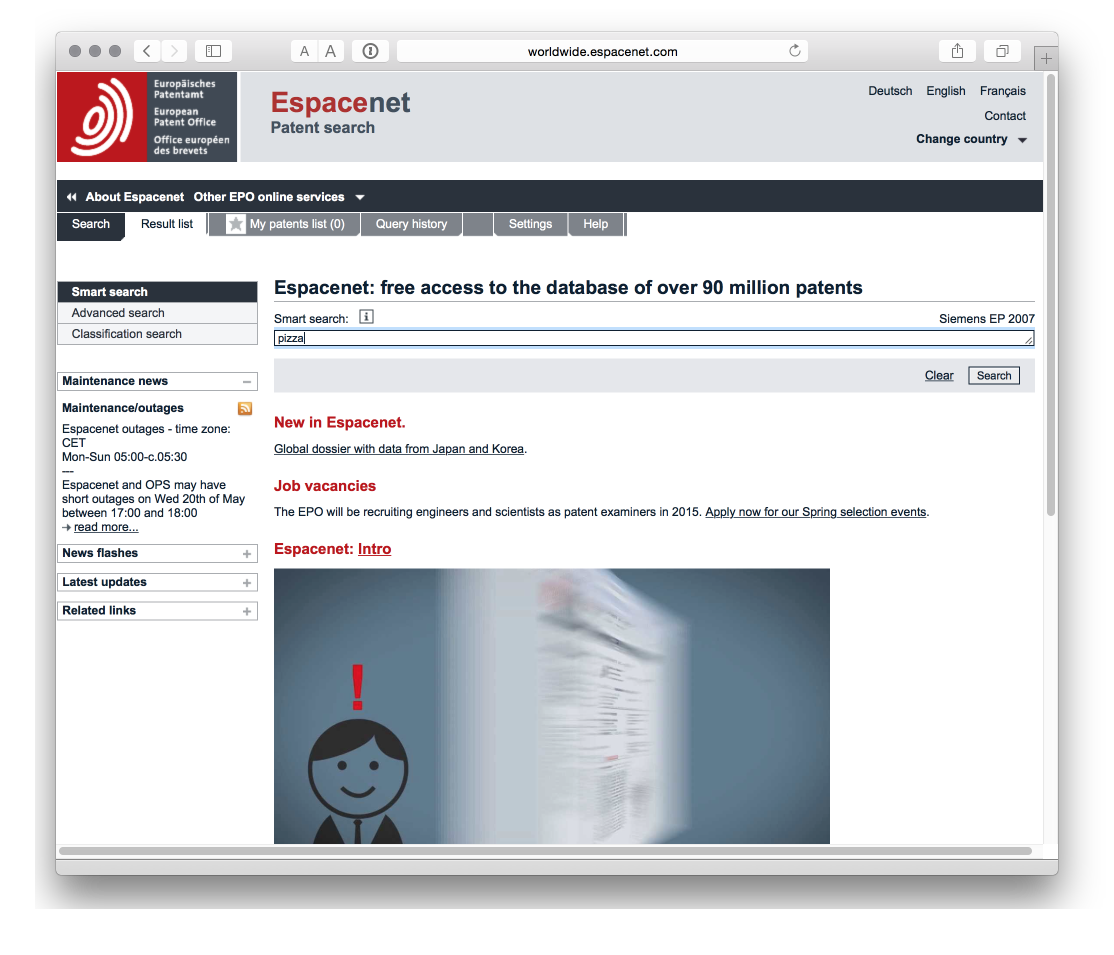
Obtaining Patent Data: Patent Databases

This article provides a quick overview of some of the main sources of free patent data. It is intended for quick reference and points to some free tools for accessing patent databases that you may not be familiar with.

It goes without saying that getting access to patent data in the first place is fundamental to patent analysis. There are quite a few free services out there and we will highlights some of the important ones. Most free sources have particular strengths or weaknesses such as the number of records that can be downloaded, the data fields that can be queried, the format the data comes back in or how clean data is in terms of the hours required to prepare for analysis. We won't go into all of that but will provide the odd pointer.

