

# Get DOI metadata

2023-10-10

## DOI API access

I was interested in getting the metadata for a DOI without using a package in Python or R. My goal was to get a better understanding of the APIs that are available to access DOI metadata.

Here is an example with a request for my first peer-reviewed publication.

```
import requests

doi_url = "https://doi.org/10.1177/1465116507076430"

headers = {"accept": "text/x-bibliography"}
r = requests.get(doi_url, headers=headers)

r.text
```

'Döring, H. (2007). The Composition of the College of Commissioners. European Union Politics

## Content types

This is the API information I was interested in.

<https://citation.crosscite.org/docs.html#sec-4>

“Currently three DOI registration agencies have implemented content negotiation for their DOIs: Crossref, DataCite and mEDRA. They support a number of meta-data content types, some of which are common to the three RAs.”

I try to avoid xml-based formats, so that the following content types are of interest to me.

- Formatted text citation // `text/x-bibliography`
- BibTeX // `application/x-bibtex`
- Citeproc JSON // `application/vnd.citationstyles.csl+json`
- Schema.org in JSON-LD // `application/vnd.schemaorg.ld+json` (only DataCite)

It is actually nice to put a request into a function. I start with a simple example.

```
def get_doi_bib(doi):
    """Return a bibliography for a given DOI."""

    url = f"https://dx.doi.org/{doi}"

    headers = {"accept": "text/x-bibliography"}
    r = requests.get(url, headers=headers)

    return r.text

get_doi_bib("10.1177/0958928715573481")
```

'Döring, H., & Schwander, H. (2015). Revisiting the left cabinet share: How to measure the p

To get the different formats I am interested in, I can rewrite the function.

```
import time
import requests

CONTENT_TYPES = {
    "bibliography": "text/x-bibliography",
    "bibtex": "application/x-bibtex",
    "json-csl": "application/vnd.citationstyles.csl+json",
    "json-ld": "application/vnd.schemaorg.ld+json",
}

def get_doi_data(doi, content_type="bibliography"):
    """Return metadata for DOI."""

    url = f"https://dx.doi.org/{doi}"
    headers = {"accept": CONTENT_TYPES.get(content_type, "bibliography")}

    return requests.get(url, headers=headers)
```

```
doi = "10.1016/j.electstud.2016.07.002"

for key in CONTENT_TYPES:
    time.sleep(2) # pause to avoid requests limit
    doi_request = get_doi_data(doi, key)
    print(f"DOI content type: {key}\n{doi_request.text}\n")
```

DOI content type: bibliography

Döring, H. (2016). Mapping established democracies: Integrated data on parties, elections and cabinets

DOI content type: bibtex

```
@article{Doring_2016,
  doi = {10.1016/j.electstud.2016.07.002},
  url = {https://doi.org/10.1016%2Fj.electstud.2016.07.002},
  year = 2016,
  month = {dec},
  publisher = {Elsevier {BV}},
  volume = {44},
  pages = {535--543},
  author = {Holger Döring},
  title = {Mapping established democracies: Integrated data on parties, elections and cabinets},
  journal = {Electoral Studies}
}
```

DOI content type: json-csl

```
{"indexed":{"date-parts":[[2022,4,3]],"date-time":"2022-04-03T16:10:26Z","timestamp":1649002200}}
```

DOI content type: json-ld

No acceptable resource available.

## Linked data json-ld

*json-ld* is only available for content registered with [DataCite](#).

A DOI for an article does not return *json-ld* metadata.

```
doi_content = get_doi_data(doi, "json-ld")
```

```
doi_content
```

```
<Response [406]>
```

```
doi_content.text
```

```
'No acceptable resource available.'
```

