Metadatenabzug per DOI

CrosRef public API: https://api.crossref.org/works/

In [3]:
```
#Bibliotheken importieren:
import urllib.request
import json
```

In [6]:
```
#URL der API einlesen und einer Variablen zuweisen:
full_url = "https://api.crossref.org/works/10.1371/journal.pcbi.1004668"
```

In [7]:
```
#Objekt erstellen und lesbar machen:
doi_json_data = urllib.request.urlopen(full_url).read()
```

In [8]:
```
#Objekt anzeigen:
print(doi_json_data)
```


```
b'{"status":"ok","message-type":"work","message-version":"1.0.0","messag
e":{"indexed":{"date-parts":[[2019,11,9]],"date-time":"2019-11-09T08:23:2
0Z","timestamp":1573287800480},"reference-count":6,"publisher":"Public Li
brary of Science (PLoS)","issue":"1","license":[{"URL":"http:\\/\\/creati
vecommons.org\\/licenses\\/by\\/4.0\\/","start":{"date-parts":[[2016,1,1
9]],"date-time":"2016-01-19T00:00:00Z","timestamp":1453161600000},"delay-
in-days":0,"content-version":"vor"}],"content-domain":{"domain":["www.plo
scompbiol.org"],"crossmark-restriction":false},"short-container-title":
["PLoS Comput Biol"],"DOI":"10.1371\\/journal.pcbi.1004668","type":"journ
al-article","created":{"date-parts":[[2016,1,19]],"date-time":"2016-01-19
T21:35:29Z","timestamp":1453239329000},"page":"e1004668","update-polic
y":"http:\\/\\/dx.doi.org\\/10.1371\\/journal.pcbi.corrections_policy","s
ource":"Crossref","is-referenced-by-count":39,"title":["A Quick Introduct
ion to Version Control with Git and GitHub"],"prefix":"10.1371","volum
e":"12","author":[{"given":"John D.","family":"Blischak","sequence":"firs
t","affiliation":[]},{"given":"Emily R.","family":"Davenport","sequenc
e":"additional","affiliation":[]},{"given":"Greg","family":"Wilson","sequ
ence":"additional","affiliation":[]}],"member":"340","published-online":
{"date-parts":[[2016,1,19]]},"reference":[{"key":"ref1","doi-asserted-b
y":"crossref","first-page":"7","DOI":"10.1186\\/1751-0473-8-7","article-t
itle":"Git can facilitate greater reproducibility and increased transpare
ncy in science","volume":"8","author":"K Ram","year":"2013","journal-titl
e":"Source Code Biol Med"},{"key":"ref2","doi-asserted-by":"crossref","fi
rst-page":"e1001745","DOI":"10.1371\\/journal.pbio.1001745","article-titl
e":"Best practices for scientific computing","volume":"12","author":"G Wi
lson","year":"2014","journal-title":"PLoS Biol"},{"key":"ref3","doi-asser
ted-by":"crossref","first-page":"57","DOI":"10.1038\\/nature11247","artic
le-title":"An integrated encyclopedia of DNA elements in the human genom
e","volume":"489","author":"B Bernstein","year":"2012","journal-title":"N
ature"},{"key":"ref4","doi-asserted-by":"crossref","first-page":"e100332
6","DOI":"10.1371\\/journal.pcbi.1003326","article-title":"Practical guid
elines for the comprehensive analysis of ChIP-seq data","volume":"9","aut
hor":"T Bailey","year":"2013","journal-title":"PLoS Comput Biol"},{"ke
y":"ref5","doi-asserted-by":"crossref","first-page":"e1000424","DOI":"10.
1371\\/journal.pcbi.1000424","article-title":"A quick guide to organizing
computational biology projects","volume":"5","author":"W Noble","year":"2
009","journal-title":"PLoS Comput Biol"},{"key":"ref6","doi-asserted-b
y":"crossref","first-page":"e1002598","DOI":"10.1371\\/journal.pcbi.10025
98","article-title":"A quick guide to software licensing for the scientis
t-programmer","volume":"8","author":"A Morin","year":"2012","journal-titl
e":"PLoS Comput Biol"}],"container-title":["PLOS Computational Biolog
y"],"original-title":[],"language":"en","link":[{"URL":"http:\\/\\/dx.plo
s.org\\/10.1371\\/journal.pcbi.1004668","content-type":"unspecified","con
tent-version":"vor","intended-application":"similarity-checking"}],"depos
ited":{"date-parts":[[2017,6,24]],"date-time":"2017-06-24T04:04:33Z","tim
estamp":1498277073000},"score":1.0,"subtitle":[],"editor":[{"given":"Fran
cis","family":"Ouellette","sequence":"first","affiliation":[]}],"short-ti
tle":[],"issued":{"date-parts":[[2016,1,19]]},"references-count":6,"journ
al-issue":{"published-online":{"date-parts":[[2016,1,19]]},"issu
e":"1"},"URL":"http:\\/\\/dx.doi.org\\/10.1371\\/journal.pcbi.1004668","r
elation":{"cites":[]},"ISSN":["1553-7358"],"issn-type":[{"value":"1553-73
58","type":"electronic"}]}}'
```