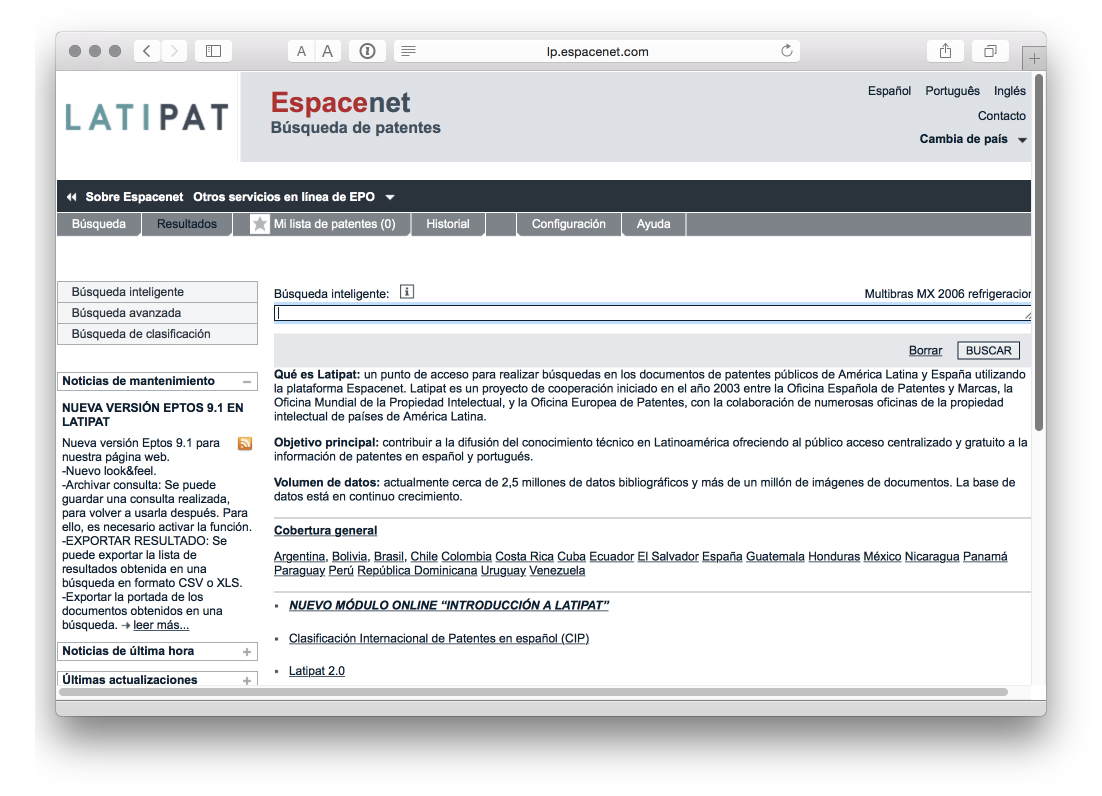
**1.** [**espacenet**](http://worldwide.espacenet.com/?locale=en_EP)

The best known free patent database from the European Patent Office.



