Power Query (M): Get public data via the *Comprehensive Knowledge Archive Network (CKAN)*

Owen price
linkedin.com/in/owenhprice

The Comprehensive Knowledge Archive Network (CKAN) is an *open-source open data portal* for storage and distribution of open data

Owen price
linkedin.com/in/owenhprice

Many governmental and public entities use CKAN to share data. Wikipedia shows some...

```
let
    tablechild = (child) => "TABLE.wikitable:nth-child(13) > * > TR > :nth-child(" & Text.From(child) & ")",
    Source = Web.BrowserContents("https://en.wikipedia.org/wiki/List_of_open_government_data_sites"),
    extract_table = Html.Table(
        Source,
        {{"Government", tablechild(1)}, {"Website", tablechild(2)}, {"Portal Solution", tablechild(3)}},
        [RowSelector = "TABLE.wikitable:nth-child(13) > * > TR"]
    ),
    ckan_only = Table.SelectRows(extract_table, each ([Portal Solution] = "CKAN"))
in
    ckan_only
```

	Government	Website	Portal Solution
1	Argentina	https://datos.gob.ar/	CKAN
2	Australia	https://data.gov.au/	CKAN
3	Denmark	https://www.opendata.dk/	CKAN
4	Italy	https://www.dati.gov.it/	CKAN
5	Japan	https://www.data.go.jp/	CKAN
6	Latvia	https://data.gov.lv/eng	CKAN
7	Paraguay	http://www.gobiernoabierto.gov.py/ http://datos.gov.py/	CKAN
8	United States	https://data.gov/	CKAN
9	Uruguay	https://catalogodatos.gub.uy/	CKAN

Owen price
linkedin.com/in/owenhprice

CKAN defines “packages” of data. Available packages can be viewed using the “package_list” action. For example:

https://openfinance.houstontx.gov/api/3/action/package_list


```
{
    "help": "https://openfinance.houstontx.gov/api/3/action/help_show?name=package_list",
    "success": true,
    "result": [
        "budget-2018","budget-2019","budget-2020",
        "budget-2021","budget-2022","budget-2023",
        "checkbook-2018","checkbook-2019","checkbook-2020",
        "checkbook-2021","checkbook-2022","checkbook-2023",
        "payroll-2018","payroll-2019","payroll-2020",
        "payroll-2021","payroll-2022","payroll-2023"
    ]
}
```


Owen price
linkedin.com/in/owenhprice

We can view the metadata for a package using the “package_show” action. For example:

https://openfinance.houstontx.gov/api/3/action/package_show?id=budget-2018


```
{
    "help": "https://openfinance.houstontx.gov/api/3/action/help_show?name=package_show",
    "success": true,
    "result": {
        "author": "City of Houston",
        "author_email": null,
        "creator_user_id": "be3daaa7-6c94-493a-912a-2a1c8eff2cd6",
        "id": "978d4a9e-f7b7-4d06-b3ee-e1980c8815c4",
        "isopen": true,
        "license_id": "odc-by",
        "license_title": "Open Data Commons Attribution License",
        "license_url": "http://www.opendefinition.org/licenses/odc-by",
        "maintainer": "City of Houston",
        "maintainer_email": null,
        "metadata_created": "2022-09-08T17:08:09.706057",
        "metadata_modified": "2022-09-08T17:08:13.610560",
        "name": "budget-2018",
        … etc
```


Owen price
linkedin.com/in/owenhprice

A package definition includes a list of resources. Each resource has a resource_id.

https://openfinance.houstontx.gov/api/3/action/package_show?id=budget-2018


```
"resources": [
    {
        "cache_last_updated": null,
        "cache_url": null,
        "ckan_url": "http://10.98.72.6",
        "created": "2022-09-08T17:08:10.021338",
… etc
        "name": "Budget 2018 Data",
        "original_url": "https://safinckanprodusgovtx.blob.core.usgovcloudapi.net/openfinance/budget-data/Budget-2018.csv?63151",
        "package_id": "978d4a9e-f7b7-4d06-b3ee-e1980c8815c4",
        "position": 0,
        "resource_id": "bf2ee126-e909-46d8-995e-049caa03c537",
… etc
        "url_type": null
    }
],
```


Owen price
linkedin.com/in/owenhprice

CKAN has a PostgreSQL back-end.