In [9]:
```
#Dictionary erstellen:
#(loads = load string)
doi_data = json.loads(doi_json_data)
```

In [10]:
```
#Typ von doi_data bestimmen:
type(doi_data)
```

Out[10]:
```
dict
```

In [11]:
```
#Schlüssel des Dictionaries anzeigen lassen:
doi_data.keys()
```

Out[11]:
```
dict_keys(['status', 'message-type', 'message-version', 'message'])
```

In [12]:
```
#Dictionary anzeigen lassen mit Einrückung (hier = 2 Leerzeichen):
print(json.dumps(doi_data, indent = 2))
```

```
{
  "status": "ok",
  "message-type": "work",
  "message-version": "1.0.0",
  "message": {
    "indexed": {
      "date-parts": [
        [
          2019,
          11,
          9
        ]
      ],
      "date-time": "2019-11-09T08:23:20Z",
      "timestamp": 1573287800480
    },
    "reference-count": 6,
    "publisher": "Public Library of Science (PLoS)",
    "issue": "1",
    "license": [
      {
        "URL": "http://creativecommons.org/licenses/by/4.0/",
        "start": {
          "date-parts": [
            [
              2016,
              1,
              19
            ]
          ],
          "date-time": "2016-01-19T00:00:00Z",
          "timestamp": 1453161600000
        },
        "delay-in-days": 0,
        "content-version": "vor"
      }
    ],
    "content-domain": {
      "domain": [
        "www.ploscompbiol.org"
      ],
      "crossmark-restriction": false
    },
    "short-container-title": [
      "PLoS Comput Biol"
    ],
    "DOI": "10.1371/journal.pcbi.1004668",
    "type": "journal-article",
    "created": {
      "date-parts": [
        [
          2016,
          1,
          19
        ]
      ],
      "date-time": "2016-01-19T21:35:29Z",
      "timestamp": 1453239329000
    },
    "page": "e1004668",
    "update-policy": "http://dx.doi.org/10.1371/journal.pcbi.corrections_
```

```
policy",
    "source": "Crossref",
    "is-referenced-by-count": 39,
    "title": [
      "A Quick Introduction to Version Control with Git and GitHub"
    ],
    "prefix": "10.1371",
    "volume": "12",
    "author": [
      {
        "given": "John D.",
        "family": "Blischak",
        "sequence": "first",
        "affiliation": []
      },
      {
        "given": "Emily R.",
        "family": "Davenport",
        "sequence": "additional",
        "affiliation": []
      },
      {
        "given": "Greg",
        "family": "Wilson",
        "sequence": "additional",
        "affiliation": []
      }
    ],
    "member": "340",
    "published-online": {
      "date-parts": [
        [
          2016,
          1,
          19
        ]
      ]
    },
    "reference": [
      {
        "key": "ref1",
        "doi-asserted-by": "crossref",
        "first-page": "7",
        "DOI": "10.1186/1751-0473-8-7",
        "article-title": "Git can facilitate greater reproducibility and
increased transparency in science",
        "volume": "8",
        "author": "K Ram",
        "year": "2013",
        "journal-title": "Source Code Biol Med"
      },
      {
        "key": "ref2",
        "doi-asserted-by": "crossref",
        "first-page": "e1001745",
        "DOI": "10.1371/journal.pbio.1001745",
        "article-title": "Best practices for scientific computing",
        "volume": "12",
        "author": "G Wilson",
        "year": "2014",
        "journal-title": "PLoS Biol"
```

```
      },
      {
        "key": "ref3",
        "doi-asserted-by": "crossref",
        "first-page": "57",
        "DOI": "10.1038/nature11247",
        "article-title": "An integrated encyclopedia of DNA elements in t
he human genome",
        "volume": "489",
        "author": "B Bernstein",
        "year": "2012",
        "journal-title": "Nature"
      },
      {
        "key": "ref4",
        "doi-asserted-by": "crossref",
        "first-page": "e1003326",
        "DOI": "10.1371/journal.pcbi.1003326",
        "article-title": "Practical guidelines for the comprehensive anal
ysis of ChIP-seq data",
        "volume": "9",
        "author": "T Bailey",
        "year": "2013",
        "journal-title": "PLoS Comput Biol"
      },
      {
        "key": "ref5",
        "doi-asserted-by": "crossref",
        "first-page": "e1000424",
        "DOI": "10.1371/journal.pcbi.1000424",
        "article-title": "A quick guide to organizing computational biolo
gy projects",
        "volume": "5",
        "author": "W Noble",
        "year": "2009",
        "journal-title": "PLoS Comput Biol"
      },
      {
        "key": "ref6",
        "doi-asserted-by": "crossref",
        "first-page": "e1002598",
        "DOI": "10.1371/journal.pcbi.1002598",
        "article-title": "A quick guide to software licensing for the sci
entist-programmer",
        "volume": "8",
        "author": "A Morin",
        "year": "2012",
        "journal-title": "PLoS Comput Biol"
      }
    ],
    "container-title": [
      "PLOS Computational Biology"
    ],
    "original-title": [],
    "language": "en",
    "link": [
      {
        "URL": "http://dx.plos.org/10.1371/journal.pcbi.1004668",
        "content-type": "unspecified",
        "content-version": "vor",
        "intended-application": "similarity-checking"
```

```
      }
    ],
    "deposited": {
      "date-parts": [
        [
          2017,
          6,
          24
        ]
      ],
      "date-time": "2017-06-24T04:04:33Z",
      "timestamp": 1498277073000
    },
    "score": 1.0,
    "subtitle": [],
    "editor": [
      {
        "given": "Francis",
        "family": "Ouellette",
        "sequence": "first",
        "affiliation": []
      }
    ],
    "short-title": [],
    "issued": {
      "date-parts": [
        [
          2016,
          1,
          19
        ]
      ]
    },
    "references-count": 6,
    "journal-issue": {
      "published-online": {
        "date-parts": [
          [
            2016,
            1,
            19
          ]
        ]
      },
      "issue": "1"
    },
    "URL": "http://dx.doi.org/10.1371/journal.pcbi.1004668",
    "relation": {
      "cites": []
    },
    "ISSN": [
      "1553-7358"
    ],
    "issn-type": [
      {
        "value": "1553-7358",
        "type": "electronic"
      }
    ]
  }
}
```