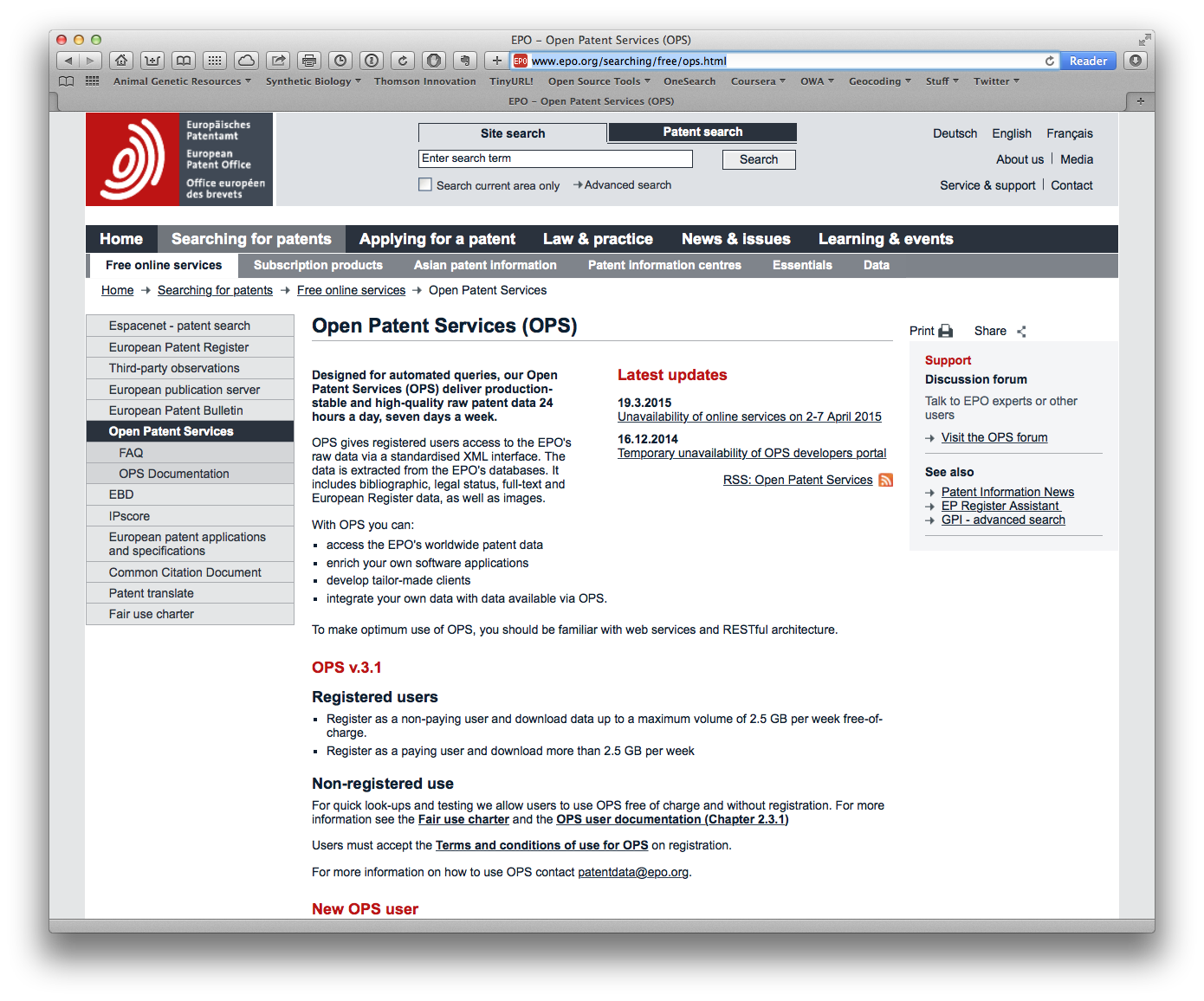
**2.** [**LATIPAT**](http://lp.espacenet.com)

For readers in Latin America (or Spain & Portugal) LATIPAT is a very useful resource.