Each resource has a table named after the resource ID. We can query it:

```
SELECT *
FROM "0b18055d-9696-4dbb-b045-20d1f9ad5a9a"
LIMIT 1
```

resource_id for “budget-2018” package

https://openfinance.houstontx.gov/api/3/action/datastore_search_sql?sql=SELECT%20*%20FROM%20%220b18055d-9696-4dbb-b045-20d1f9ad5a9a%22%20LIMIT%201


```
{
…
    "success": true,
    "result": {
        "sql": "SELECT * FROM \"0b18055d-9696-4dbb-b045-20d1f9ad5a9a\" LIMIT 1",
        "records": [
            {
                "Fiscal Year": "2023",
                "Fund ID": "1000",
                "Fund Name": "General Fund",
                "Fund Type": "General Funds",
                "Department ID": "1600",
                "Department Name": "Municipal Courts Department",
                "Department Type": "Public Safety",
                "GL Account": "520108",
                "GL Description": "Information Resource Services"
```


Owen price
linkedin.com/in/owenhprice

Because of all this, Power Query can be used to extract all resources at once

Packages

Package metadata

API result against the resource_id

The columns returned from the resource

The data returned from the resource

= extract_query_records

	package_name	package_definition	resource_id	data	query_fields	query_records
1	budget-2018	Record	bf2ee126-e909-46d8-995e-049caa03c537	Record	List	Table
2	budget-2019	Record	7675d85e-b51f-4ea2-98dc-825ae417bddd	Record	List	Table
3	budget-2020	Record	57c276b6-7eae-40f9-ab52-a5d7081f2fe2	Record	List	Table
4	budget-2021	Record	680c17de-d656-46d0-9c93-1dc2a9167f1b	Record	List	Table
5	budget-2022	Record	2df4eb1c-5ffc-46b1-8d34-bd9d9ff9db35	Record	List	Table
6	budget-2023	Record	0b18055d-9696-4dbb-b045-20d1f9ad5a9a	Record	List	Table
7	checkbook-2018	Record	8f368b3c-a2fd-4d1e-a359-770623020bba	Record	List	Table
8	checkbook-2019	Record	7e73588d-f9e4-4463-a382-97cf79d5a70d	Record	List	Table
9	checkbook-2020	Record	835823db-e4d8-44b9-9931-f5f7d02bd68f	Record	List	Table
10	checkbook-2021	Record	0e83e31d-61d5-44ca-af4a-b267e4dd8fc5	Record	List	Table
11	checkbook-2022	Record	1c8eaf09-9828-4150-baf5-5ed619070e72	Record	List	Table
12	checkbook-2023	Record	84ad3510-149f-432c-aa28-75a11909ade5	Record	List	Table
13	payroll-2018	Record	73ec3be1-5eef-479d-bdce-2d9619fca381	Record	List	Table
14	payroll-2019	Record	522da494-62ad-4520-93b8-4d64ff773283	Record	List	Table
15	payroll-2020	Record	75177de4-7ce3-4cb7-9da3-08da50eb6b9a	Record	List	Table
16	payroll-2021	Record	b2af1219-de9e-44fb-b968-0fdd304b3893	Record	List	Table
17	payroll-2022	Record	a8d64655-5397-4ec5-9717-84430a00a27e	Record	List	Table
18	payroll-2023	Record	386bf9d8-70cc-4079-93af-16af354a7189	Record	List	Table

_id	Fiscal Year	Fund ID	Fund Name	Fund Type	Department ID	Department Name	Department Type	GL Account	GL Description	GL Category
1	2018	1000	General Fund	General Funds	2500	General Services Department	Development & Maintenance Services	500030	Salary Part Time - Civilian	Personnel Service
2	2018	1000	General Fund	General Funds	7000	Planning & Development	Development & Maintenance Services	511110	Fuel	Supplies
3	2018	1011	WorkersCompensation	Service Chargeback Funds	8000	Human Resources Dept.	Administrative Services	511040	Audiovisual Supplies	Supplies
4	2018	2002	Health Special	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services	456255	Miscellaneous Operating Revenue	Miscellaneous/Ot
5	2018	2002	Health Special	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services	511120	Clothing	Supplies
6	2018	2009	Swimming Pool Safety	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services	504030	Unemployment Claims - Administration	Personnel Service
7	2018	2009	Swimming Pool Safety	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services	511080	General Laboratory Supplies	Supplies

Owen price
linkedin.com/in/owenhprice

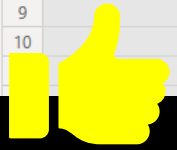
And we can extract each resource type to a separate table

```
// GetPackage
(package as text) as table =>
let
    Source = Table.SelectRows(
        GetCKANResources,
        each Text.StartsWith([package_name],package))[query_records],
    result = Table.Combine(Source)
in
    result
```

Query shown on previous page

= GetPackage("budget")

