine-readable). Furthermore, while some open data initiatives are currently still active, others – such as the Cornell University Open Data Initiative – may not be.

Overall, this seems to be an area where little systematic work has been done. In this sense, UBC has an opportunity to be a leader in this space. Given that UBC has undertaken a major Data Governance initiative for data that is used internally, it may be well suited to initiate an open data strategy as well. Developing a small-scale open data movement that is well thought-out and coordinated, with a clear plan for longevity and sustainability, would be valuable and show leadership in this space.

## 4. Appendix A

The Appendix contains information on Canadian U15 universities and Carleton University, top 10 American universities (according to 2018 QS National University Rankings), and top 3 Australian universities (according to 2018 QS Australian University Rankings).

### **University of Calgary (Canadian U15):**

- Fact Book that provides statistical information about the institution. The primary purpose of the Fact Book is to provide planners and decision-makers with the information they need to assess and improve the programs and services.
- The Fact Book contains information about:
  - Student summary on undergraduate, graduate enrollment
  - Student detail data on the number of full time, part time, co-op students
  - Degrees and diplomas granted by gender and faculty
  - Course enrollees by faculty, department, and level
  - Grade in terms of percentage grade distribution
  - Staff headcounts by full time, part time academic staff by rank
  - Library collections

- Link: [https://oia.ucalgary.ca/fact-books/fact-book-20172018#quickset-field\\_collection\\_quicktabs\\_2](https://oia.ucalgary.ca/fact-books/fact-book-20172018#quickset-field_collection_quicktabs_2)
- *Managed by office of institutional analysis*

#### **University of Alberta (Canadian U15):**

- University library shares some datasets, including metadata sets, financial, desk tracker data in the UAL Dataverse network
  - Link: <https://dataverse.library.ualberta.ca/dataverse/ualib>
- The Fact Book contains information about:
  - Student headcounts in respective undergraduate and graduate programs
  - Staff headcounts in terms of full time, part time, and position
  - Link: <https://www.ualberta.ca/reporting/statistical-reports>
  - *Managed by strategic analysis and data warehousing*

#### **Dalhousie University (Canadian U15):**

- Has an page online with enrollment reports over the last 11 years
  - Enrolment by faculty, by gender, by place of permanent residence, by major
  - Link: <https://www.dal.ca/dept/oia/Advanced-analysis-planning/enrollment.html>
  - *Managed by Dalhousie Analytics*

#### **University of Manitoba (Canadian U15):**

- Their Office of Institutional Analysis shares information about:
  - Enrolment reports (undergrad & grad, Indigenous, International)
  - Courses, degrees, grades, demographics
  - Program costs, tuition fees
  - Faculty and staff
  - Alumni
  - Finance
  - Publications
  - Link: <http://umanitoba.ca/admin/oia/index.html>
  - *Managed by Office of Institutional Analysis*

#### **University of Winnipeg (Canadian U15):**

- The Fact Book contains information about:
  - Headcount of undergraduate and graduate students, International and Indigenous students
  - Link: <https://www.uwinnipeg.ca/institutional-analysis/university-data-and-statistics.html>
  - *Managed by Office of Institutional Analysis*

#### **McGill University (Canadian U15):**

- The Fact Book contains information about:
  - Enrolment reports (undergraduate, graduate, degrees, doctoral, postdoctoral)
  - Link: <https://mcgill.ca/es/registration-statistics>
  - *Managed by Enrolment Services*