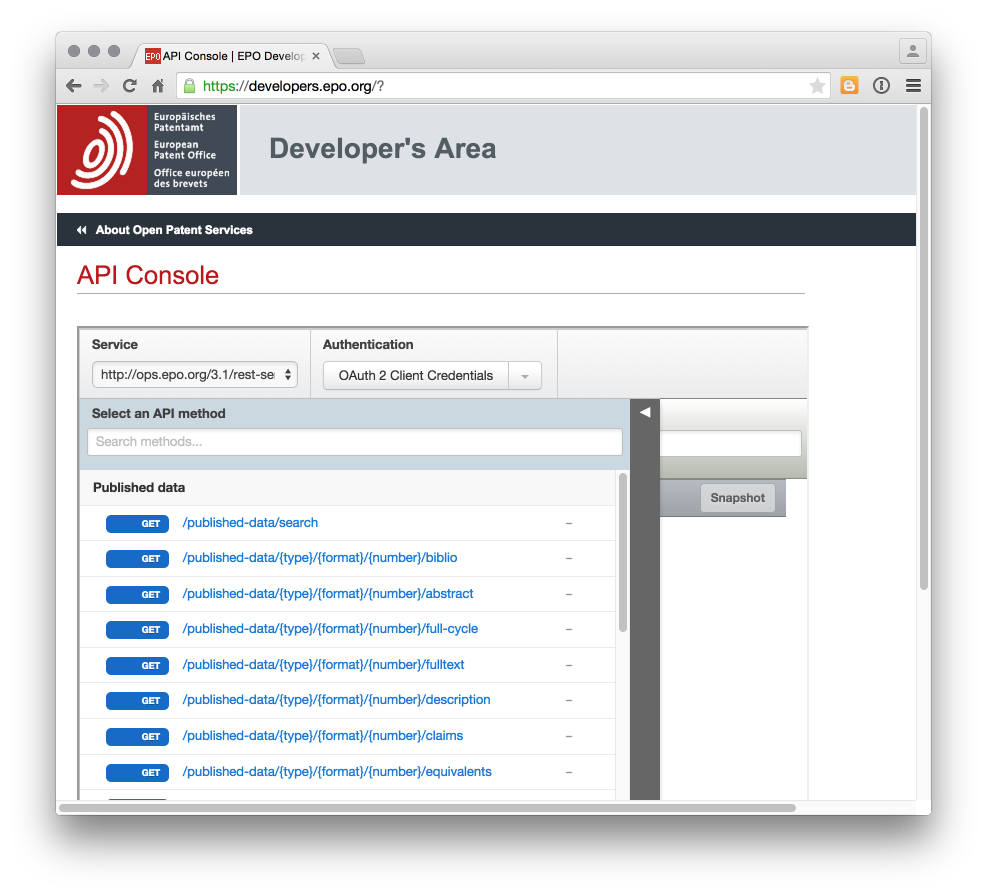


**3.** [**EPO Open Patent Services**](http://www.epo.org/searching/free/ops.html)

Access patent data through the EPO API free of charge.



The developer portal allows you to test your API queries and is recommended.

