Introduction to opendatatoronto

`opendatatoronto` is an R interface to the City of Toronto Open Data Portal. The goal of the package is to help read data directly into R without needing to manually download it via the portal.

In the portal, datasets are called **packages**. You can see a list of available packages by using `list_packages()`. This will show metadata about the package, including what topics (i.e. tags) the package covers, a description of it, any civic issues it addresses, how many resources there are (and their formats), how often it is is refreshed and when it was last refreshed.

```
library(opendatatoronto)

packages <- list_packages(limit = 10)

packages
#> # A tibble: 10 x 11
#>    title  id    topics civic_issues publisher excerpt dataset_category
#>    <chr>  <chr> <chr>  <chr>        <chr>     <chr>   <chr>
#>  1 Dines~ b6b4~ "c(\"~ "NULL"       Toronto ~ "Snaps~ Table
#>  2 Toron~ toro~ "Tran~ "NULL"       Parks, F~ "This ~ Table
#>  3 Commi~ 260e~ "City~ "NULL"       City Pla~ "This ~ Table
#>  4 Toron~ c6d6~ "City~ "NULL"       City Man~ "This ~ Table
#>  5 Noise~ nois~ "Busi~ "NULL"       Municipa~ "This ~ Table
#>  6 Traff~ traf~ "Tran~ "NULL"       Transpor~ "This ~ Table
#>  7 Prope~ 1aca~ "Loca~ "NULL"       Informat~ "This ~ Map
#>  8 Green~ 9a28~ "c(\"~ "NULL"       Parks, F~ "Parks~ Map
#>  9 Munic~ 5da2~ "c(\"~ "c(\"Afford~ Municipa~ "This ~ Document
#> 10 Lobby~ 6a87~ "City~ "NULL"       Lobbyist~ "The L~ Document
#> # i 4 more variables: num_resources <int>, formats <chr>,
#> #   refresh_rate <chr>, last_refreshed <date>
```

Or, you can search packages by title using `search_packages()`:

```
apartment_packages <- search_packages("Apartment")

apartment_packages
#> # A tibble: 2 x 11
#>   title   id    topics civic_issues publisher excerpt dataset_category
#>   <chr>   <chr> <chr>  <chr>        <chr>     <chr>   <chr>
#> 1 Apartm~ 4ef8~ "c(\"~ "NULL"       Municipa~ This d~ Table
#> 2 Apartm~ 2b98~ "c(\"~ "c(\"Afford~ Municipa~ This d~ Table
#> # i 4 more variables: num_resources <int>, formats <chr>,
#> #   refresh_rate <chr>, last_refreshed <date>
```

You can also see metadata for one specific package using `show_package()`:

```
show_package("996cfe8d-fb35-40ce-b569-698d51fc683b")
#> # A tibble: 4 x 11
#>   title   id    topics civic_issues publisher excerpt dataset_category
#>   <chr>   <chr> <chr>  <chr>        <chr>     <chr>   <chr>
#> 1 TTC Su~ 996c~ Trans~ <NA>         Toronto ~ TTC Su~ Document
#> 2 TTC Su~ 996c~ Trans~ <NA>         Toronto ~ TTC Su~ Document
#> 3 TTC Su~ 996c~ Trans~ <NA>         Toronto ~ TTC Su~ Document
#> 4 TTC Su~ 996c~ Trans~ <NA>         Toronto ~ TTC Su~ Document
#> # i 4 more variables: num_resources <int>, formats <chr>,
#> #   refresh_rate <chr>, last_refreshed <date>
```

Within a package, there are a number of **resources** - e.g. CSV, XSLX, JSON, SHP files, and more. Resources are the actual "data".

For a given package, you can get a list of resources using `list_package_resources()`, either by using a package found via `search_packages()` or `list_packages()`:

```
apartment_building_registration_package <- search_packages("Apartment Building Registration")

apartment_building_registration_resources <- apartment_building_registration_package %>%
  list_package_resources()

apartment_building_registration_resources
#> # A tibble: 4 x 4
#>   name                                      id    format last_modified
#>   <chr>                                     <chr> <chr>  <date>
#> 1 Apartment Building Registration Data      3ad7~ CSV    2025-03-05
#> 2 Apartment Building Registration Data.csv  97b8~ CSV    2025-03-05
#> 3 Apartment Building Registration Data.xml  b1b6~ XML    2025-03-05
#> 4 Apartment Building Registration Data.json 005b~ JSON   2025-03-05
```

or by passing the package's portal URL directly:

```
list_package_resources("https://open.toronto.ca/dataset/apartment-building-registration/")
#> # A tibble: 4 x 4
#>   name                                      id    format last_modified
#>   <chr>                                     <chr> <chr>  <date>
#> 1 Apartment Building Registration Data      3ad7~ CSV    2025-03-05
#> 2 Apartment Building Registration Data.csv  97b8~ CSV    2025-03-05
#> 3 Apartment Building Registration Data.xml  b1b6~ XML    2025-03-05
#> 4 Apartment Building Registration Data.json 005b~ JSON   2025-03-05
```

Finally (and most usefully!), you can download the resource (i.e., the actual data) directly into R using `get_resource()`:

```
library(dplyr)

apartment_building_registration_data <- apartment_building_registration_resources %>%
  filter(name == "Apartment Building Registration Data") %>%
  get_resource()

apartment_building_registration_data
#> # A tibble: 3,597 x 70
#>     `_id` AIR_CONDITIONING_TYPE AMENITIES_AVAILABLE
#>     <int> <chr>                 <chr>
#>  1 134649 NONE                  <NA>
#>  2 134650 NONE                  <NA>
#>  3 134651 NONE                  <NA>
#>  4 134652 NONE                  <NA>
#>  5 134653 NONE                  <NA>
#>  6 134654 INDIVIDUAL UNITS      <NA>
#>  7 134655 NONE                  <NA>
#>  8 134656 NONE                  <NA>
#>  9 134657 NONE                  <NA>
#> 10 134658 INDIVIDUAL UNITS      Indoor recreation room
#> # i 3,587 more rows
#> # i 67 more variables: ANNUAL_FIRE_ALARM_TEST_RECORDS <chr>,
#> #   ANNUAL_FIRE_PUMP_FLOW_TEST_RECORDS <chr>,
#> #   APPROVED_FIRE_SAFETY_PLAN <chr>, BALCONIES <chr>,
#> #   BARRIER_FREE_ACCESSIBILTY_ENTR <chr>, BIKE_PARKING <chr>,
#> #   CONFIRMED_STOREYS <int>, CONFIRMED_UNITS <int>,
#> #   DATE_OF_LAST_INSPECTION_BY_TSSA <chr>, ...
```

The `opendatatoronto` package can currently handle the download of CSV, XLS/XLSX, XML, JSON, SHP, and GeoJSON resources, as well as ZIP resources that contain multiple files. For more information, see the following vignettes:

- Retrieving multi-sheet XLS/XLSX resources
- Retrieving multi-file ZIP resources
- Working with spatial data from the portal