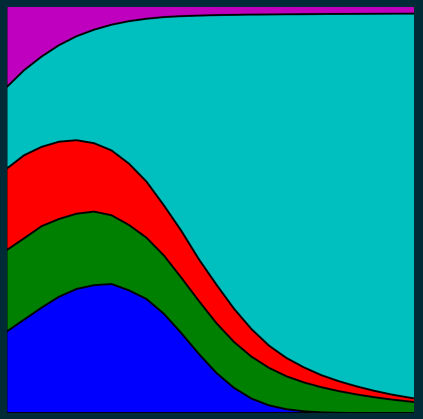
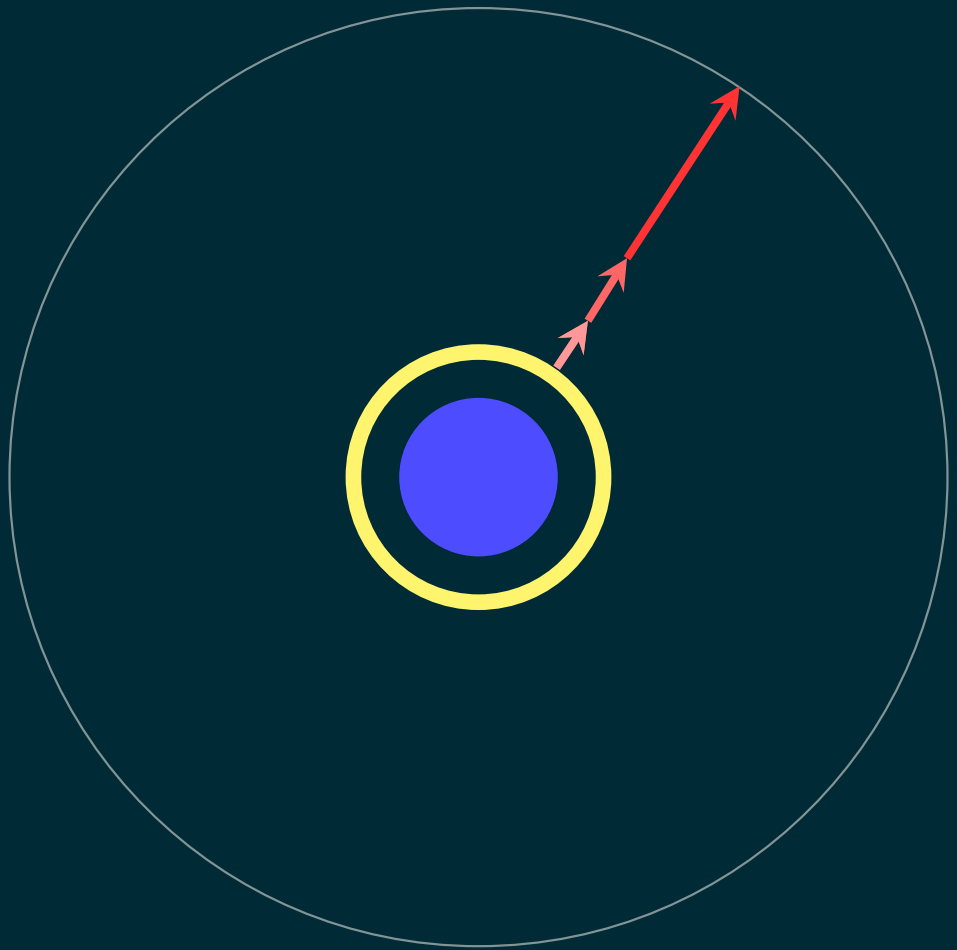
Accessing open research literature with Python