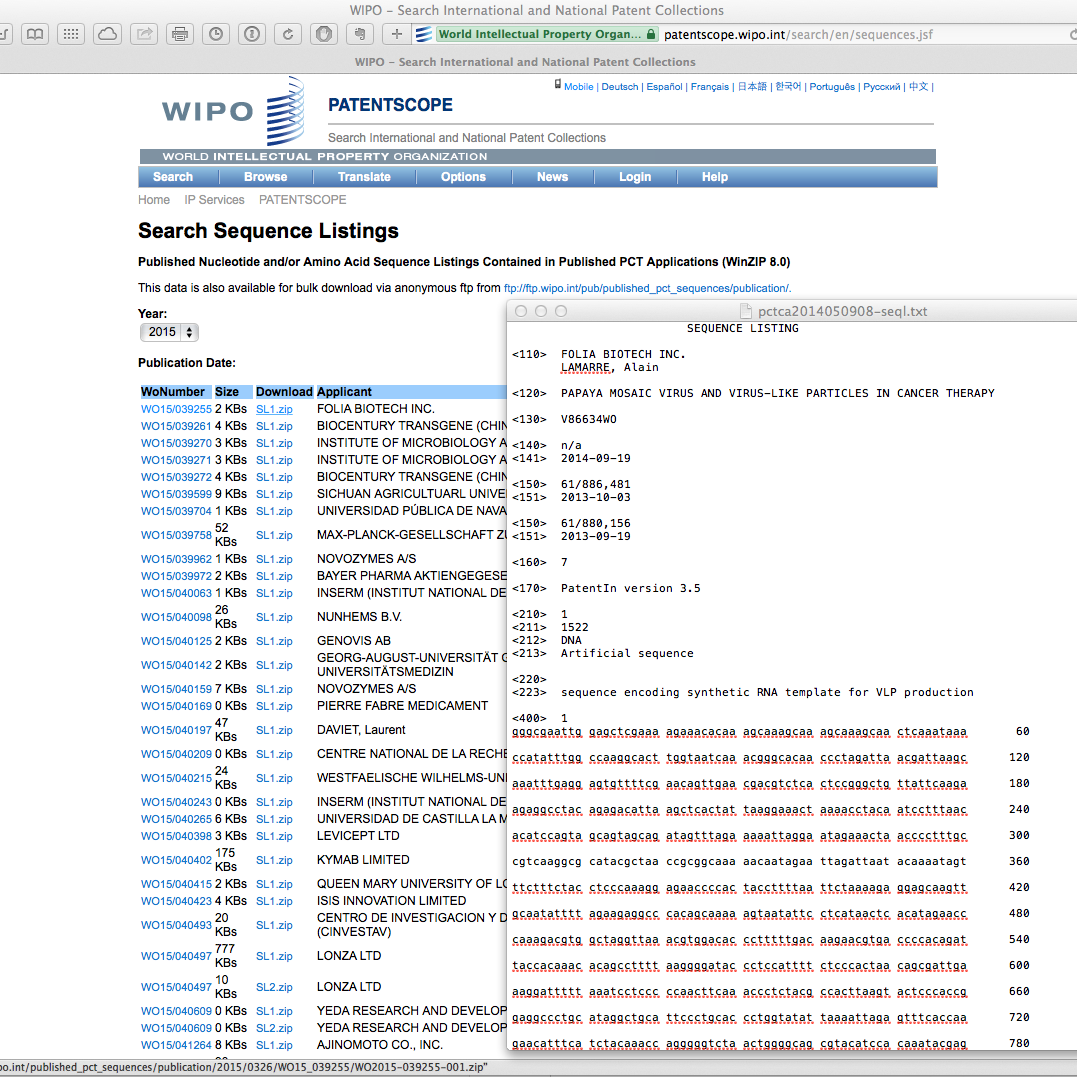


**4.** [**Patentscope**](https://patentscope.wipo.int/search/en/search.jsf)

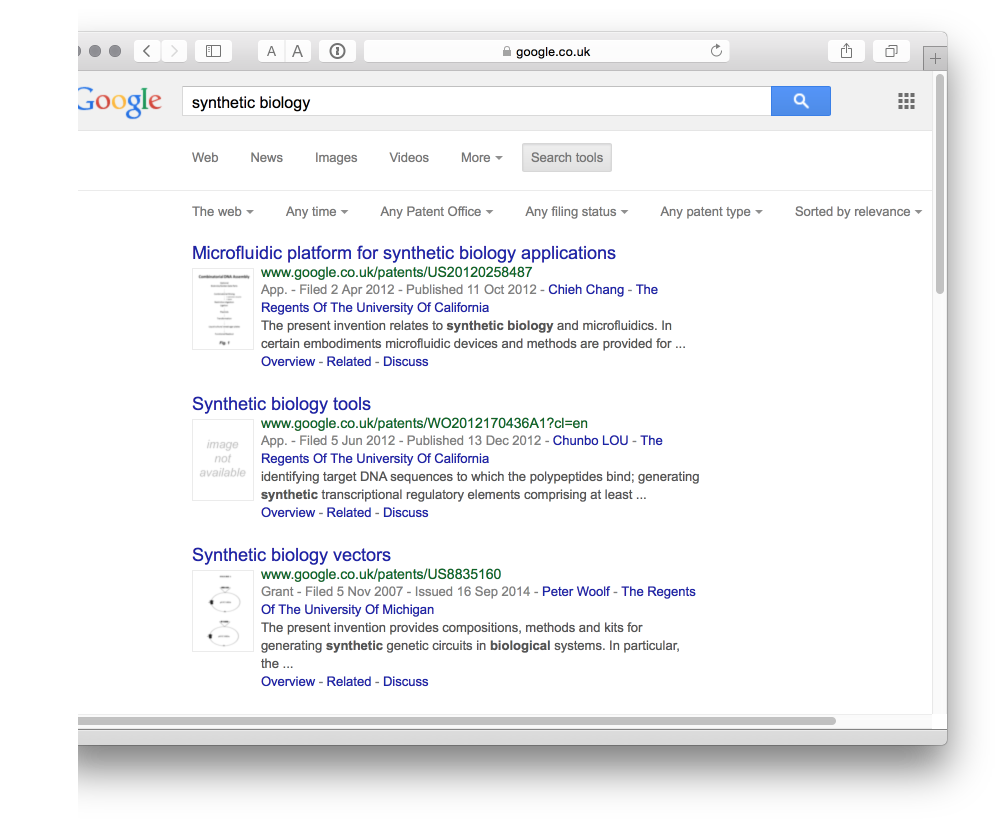
The WIPO Patentscope database provides access to Patent Cooperation Treaty data including downloads of a selection of fields (upto 10,000 records), a very useful [search expansion translation tool](https://patentscope.wipo.int/search/en/clir/clir.jsf?new=true), and [translation](https://www3.wipo.int/patentscope/translate/translate.jsf?interfaceLanguage=en).



