

✓ Week 2 - Manipulating Structured Data

(Be sure to Copy to Drive first)

This will build on content introduced in week 1. In week 1 we learned that information can be presented on the web in HTML (ie, with formatting and graphics) or as data that can be read programmatically via Python through an *API* (Application Program Interface).

This week we will build on what we learned by creating a dataset from a few different APIs.

With the library installed we will now query the API to look at the same article we saw above.

```
1 from crossref.restful import Works
2 works = Works()
3 item = works.doi("10.5334/johd.109")
```

Run the next cell to see how the information looks a lot like JSON.

```
1 item
```

```
⇒ {'indexed': {'date-parts': [[2025, 5, 14]],
  'date-time': '2025-05-14T13:20:25Z',
  'timestamp': 1747228825714,
  'version': '3.40.5'},
  'reference-count': 8,
  'publisher': 'Ubiquity Press, Ltd.',
  'content-domain': {'domain': [], 'crossmark-restriction': False},
  'short-container-title': [],
  'published-print': {'date-parts': [[2023, 6, 29]]},
  'DOI': '10.5334/johd.109',
  'type': 'journal-article',
  'created': {'date-parts': [[2023, 6, 29]],
  'date-time': '2023-06-29T10:38:10Z',
  'timestamp': 1688035090000},
  'source': 'Crossref',
  'is-referenced-by-count': 0,
  'title': ['A Mixed Method Twitter Methodology and Anonymous Corpus'],
  'prefix': '10.5334',
  'volume': '9',
  'author': [{'ORCID': 'https://orcid.org/0000-0001-9229-8569',
  'authenticated-orcid': False,
  'given': 'Tim',
  'family': 'Ribaric',
  'sequence': 'first',
  'affiliation': []}],
  'member': '3285',
  'reference': [{'volume': 9, 'title': 'Libraries as dysfunctional organizations and
```

```

reference : 11 volume-title : 'Libraries as dysfunctional organizations and
workplaces',
  'year': '2022',
  'key': 'key20230629063807_B1'},
{'key': 'key20230629063807_B2',
  'unstructured': 'Ettarh, F. (2018). Vocational Awe and Librarianship: The
Lies We Tell Ourselves – In the Library with the Lead Pipe. Retrieved from
https://www.inthelibrarywiththeleadpipe.org/2018/vocational-awe/ (last
accessed: April 28, 2023).'},
{'year': '2014',
  'key': 'key20230629063807_B3',
  'article-title': 'VADER: A Parsimonious Rule-Based Model for Sentiment
Analysis of Social Media Text [Full Paper]'},
{'key': 'key20230629063807_B4',
  'unstructured': '@lis_grievances bot. (n.d.). LIS Grievances–About.
Retrieved from https://lisgrievances.com/about.html (last accessed: April 17,
2023).'},
{'key': 'key20230629063807_B5',
  'unstructured': 'Ribaric, T. (2022a). LIS Grievances. Retrieved from
https://github.com/elibtronic/lis\_grievances (last accessed: April 28,
2023).'},
{'key': 'key20230629063807_B6',
  'first-page': '167',
  'volume-title': 'Libraries as Dysfunctional Organizations and Workplaces',
  'year': '2022',
  'edition': '1st ed.'},
{'key': 'key20230629063807_B7',
  'unstructured': 'Roesslein, J. (2020). Tweepy: Twitter for Python.
Retrieved from https://Github.Com/Tweepy/Tweepy (last accessed: April 28,
2023).'},
{'key': 'key20230629063807_B8',
  'first-page': '1',

```

✓ Parts of the data

By doing a key lookup, we can find the associated value

```
1 item['title']
```

```
➤ ['A Mixed Method Twitter Methodology and Anonymous Corpus']
```

Q1 Print the ISSN that the article is from

```
1 #Q1
2 item['ISSN']
```

```
➤ ['2059-481X']
```

Q2 Print the number of references this article has

```

1 #Q2
2 item['references-count']

8

```

Q3 Print the volume that this article is from

```

1 #Q3
2 item['volume']

'9'

```

✓ Two levels deep

For example we can find the family name of the author with something like the following

```

1 #We are interested in the first item in this list
2 #which is why we index at 0
3 #if this piece had more than 1 author the list above
4 #would be bigger
5 author_info = item['author'][0]
6 author_info['family']
7

'Ribaric'

```

Q4 Print the ORCID for this author

```

1 author_info = item['author'][0]
2 author_info['ORCID']

'https://orcid.org/0000-0001-9229-8569'

```

Q5 Print the list of references this article has. You just need to modify line 1.

```

1 references = item['reference']
2
3 for reference in references:
4     #Even more JSON objects!
5     print(reference)

{'volume-title': 'Libraries as dysfunctional organizations and workplaces', 'v

```

{ 'volume-title': 'Libraries as dysfunctional organizations and workplaces',
{ 'key': 'key20230629063807_B2', 'unstructured': 'Ettarh, F. (2018). Vocational
{ 'year': '2014', 'key': 'key20230629063807_B3', 'article-title': 'VADER: A Pal
{ 'key': 'key20230629063807_B4', 'unstructured': '@lis_grievances bot. (n.d.).
{ 'key': 'key20230629063807_B5', 'unstructured': 'Ribaric, T. (2022a). LIS Gri
{ 'key': 'key20230629063807_B6', 'first-page': '167', 'volume-title': 'Librari
{ 'key': 'key20230629063807_B7', 'unstructured': 'Roesslein, J. (2020). Tweepy
{ 'key': 'key20230629063807_B8', 'first-page': '1', 'article-title': 'New Libra