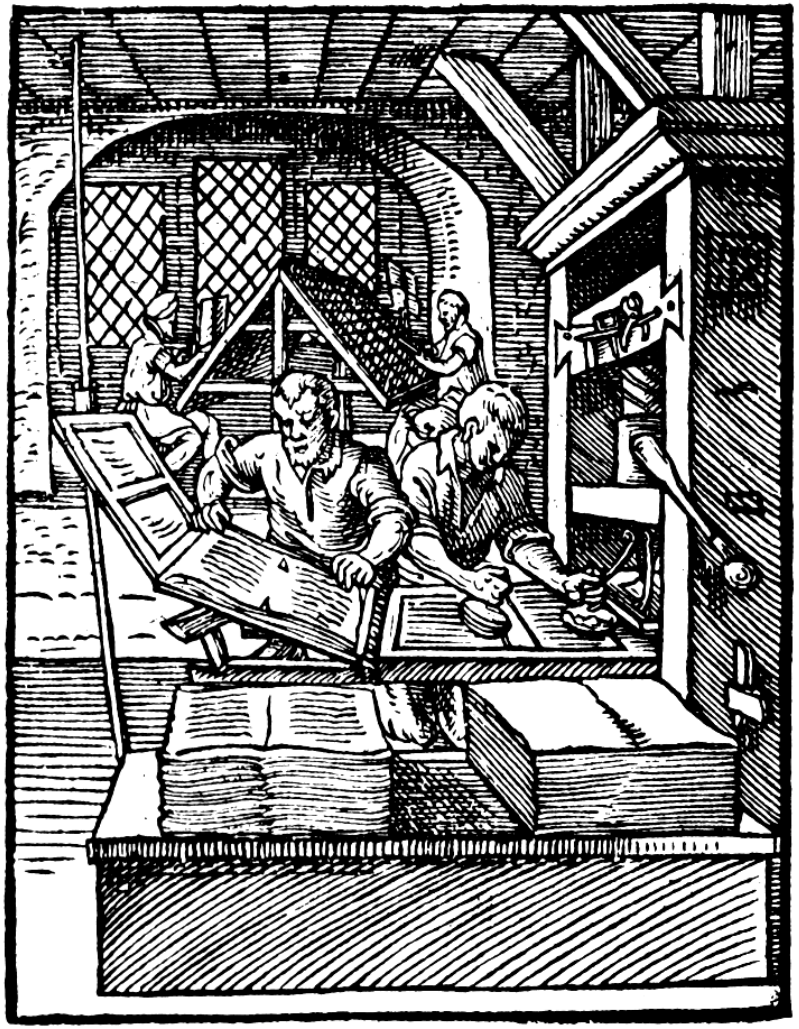
@NikoletaGlyn

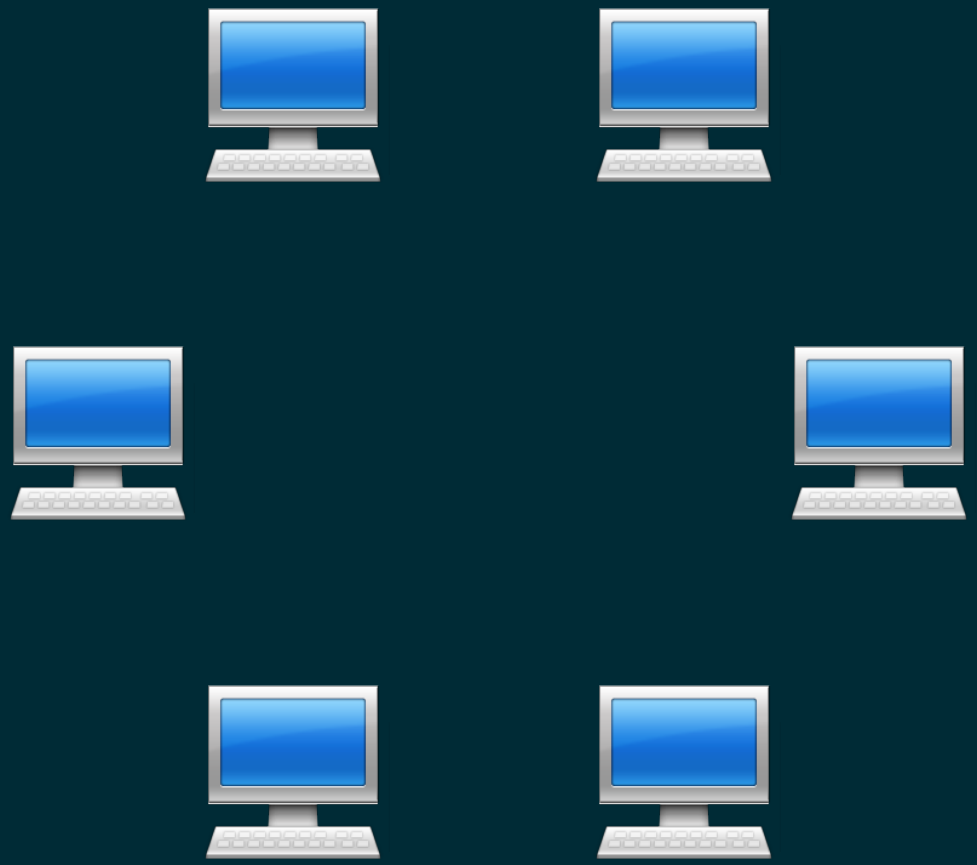


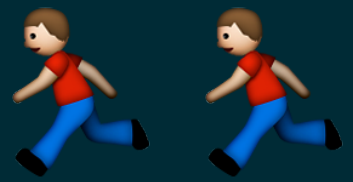
Software
Sustainability
Institute

http://matt.might.net/articles/phd-school-in-pictures/









IEEE - Search Results

Secure | https://www.ieee.org/searchresults/index.html?cx=006539740418318249752%3Af2h38l7gvis&cof=FORID%3A11&qp=&ie=UTF-8&oe=UTF-8&q=Namibia

Apps | Vince Knight | Home - Intranet | GitHub | Axelbib | nbdime/README | Game Theory Expl | Google Calendar | ear | Training the tagge

Search Results

Refine results for: **Namibia**

All results | Societies & Communities | Subscriptions
Membership & Services | Conferences | Multimedia
Future Directions | Standards | Tutorials & Webinars
IEEE Spectrum | News

Showing 1 - 10 of 54,900,000 for Namibia (0.44 seconds)

Open Data Portal - A Technical Enabler to Drive Innovation in **Namibia**
Promising a world of endless opportunities, open data is increasingly becoming a global phenomenon, attracting interests from many governments, such as the ...
www.computer.org

Namibia Science Lab - EPICS in IEEE
Sep 8, 2015 ... In **Namibia**, science is a difficult subject to teach due to a lack of resources and qualified, trained teachers. Schools that do offer physical ...
epics.ieee.org

Beetles, Cacti, and Killer Plants Inspire Energy Efficiency - IEEE ...
Feb 24, 2016 ... Photo: Martin Harvey/Alamy Darkling beetle harvesting fog in the **Namibian** desert. Joanna Aizenberg, whose Harvard lab produced the work, ...
spectrum.ieee.org

IEEE Women in Engineering (WIE)
File Format: PDF/Adobe Acrobat
Changing Lives - **Namibia**. 19. You will reach 160 countries as a recognized leader in supporting women and the education of young girls in engineering and ...
www.ieee.org

$$100x(0.5min + 1.5min) + 10X(0.5min) =$$
$$110min \Rightarrow 1h \text{ and } 50min$$

API

'QUERY'

```
http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?
```

'QUERY'

```
http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?
```

```
...ti=Namibia
```