### **McMaster University (Canadian U15):**

- The Office of Institutional Research and Analysis published a Fact Book that contains information about:
  - University statistics on admission, degrees awarded, enrolment, faculty, retention and graduation data
  - Link: <https://www.mcmaster.ca/vpacademic/documents/FactBook2017-2018.pdf>
  - *Managed by Office of Institutional Research and Analysis*
- The Office of Institutional Research and Analysis also offers a page on the most frequently requested data and statistical information, including:
  - Business intelligence (data on areas of strategic importance to the university)
  - Canadian Graduate and Professional Student Survey, CGPSS (information about graduate student satisfaction and experience)
  - Canadian University Survey Consortium, CUSC (survey about what students expect of their university)
  - Common University Data Ontario, CUDO (key data such as enrolment, student satisfaction, fees, class size, research, graduation, and employment about Ontario's universities)
  - Link: <https://ira.mcmaster.ca/category/data/>
  - *Managed by Office of Institutional Research and Analysis*

### **University of Ottawa (Canadian U15):**

- The Fact Book contains information about:
  - Enrolment by program, retention and graduation, degree, and academic and support staff
  - Link: <https://www.uottawa.ca/institutional-research-planning/resources/facts-figures>
  - *Managed by Institutional Research and Planning*

### **Queen's University (Canadian U15):**

- The Fact Book contains information about:
  - Institutional data on student enrolment, faculty and staff, research intensity
  - Link: <https://www.queensu.ca/planningandbudget/institutional-data>
  - *Managed by Office of Planning and Budgeting*

### **University of Saskatchewan (Canadian U15):**

- The Fact Book contains information about:
  - Student headcount and demographics, student qualifications
  - Faculty and staff
  - Link: <https://www.usask.ca/isa/statistics/students/>
  - *Managed by Information and Communications Technology Reporting and Data Systems*
- They also have a Data management policy, excerpts include:
  - All units and members of the university community must access and use university data in ways that safeguard the data and protect the institution.
  - Units and members of the university community must ensure:



- Compliance with regulatory requirements, as well as third-party and other contractual data obligations.
- Data is used for the purposes for which it is collected and any restrictions for its use are observed.
- Data is collected, stored, and disposed of in ways appropriate to the risk and impact of unintended disclosure.
- For research data, the principal investigator is accountable for all decisions regarding their research data.
- For decisions regarding institutional data, such as access, classification and appropriate use, members of the university community must consult the designated individual that has accountability for the data. These roles and accountabilities are defined in the Data Governance Framework.
- Link: <https://policies.usask.ca/policies/operations-and-general-administration/data-management.php#Procedures>
- *Managed by Information and Communications Technology*

#### **University of Toronto (Canadian U15):**

- The Fact Book contains information about:
  - Student enrolment, faculty & staff, alumni, tuition fees
  - Financial and planning reports
  - Link: <https://www.utoronto.ca/about-u-of-t/reports-and-accountability>
  - *Not clear who is responsible*

#### **University of Waterloo (Canadian U15):**

- Has an Open Data Initiative:
  - Datasets are filed in Dataverse: <https://github.com/uWaterloo/Datasets>
  - To use their Open Data, users must go through a series of steps:
    - Step 1 – comply with their license:
      - User is free to copy, publish, distribute and transmit the information; adapt, modify, or translate the information; and use the information commercially
      - The Agreement does not cover the use of personal information, information or records that are not accessible pursuant to the Freedom of Information and Protection of Privacy Act of Ontario, third-party rights the Information Provider is not authorized to license, and the Information Provider’s logos, trademarks, slogans or crests where they may be included in the Information
      - Link to license: <https://uwaterloo.ca/open-data/university-waterloo-open-data-license-agreement-v1>
    - Step 2 – register for an Application Programming Interface (API) key:
      - Information required are full name and email address
      - Link: <https://uwaterloo.ca/api/register>
    - Step 3 – read the documentation:
      - To access the API, all calls are made to the URL <http://api.uwaterloo.ca/v2/>
      - In order to make an API call, the user must have a valid API key

- Link: <https://uwaterloo.ca/open-data/>
- *Managed by Information Systems and Technology (IST)*
- Also has a Fact Book which contains information about:
  - Enrolment and degrees, student headcounts, faculty performance indicators
  - Link: <https://uwaterloo.ca/institutional-analysis-planning/>
  - *Managed by Institutional Analysis and Planning*