---

A DOI for a dataset returns *json-ld* metadata.

```
doi_content = get_doi_data("10.7910/DVN/BUWAAF", "json-ld")
```

```
doi_content
```

```
<Response [200]>
```

```
doi_content.text
```

```
'{\n  "@context": "http://schema.org",\n  "@type": "Dataset",\n  "@id": "https://doi.org/10.7910/DVN/BUWAAF",\n  "version": "Version 2022",\n  "author": [\n    {\n      "name": "Paul Bederke",\n      "givenName": "Paul",\n      "familyName": "Bederke",\n      "affiliation": {\n        "@type": "Organization",\n        "name": "(University of Konstanz)"\n      }\n    }\n  ],\n  "url": "https://dataverse.harvard.edu/citation?persistentId=doi:10.7910/DVN/BUWAAF",\n  "name": "Party Facts - Version 2022",\n  "description": "A dataset of party facts from the University of Konstanz, Germany, covering the years 2010 to 2022. The dataset is part of the 'Party Facts' project, which aims to provide a comprehensive overview of party activities and structures in Germany. The data is sourced from various public and private databases, including the German Bundestag, state parliaments, and academic research. The dataset is structured as a JSON-LD document, allowing for easy integration and analysis. The version number 2022 indicates that the data is up-to-date and reflects the latest available information. The dataset is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike license, which allows for reuse and distribution of the data, provided that the original source is properly acknowledged and that any derivative works are also licensed under the same terms. The dataset is available for download and use by anyone interested in party politics and research. The data is presented in a clear and concise format, making it easy to understand and use. The dataset is a valuable resource for researchers, policymakers, and anyone interested in the political system of Germany. The dataset is a testament to the commitment of the University of Konstanz to open science and data sharing. The dataset is a valuable resource for anyone interested in party politics and research. The dataset is a testament to the commitment of the University of Konstanz to open science and data sharing.'
```

```
doi_content.json()
```

```
{ '@context': 'http://schema.org',\n  '@type': 'Dataset',\n  '@id': 'https://doi.org/10.7910/dvn/buwaaf',\n  'url': 'https://dataverse.harvard.edu/citation?persistentId=doi:10.7910/DVN/BUWAAF',\n  'name': 'Party Facts - Version 2022',\n  'author': [{ 'name': 'Paul Bederke',\n                'givenName': 'Paul',\n                'familyName': 'Bederke',\n                'affiliation': { '@type': 'Organization',\n                                 'name': '(University of Konstanz)' },\n              }],\n  'description': 'A dataset of party facts from the University of Konstanz, Germany, covering the years 2010 to 2022. The dataset is part of the 'Party Facts' project, which aims to provide a comprehensive overview of party activities and structures in Germany. The data is sourced from various public and private databases, including the German Bundestag, state parliaments, and academic research. The dataset is structured as a JSON-LD document, allowing for easy integration and analysis. The version number 2022 indicates that the data is up-to-date and reflects the latest available information. The dataset is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike license, which allows for reuse and distribution of the data, provided that the original source is properly acknowledged and that any derivative works are also licensed under the same terms. The dataset is available for download and use by anyone interested in party politics and research. The data is presented in a clear and concise format, making it easy to understand and use. The dataset is a valuable resource for researchers, policymakers, and anyone interested in the political system of Germany. The dataset is a testament to the commitment of the University of Konstanz to open science and data sharing.'
```

```

    '@type': 'Person',
    '@id': 'https://orcid.org/0000-0001-7555-8656'},
{'name': 'Holger Döring',
 'givenName': 'Holger',
 'familyName': 'Döring',
 'affiliation': {'@type': 'Organization',
 'name': '(GESIS - Leibniz Institute for the Social Sciences)'},
 '@type': 'Person',
 '@id': 'https://orcid.org/0000-0002-6616-8805'},
{'name': 'Sven Regel',
 'givenName': 'Sven',
 'familyName': 'Regel',
 'affiliation': {'@type': 'Organization',
 'name': '(WZB Berlin Social Science Center)'},
 '@type': 'Person'}],
'editor': {'name': 'Paul Bederke',
 'givenName': 'Paul',
 'familyName': 'Bederke',
 'affiliation': {'@type': 'Organization', 'name': '(University of Konstanz)'},
 'contributorType': 'ContactPerson',
 '@type': 'Person'},
'description': 'Party Facts links datasets on political parties',
'datePublished': '2022',
'schemaVersion': 'http://datacite.org/schema/kernel-4',
'publisher': {'@type': 'Organization', 'name': 'Harvard Dataverse'},
'provider': {'@type': 'Organization', 'name': 'datacite'}}

```

## Redirects

Finally, I was interested if there is a redirect to a DOI registration agency.

<https://citation.crosscite.org/docs.html#sec-2>

“Content negotiated requests to doi.org that ask for a content type which isn’t”text/html” will be redirected to a metadata service hosted by the DOI’s registration agency. Crossref, DataCite and mEDRA support content negotiated DOIs via their respective resolvers.”

However, it seems that the content is returned by *doi.org*.

```

for request_history in doi_content.history:
    print(request_history.status_code, request_history.url)

```

302 <https://dx.doi.org/10.7910/DVN/BUWAAF>

A request to **http://doi.org** is redirected to **https://doi.org**, but there is no further redirect to a DOI registration agency.

```
doi_url = f"http://doi.org/{doi}"
headers = {"accept": "text/x-bibliography"}

r = requests.get(doi_url, headers=headers)

for request_history in r.history:
    print(request_history.status_code, request_history.url)
```

301 <http://doi.org/10.1016/j.electstud.2016.07.002>  
302 <https://doi.org/10.1016/j.electstud.2016.07.002>

---