'QUERY'

```
http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?
```

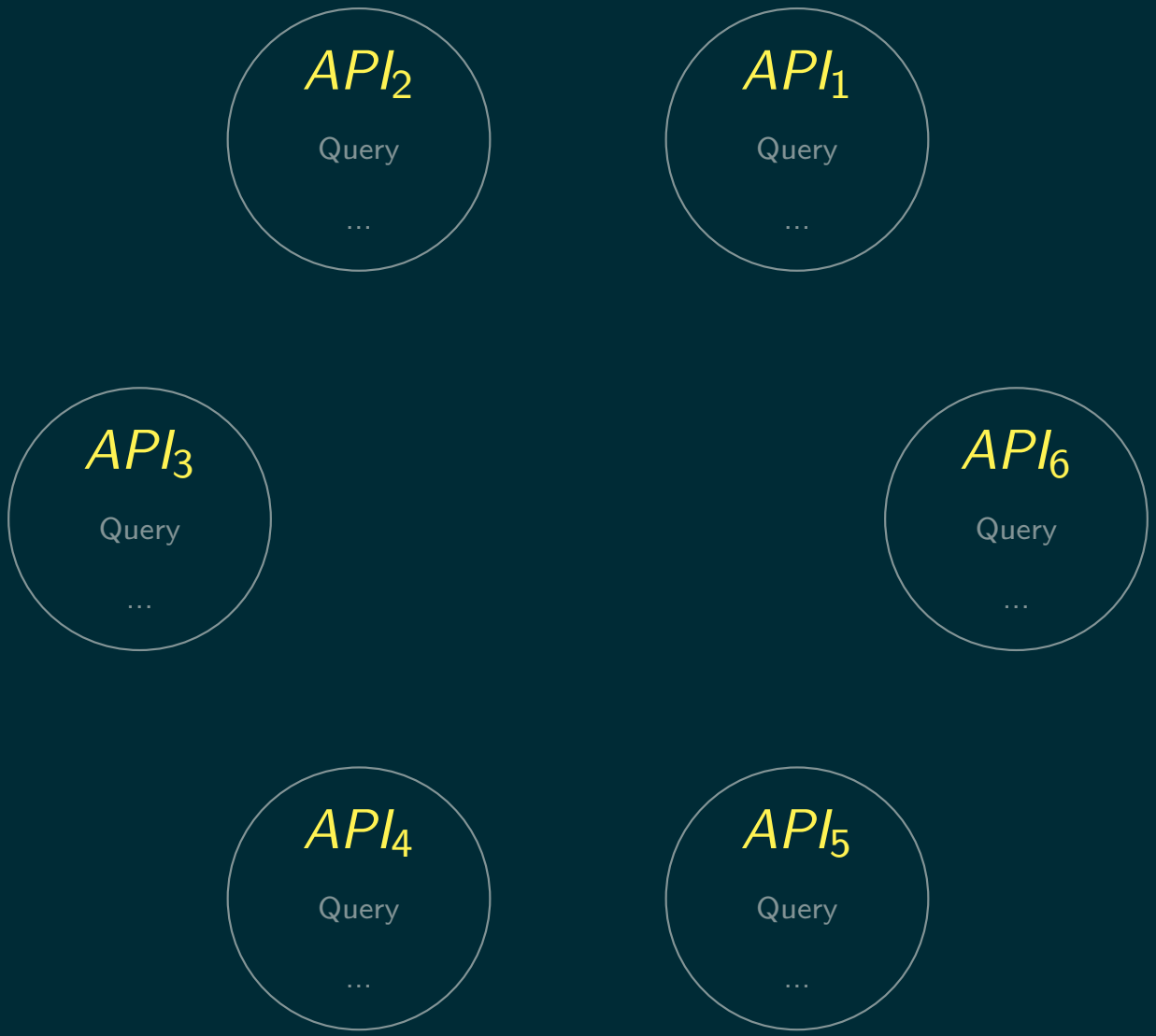
```
...ti=Namibia
```

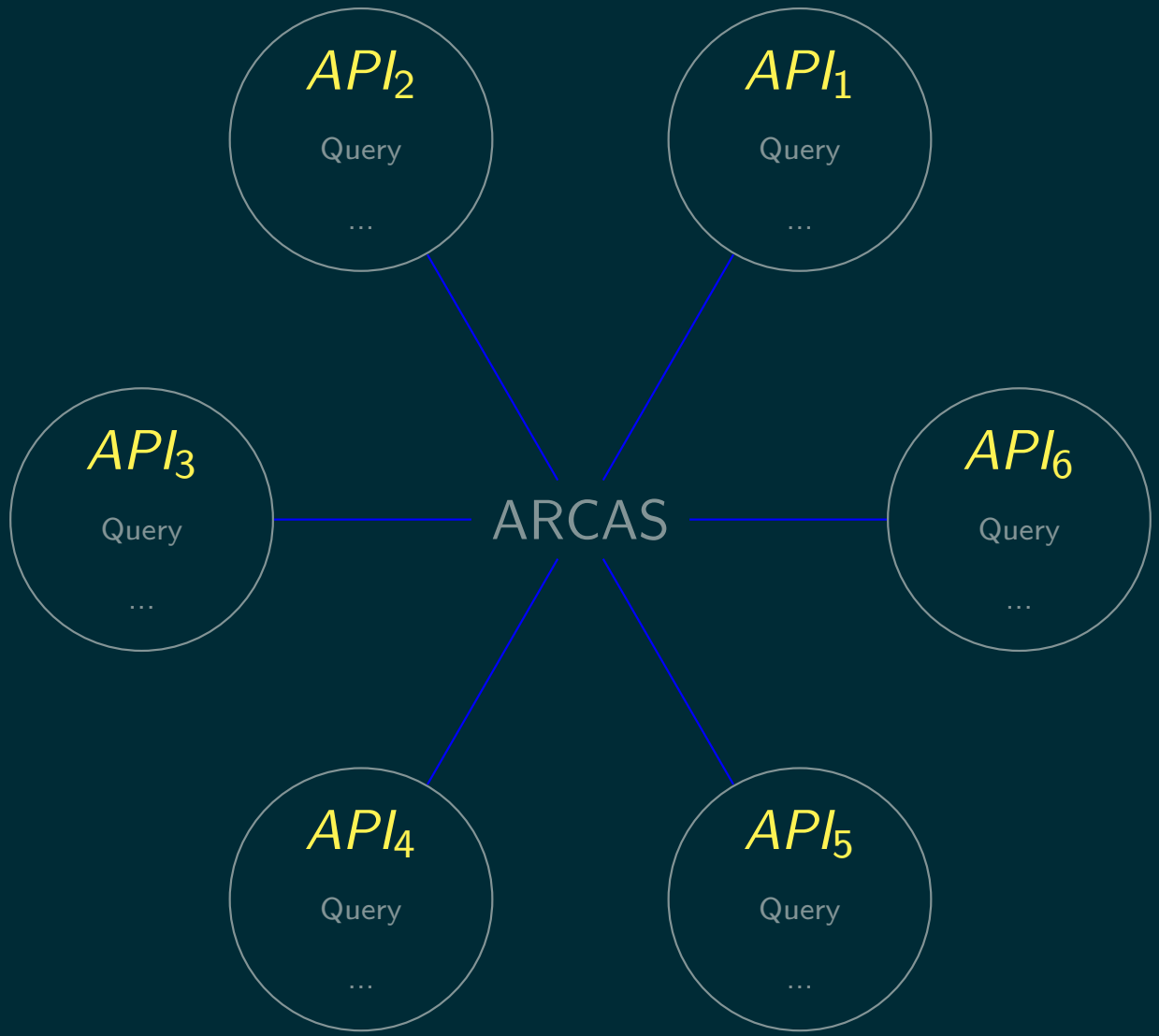
```
...ti=Namibia&hc=100
```

'RESULTS'

```
|--<root>
    |- ...
|-<document>
    |-<rank>1</rank>
    |-<title>
        |-<![CDATA[
            Implementation of 220 ...]]>
    |-</title>
    |-<authors>
        |- ...
    |- ...
|-</document>
|-<document>
```

$$15min + 1min + 50min = 66min \Rightarrow 1h \text{ and } 5min$$


API_2
Query
...
API_1
Query
...
API_3
Query
...
API_6
Query
...
API_4
Query
...
API_5
Query
...


API_2
Query
...
API_1
Query
...
API_3
Query
...
ARCAS
API_6
Query
...
API_4
Query
...
API_5
Query
...

```
pip install arcas
```