#### **University of Western Ontario (Canadian U15):**

- Has a data book on academic unit
  - Link: <https://www.ipb.uwo.ca/facts.php>
  - *Managed by Office of Institutional Planning and Budgeting*

#### **University of Carlton (Canadian non-U15):**

- A page dedicated to available data, divided into 4 sections:
  - Surveys
    - International quantitative information sorted according to topic, such as crime and justice, drugs and alcohol, education, health, etc.
  - Online data repositories
    - Carleton University Data Repository Dataverse is the research data repository for Carleton University managed by library IT services department in conjunction with data services (research data)
    - Scholars Portal Dataverse network (Ontario Universities data collection)
    - Abacus Dataverse Network (BC)
    - International
  - Aggregate data
    - Presented as tables and time series, they are available for academic research or teaching
      - Beyond 20/20: data is restricted by Carleton IP address to current Carleton students, faculty, and staff
      - CANSIM: Canadian Socio-economic information and Management Database) is Statistics Canada's computerized database of time series covering social and economic aspects of Canadian life
  - Free data sources
    - International free data sources
- Data Use Restrictions:
  - Data files housed with Data Services are subject to licensing agreements
  - Most agreements stipulate that the data may be used by students, faculty and staff of Carleton University for academic and personal research only. Any use of these data for commercial applications or contract purposes will violate our licenses; use by members of the community outside of Carleton University could also violate our licenses
  - Read more at: <https://library.carleton.ca/about/policies/data-use-restrictions>
- Link: <https://library.carleton.ca/find/data/available-data>
- *Managed by: data services department in the Macodrum Library*

### Harvard University (USA):

- Harvard has an Open Data Page; however, some data may only be visible to the Harvard University community
  - There are 23 datasets that are filed in its Dataverse:
    - Link: <https://dataverse.harvard.edu/dataverse/harvardopendata>
    - The data were added from 2015-2016
    - The data are made available under a CC0 waiver, open to the world
    - The datasets contain information about:
      - Annual report on fire safety 2015
      - Harvard worldwide activities 2016
      - Identity and Access Management Attributes 2016
      - GIS data (information on buildings and grounds related to Harvard University)
      - Annual security report
      - Student enrollment information
  - Link to Open Data Page: <https://data.harvard.edu>
  - *Managed by: Harvard University Information Technology (UCIO)*
- The university also has an Open Data Catalog featuring dozens of publicly-available datasets from around Harvard University
  - Data include:
    - Academics, admissions, campus, faculty/staff, finances, library, safety, student life
    - Files types include API, database, CSV, excel, graph, HTML, PDF, visualization
    - Link: <https://hodp.org/catalog/?category=safety>
    - *Managed by: Harvard University Information Technology (UCIO)*
- The university library further has a library catalogue with bibliographic dataset:
  - The dataset has over 12 million bibliographic records for materials held by the Harvard Library, including:
    - Books, journals, electronic resources, archival materials, manuscripts, audio, scores, and other materials
    - Each item has its own citation metadata
  - Link:  
<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LZDQYN>
  - *Managed by: Harvard University Information Technology (UCIO)*

### Stanford University (USA):

- Offers data related to GIS
  - Link: <https://library.stanford.edu/research/stanford-geospatial-center/data>
  - *Managed by: Stanford University Library*
- Offers “Stanford Common Data Set” (equivalent to Fact Book)
  - Link: <https://ucomm.stanford.edu/cds/>
  - *Managed by: Office of University Communications*

### Yale University (USA):

- Yale YODA Project (Yale University Open Data Access)
  - Advocates for the responsible sharing of clinical research data
  - The mission of the YODA project is to increase access to clinical research data and to generate new knowledge
  - A signed Data Use Agreement (DUA) is required for data access through the YODA Project
  - Managed by: *Centre for Outcomes Research and Evaluation*
  - Link: <https://yoda.yale.edu>
- The university also has a Fact Book which contains information about:
  - Alumni, enrollment, tuition, financial aid, etc.
  - [https://oir.yale.edu/sites/default/files/factsheet\\_2015-16\\_0.pdf](https://oir.yale.edu/sites/default/files/factsheet_2015-16_0.pdf)