	_id	Fiscal Year	Fund ID	Fund Name	Fund Type	Department ID	Department Name	Department Type
1	1	2018	1000	General Fund	General Funds	2500	General Services Department	Development & Maintenance Services
2	2	2018	1000	General Fund	General Funds	7000	Planning & Development	Development & Maintenance Services
3	3	2018	1011	WorkersCompensation	Service Chargeback Funds	8000	Human Resources Dept.	Administrative Services
4	4	2018	2002	Health Special	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services
5	5	2018	2002	Health Special	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services
6	6	2018	2009	Swimming Pool Safety	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services
7	7	2018	2009	Swimming Pool Safety	Special Revenue Funds	3800	Houston Health Department	Human & Cultural Services
8	8	2018	2301	Building Inspection	Special Revenue Funds	2000	Houston Public Works - HPW	Development & Maintenance Services
9	9	2018	2301	Building Inspection	Special Revenue Funds	2000	Houston Public Works - HPW	Development & Maintenance Services
10	10	2018	2301	Building Inspection	Special Revenue Funds	2000	Houston Public Works - HPW	Development & Maintenance Services
11	11	2018	2301	Building Inspection	Special Revenue Funds	2000	Houston Public Works - HPW	Development & Maintenance Services



Owen price
linkedin.com/in/owenhprice

TAKEAWAYS

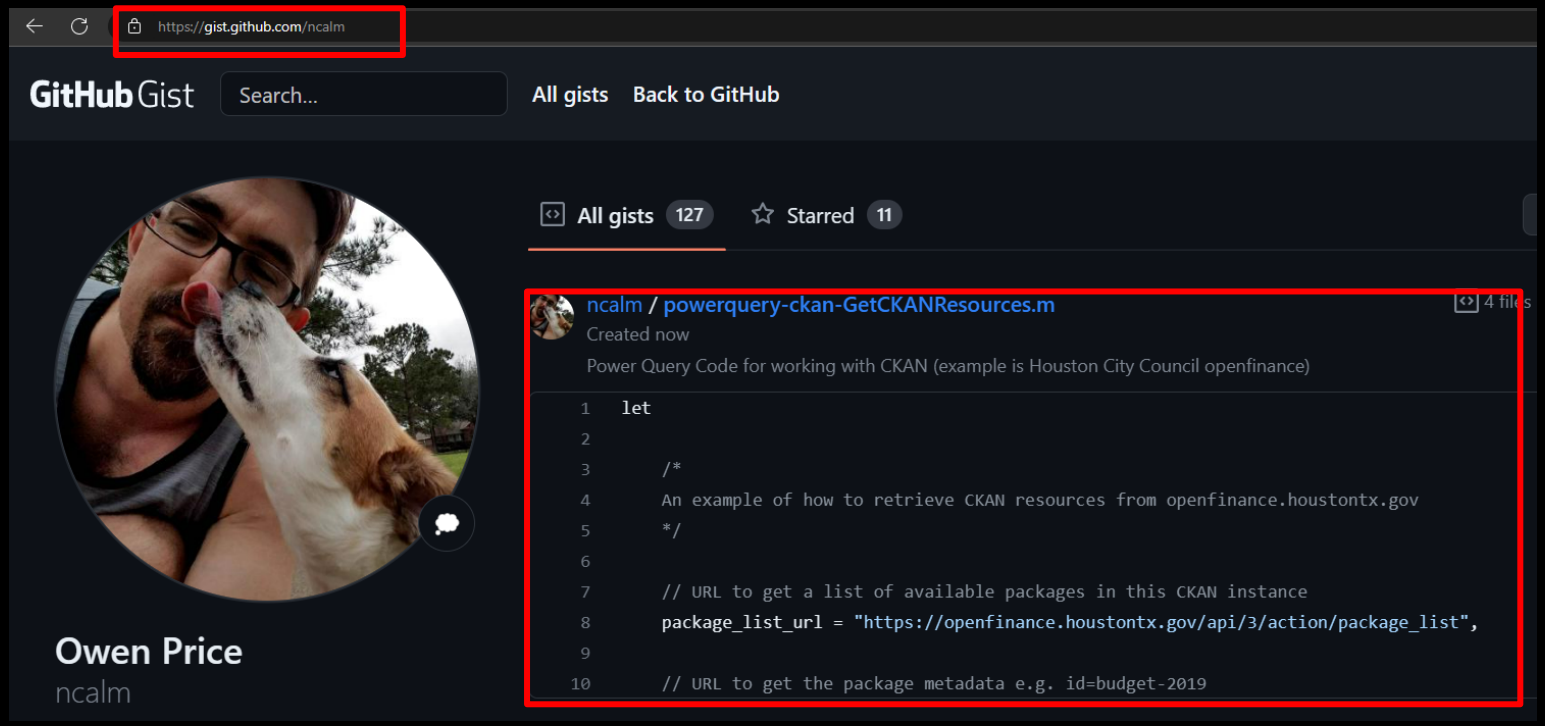
1. *The Comprehensive Knowledge Archive Network is used by many entities to share public data*

2. *They all use a similar API structure with standardized retrieval functions*

3. *Power Query can help us retrieve several related resources at once*

Owen price
linkedin.com/in/owenhprice

Grab the code



Owen price
linkedin.com/in/owenhprice