In [14]:
```
#Nur den Inhalt von message ausgeben lassen:
doi_data["message"]
```

Out[14]:

```
{'indexed': {'date-parts': [[2019, 11, 9]],
  'date-time': '2019-11-09T08:23:20Z',
  'timestamp': 1573287800480},
 'reference-count': 6,
 'publisher': 'Public Library of Science (PLoS)',
 'issue': '1',
 'license': [{'URL': 'http://creativecommons.org/licenses/by/4.0/',
   'start': {'date-parts': [[2016, 1, 19]],
    'date-time': '2016-01-19T00:00:00Z',
    'timestamp': 1453161600000},
   'delay-in-days': 0,
   'content-version': 'vor'}],
 'content-domain': {'domain': ['www.ploscompbiol.org'],
  'crossmark-restriction': False},
 'short-container-title': ['PLoS Comput Biol'],
 'DOI': '10.1371/journal.pcbi.1004668',
 'type': 'journal-article',
 'created': {'date-parts': [[2016, 1, 19]],
  'date-time': '2016-01-19T21:35:29Z',
  'timestamp': 1453239329000},
 'page': 'e1004668',
 'update-policy': 'http://dx.doi.org/10.1371/journal.pcbi.corrections_pol
icy',
 'source': 'Crossref',
 'is-referenced-by-count': 39,
 'title': ['A Quick Introduction to Version Control with Git and GitHu
b'],
 'prefix': '10.1371',
 'volume': '12',
 'author': [{'given': 'John D.',
   'family': 'Blischak',
   'sequence': 'first',
   'affiliation': []},
  {'given': 'Emily R.',
   'family': 'Davenport',
   'sequence': 'additional',
   'affiliation': []},
  {'given': 'Greg',
   'family': 'Wilson',
   'sequence': 'additional',
   'affiliation': []}],
 'member': '340',
 'published-online': {'date-parts': [[2016, 1, 19]]},
 'reference': [{'key': 'ref1',
   'doi-asserted-by': 'crossref',
   'first-page': '7',
   'DOI': '10.1186/1751-0473-8-7',
   'article-title': 'Git can facilitate greater reproducibility and incre
ased transparency in science',
   'volume': '8',
   'author': 'K Ram',
   'year': '2013',
   'journal-title': 'Source Code Biol Med'},
  {'key': 'ref2',
   'doi-asserted-by': 'crossref',
   'first-page': 'e1001745',
   'DOI': '10.1371/journal.pbio.1001745',
   'article-title': 'Best practices for scientific computing',
   'volume': '12',
   'author': 'G Wilson',
   'year': '2014',
```

```
    'journal-title': 'PLoS Biol'},
   {'key': 'ref3',
    'doi-asserted-by': 'crossref',
    'first-page': '57',
    'DOI': '10.1038/nature11247',
    'article-title': 'An integrated encyclopedia of DNA elements in the hu
man genome',
    'volume': '489',
    'author': 'B Bernstein',
    'year': '2012',
    'journal-title': 'Nature'},
   {'key': 'ref4',
    'doi-asserted-by': 'crossref',
    'first-page': 'e1003326',
    'DOI': '10.1371/journal.pcbi.1003326',
    'article-title': 'Practical guidelines for the comprehensive analysis
of ChIP-seq data',
    'volume': '9',
    'author': 'T Bailey',
    'year': '2013',
    'journal-title': 'PLoS Comput Biol'},
   {'key': 'ref5',
    'doi-asserted-by': 'crossref',