### **Princeton University (USA):**

- Has a page dedicated to presentations and reports:
  - Annual report of the Treasurer (focuses on concept of sustainability and commitment to responsible stewardship of financial and environmental resources)
  - Annual report on the University research board
  - Annual report of the priorities committee (recommendations and details concerning the Princeton University operating budget)
  - Princeton profile (answers the most frequently asked questions about the university and diverse resources available on campus)
  - Attitudes on sexual misconduct at Princeton
  - Link: <https://ir.princeton.edu/university-factbook/other-university-data/presentations-and-reports>
  - *Managed by: Office of Institutional Research*
- Has a page on common data set:
  - a collaborative effort among data providers in the higher education community and publishers as represented by the College Board, Peterson's and US News and World report
  - aim is to improve the quality and accuracy of information provided to all involved in a student's transition into higher education
  - equivalent to Fact Book
  - Link: <https://ir.princeton.edu/other-university-data/common-data-set>
  - *Managed by: Office of Institutional Research*
- Has a page on Information Security Policy:
  - The purpose of this policy is to provide a security framework that will ensure the protection of University Information from unauthorized access, loss or damage while supporting the open, information-sharing needs of our academic culture
  - Link: <https://ir.princeton.edu/data/information-security-policy>

### **Columbia University (USA):**

- Has a page equivalent to fact book with information on:
  - Student admissions, enrollment, degrees and certificates, tuition and financial aid
  - Faculty and staff
  - University libraries, research, finance, leadership
  - Link: <https://www.columbia.edu/content/statistics-and-facts>

- *Managed by: Office of Planning and Institutional Research*

### **Duke University (USA):**

- Has a common data set (equivalent of fact book) with information on:
  - Enrollment information, financial aid, student head count, etc.
  - Link: <https://finance.provost.duke.edu/common-data-set>
- Has a separate page dedicated to:
  - budget and finance policies and documents, human resources, and institutional research policies and surveys
  - Link: <https://finance.provost.duke.edu/related-policies-and-key-documents#institutional-research>
- *Managed by: Office of the Provost, Finance, and Administration*

### **University of Pennsylvania (USA):**

- Has a page dedicated to GIS data
  - Link: <https://guides.library.upenn.edu/data>
  - *Managed by: UPenn library*
- Has a fact book with information on:
  - Budget and finance, student headcounts, etc.
  - <http://www.upenn.edu/ir/facts.html>
  - *Managed by Office of Institutional Research and Analysis*

### **Cornell University (USA):**

- Has a Cornell Open Data Initiative (CODI):
  - CODI is an attempt to create an open and easily accessible data repository for the Cornell community
  - First started in 2016 with the help of students from Cornell AppDev, Cornell Data Science, and Cornell University Sustainable Design
  - Current open datasets include map data (buildings, carshare, bluelights, bike racks, parking, park mobile, virtual tour, info booths), transit data (route schedule, stop schedule, stop locations), calendar data (Cornell events), dining data
  - Based on the data available, CODI seems to have a strong app development and/or economic perspective
  - CODI also appears to be a group initiative rather than a university wide program
  - Link: <https://codi.engineering.cornell.edu>
  - *Managed by: Cornell AppDev (IT department)*
- Has a DataStaR program:
  - An initiative to create a dataset registry that will support data sharing among researchers from different institutions across the nation and world
  - Link: <https://www.library.cornell.edu/datastar>
  - *Managed by: Cornell University Library*