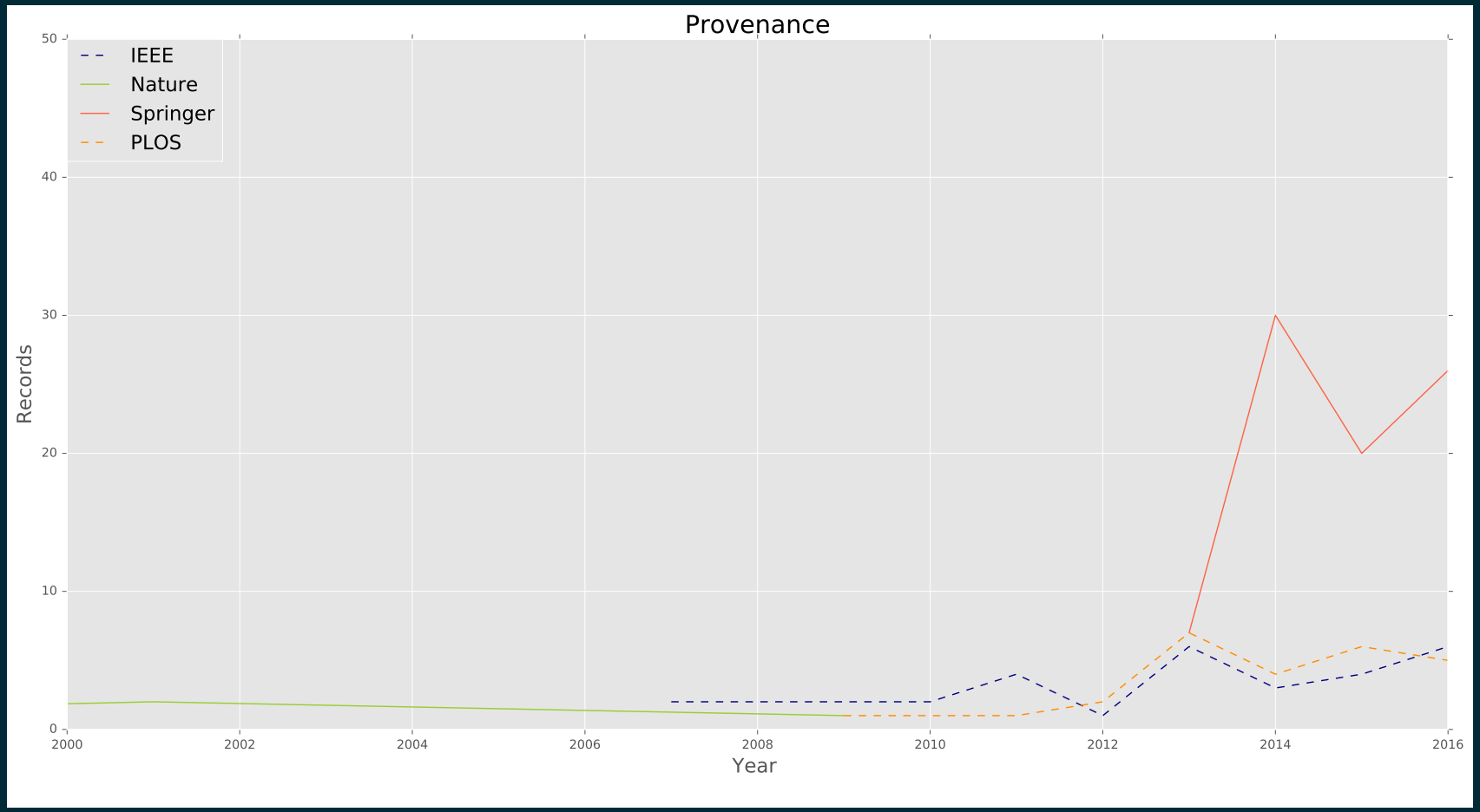
```
import pandas as pd
from arcas import *

def get_arguments(word, start, count):
    arguments = {'-a': None, '-t': word, '-s': start,
                 '-r': count, '-y': None, '-b': None}
    return arguments

def main_program(arguments):
    parameters = pp.parameters_fix(arguments=arguments)
    url = pp.create_url_search(parameters=parameters)
    response = pp.make_request(url)
    root = pp.get_root(response)
    article = pp.parse(root)
    return article, url
```

```
words = ["Namibia"]

apis = {"ieee": Ieee, "nature": Nature, "arxiv": Arxiv,
        "springer": Springer, "plos": Plos}
list_apis = ['plos', 'arxiv', 'ieee', 'nature', 'springer']
count = 10
```

```
for p in list_apis:
    pp = apis[p]()
    start = 1
    while start < 100:
        arguments = get_arguments(words[0], start, count)
        raw_articles, url = main_program(arguments)

        dfs = []
        if raw_articles is not False:
            for raw_article in raw_articles:
                try:
                    df = pp.to_dataframe(raw_article)
                    dfs.append(df)
                except ValueError:
                    ValueError()
            try:
                df = pd.concat(dfs, ignore_index=True)
                pp.export(df, filename='articles/{}-{}-{}.json'.format(p, words[0], start))
            except ValueError:
                print(None)
        start += 10
```

```
{
"key":{"0":, "1":, "2": },
"unique_key":{"0":, "1":, "2":},
"title":{"0":,
         "1":,
         "2":},
"author":{"0":,"1":,"2":},
"abstract":{"0":,
            "1":,
            "2":},
"date":{"0":,"1":, "2":},
"journal":{"0":,"1":,"2":},
"provenance":{"0":,"1":,"2":}
}
```

$$15min + 5min = 20min$$


Provenance
IEEE
Nature
Springer
PLOS
Records
Year
0
10
20
30
40
50
2000
2002
2004
2006
2008
2010
2012
2014
2016

@NikoletaGlyn
https://github.com/Nikoleta-v3
ttps://github.com/Nikoleta-v3/Arcas