    'first-page': 'e1000424',
    'DOI': '10.1371/journal.pcbi.1000424',
    'article-title': 'A quick guide to organizing computational biology pr
ojects',
    'volume': '5',
    'author': 'W Noble',
    'year': '2009',
    'journal-title': 'PLoS Comput Biol'},
   {'key': 'ref6',
    'doi-asserted-by': 'crossref',
    'first-page': 'e1002598',
    'DOI': '10.1371/journal.pcbi.1002598',
    'article-title': 'A quick guide to software licensing for the scientis
t-programmer',
    'volume': '8',
    'author': 'A Morin',
    'year': '2012',
    'journal-title': 'PLoS Comput Biol'}],
  'container-title': ['PLOS Computational Biology'],
  'original-title': [],
  'language': 'en',
  'link': [{'URL': 'http://dx.plos.org/10.1371/journal.pcbi.1004668',
    'content-type': 'unspecified',
    'content-version': 'vor',
    'intended-application': 'similarity-checking'}],
  'deposited': {'date-parts': [[2017, 6, 24]],
   'date-time': '2017-06-24T04:04:33Z',
   'timestamp': 1498277073000},
  'score': 1.0,
  'subtitle': [],
  'editor': [{'given': 'Francis',
    'family': 'Ouellette',
    'sequence': 'first',
    'affiliation': []}],
  'short-title': [],
  'issued': {'date-parts': [[2016, 1, 19]]},
  'references-count': 6,
  'journal-issue': {'published-online': {'date-parts': [[2016, 1, 19]]},
```

```
 'issue': '1'},
'URL': 'http://dx.doi.org/10.1371/journal.pcbi.1004668',
'relation': {'cites': []},
'ISSN': ['1553-7358'],
'issn-type': [{'value': '1553-7358', 'type': 'electronic'}]}
```

In [15]:
```
#Schlüssel von message anzeigen lassen:
doi_data["message"].keys()
```

Out[15]:
```
dict_keys(['indexed', 'reference-count', 'publisher', 'issue', 'license',
'content-domain', 'short-container-title', 'DOI', 'type', 'created', 'pag
e', 'update-policy', 'source', 'is-referenced-by-count', 'title', 'prefi
x', 'volume', 'author', 'member', 'published-online', 'reference', 'conta
iner-title', 'original-title', 'language', 'link', 'deposited', 'score',
'subtitle', 'editor', 'short-title', 'issued', 'references-count', 'journ
al-issue', 'URL', 'relation', 'ISSN', 'issn-type'])
```

In [25]:
```
#Nur den Titel ausgeben lassen (kurzer Weg):
doi_data["message"]["title"]
#Achtung: kommt als Liste zurück = []

#Langer Weg:
#message_data = doi_data["message"]
#message_data["title"]
```

Out[25]:
```
['A Quick Introduction to Version Control with Git and GitHub']
```

In [27]:
```
#Typ überprüfen:
type(doi_data["message"]["title"])
```

Out[27]:
```
list
```

In [26]:
```
#Titel als String ausgeben lassen:
doi_data["message"]["title"][0]
```

Out[26]:
```
'A Quick Introduction to Version Control with Git and GitHub'
```

In [28]:
```
#Typ überprüfen:
type(doi_data["message"]["title"][0])
```

Out[28]:
```
str
```

Übung

Für 4 DOIs jeweils "title", "short-container-title" und "doi" ausgeben lassen.