### **Rice University (USA):**

- Part of the Common Data Set initiative (equivalent to Fact Book)
  - Link: <https://oir.rice.edu/additional-resources/common-data-set>
  - *Managed by: Office of Institutional Research*

- Has a page dedicated to sharing your research data
  - Outlines Rice’s Research Data Management Policy for information about intellectual property rights and research data
  - Link: <https://library.rice.edu/services/sharing-data>
  - *Managed by: Rice University Library*

#### **University of Chicago (USA):**

- Has a program called “Center for Translational Data Science”
  - A research center pioneering translational data science to advance science
  - Develops instruments to integrate commons of complex data with cloud computing technology
  - Link: <https://ctds.uchicago.edu/what-we-do>
- The library has data sets on GIS systems and city of Chicago
  - Link: <https://guides.lib.uchicago.edu/c.php?g=565143&p=3892726>
  - *Managed by: Chicago University library*

#### **University of Sydney (Australia):**

- Has a university library page dedicated to research data management:
  - Link: <https://library.sydney.edu.au/research/manage-data.html>
  - *Managed by: University of Sydney Library*

#### **University of Melbourne (Australia):**

- Has a research team within the university interested in open data:
  - The Open Data Initiative Team evaluates what attitudes workers within government agencies have towards open data, how prepared they are to opening their data, and what constraints or barriers prevent them from doing so
  - This is not a university level data catalogue
  - Link: <https://networkedsociety.unimelb.edu.au/research/projects/active/open-data-practice>
  - *Managed by: Networked Society Institute*

#### **Australian National University (Australia):**

- Has a page dedicated to general research management strategies
  - Link: <https://anulib.anu.edu.au/research-learn/research-data-management>
  - *Managed by: Australian National University Library*
- Has a page on what open access publishing is how to contribute to the university’s repository
  - Link: <https://anulib.anu.edu.au/research-learn/publishing-sharing/open-access>

## 5. Works Cited

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A Systematic Review of Open Government Data Initiatives. *Government Information Quarterly*, 32, 399-418. Retrieved July 01, 2019, from <https://www.sciencedirect.com/science/article/pii/S0740624X1500091X>.

- Borgman, C. (2018). Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier. *Berkeley Technology Law Journal*, 33(2), 365-412. doi: 10.15779/Z38B56D489
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems Management*, 29(4), 258-268. doi: 10.1080/10580530.2012.716740
- Keßler, C., & Kauppinen, T. (2012). Linked Open Data University of Münster – Infrastructure and Applications. doi: 10.1007/978-3-662-46641-4\_43
- Krumova, M. (2017). Higher Education 2.0 and Open Data: A Framework for University Openness and Co-creation Performance. *ICEGOV*, 562-563. Retrieved July 21, 2019, from <https://dl.acm.org/citation.cfm?id=3047316&dl=ACM&coll=DL>
- Ma Y., Xu B., Bai Y., Li Z. (2012) Building Linked Open University Data: Tsinghua University Open Data as a Showcase. In: Pan J.Z. et al. (eds) *The Semantic Web. JIST 2011. Lecture Notes in Computer Science*, vol 7185. Springer, Berlin, Heidelberg
- Open Data 101. (2017, December). Retried July 2, 2019, from <https://open.canada.ca/en/open-data-principles>
- Open Data Movement Reaches Turning Point (2016, May). Retrieved July 2, 2019, from <https://gijn.org/2016/05/13/open-data-movement-reaches-turning-point/>
- ‘Open data’: Why it’s making waves across Canada. (2016, May). Retrieved July 2, 2019, from <https://www.cbc.ca/news/technology/open-data-movement-canada-1.3557562>
- Zablith F., Fernandez M., Rowe M. (2012) The OU Linked Open Data: Production and Consumption. In: García-Castro R., Fensel D., Antoniou G. (eds) *The Semantic Web: ESWC 2011 Workshops. ESWC 2011. Lecture Notes in Computer Science*, vol 7117. Springer, Berlin, Heidelberg