In [90]:
```
#DOIs an die Stamm-URL anhängen:

base_url = "https://api.crossref.org/works/"

dois = ["10.1371/journal.pcbi.1004668",
        "10.21105/joss.01035",
        "10.1038/35057062",
        "10.21105/joss.01006"]

for doi in dois:
    print(base_url + doi)
```

```
https://api.crossref.org/works/10.1371/journal.pcbi.1004668
https://api.crossref.org/works/10.21105/joss.01035
https://api.crossref.org/works/10.1038/35057062
https://api.crossref.org/works/10.21105/joss.01006
```

In [87]:
```
#Für die so erstellten Links die gewünschten Angaben ausgebenlassen:

for doi in dois:
    print(base_url + doi)
    full_url = base_url + doi
    doi_json_data = urllib.request.urlopen(full_url).read()
    doi_data = json.loads(doi_json_data)
    print(doi_data["message"]["title"][0])
    print(doi_data["message"]["short-container-title"][0])
    print(doi)
    print("\n")
```

```
https://api.crossref.org/works/10.1371/journal.pcbi.1004668
A Quick Introduction to Version Control with Git and GitHub
PLoS Comput Biol
10.1371/journal.pcbi.1004668


https://api.crossref.org/works/10.21105/joss.01035
nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Clim
atology Data Client for R
JOSS
10.21105/joss.01035


https://api.crossref.org/works/10.1038/35057062
Initial sequencing and analysis of the human genome
Nature
10.1038/35057062


https://api.crossref.org/works/10.21105/joss.01006
SeqTools: A python package for easy transformation, combination and evalu
ation of large datasets.
JOSS
10.21105/joss.01006
```

In [101]:
```
#Feldnamen voranstellen:

for doi in dois:
    full_url = base_url + doi
    doi_json_data = urllib.request.urlopen(full_url).read()
    doi_data = json.loads(doi_json_data)
    print("Titel: " + doi_data["message"]["title"][0])
    print("Zeitschrift: " + doi_data["message"]["short-container-title"]
[0])
    print("DOI: " + doi)
    print("CrossRef-URL: " + base_url + doi)
    print()
```

```
Titel: A Quick Introduction to Version Control with Git and GitHub
Zeitschrift: PLoS Comput Biol
DOI: 10.1371/journal.pcbi.1004668
CrossRef-URL: https://api.crossref.org/works/10.1371/journal.pcbi.1004668

Titel: nasapower: A NASA POWER Global Meteorology, Surface Solar Energy a
nd Climatology Data Client for R
Zeitschrift: JOSS
DOI: 10.21105/joss.01035
CrossRef-URL: https://api.crossref.org/works/10.21105/joss.01035

Titel: Initial sequencing and analysis of the human genome
Zeitschrift: Nature
DOI: 10.1038/35057062
CrossRef-URL: https://api.crossref.org/works/10.1038/35057062

Titel: SeqTools: A python package for easy transformation, combination an
d evaluation of large datasets.
Zeitschrift: JOSS
DOI: 10.21105/joss.01006
CrossRef-URL: https://api.crossref.org/works/10.21105/joss.01006
```

In [105]:
```
#Angaben verketten:

for doi in dois:
    full_url = base_url + doi
    doi_json_data = urllib.request.urlopen(full_url).read()
    doi_data = json.loads(doi_json_data)
    print(doi_data["message"]["title"][0]
          + ". In: " + doi_data["message"]["short-container-title"][0]
          + ". DOI: " + doi)
    print()
```

```
A Quick Introduction to Version Control with Git and GitHub. In: PLoS Com
put Biol. DOI: 10.1371/journal.pcbi.1004668

nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Clim
atology Data Client for R. In: JOSS. DOI: 10.21105/joss.01035

Initial sequencing and analysis of the human genome. In: Nature. DOI: 10.
1038/35057062

SeqTools: A python package for easy transformation, combination and evalu
ation of large datasets.. In: JOSS. DOI: 10.21105/joss.01006
```