Obtaining [sequence data from Patentscope](https://patentscope.wipo.int/search/en/sequences.jsf). Note that this rapidly becomes gigabytes of data.



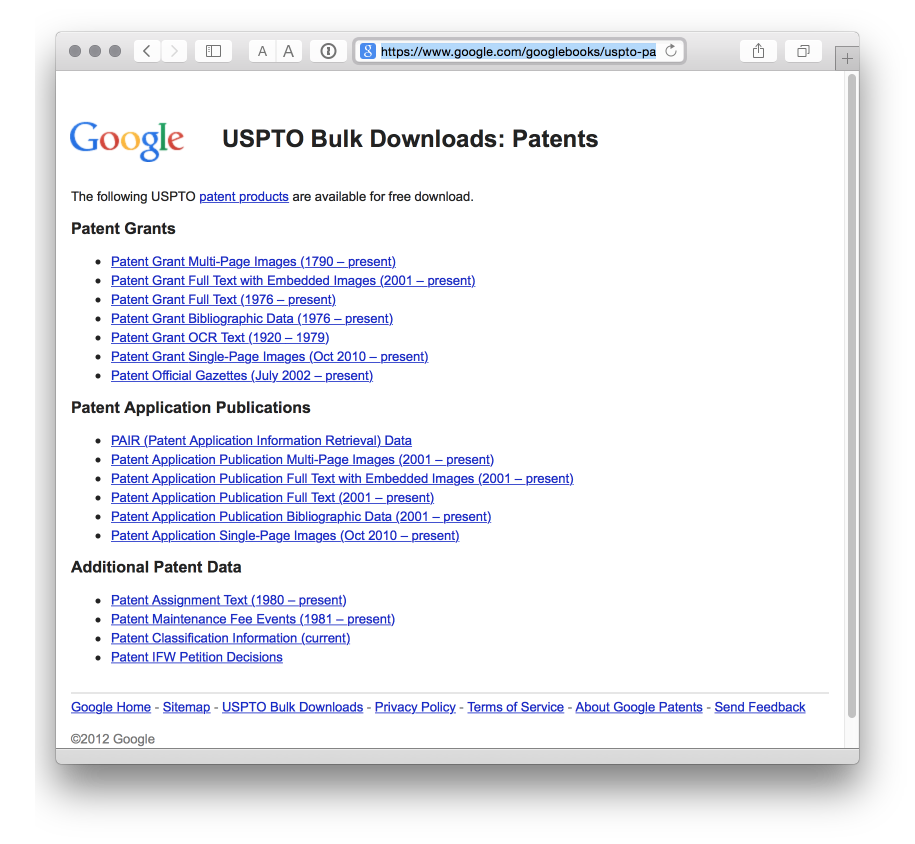
**5.** [**Google Patents**](http://www.google.com/patents)



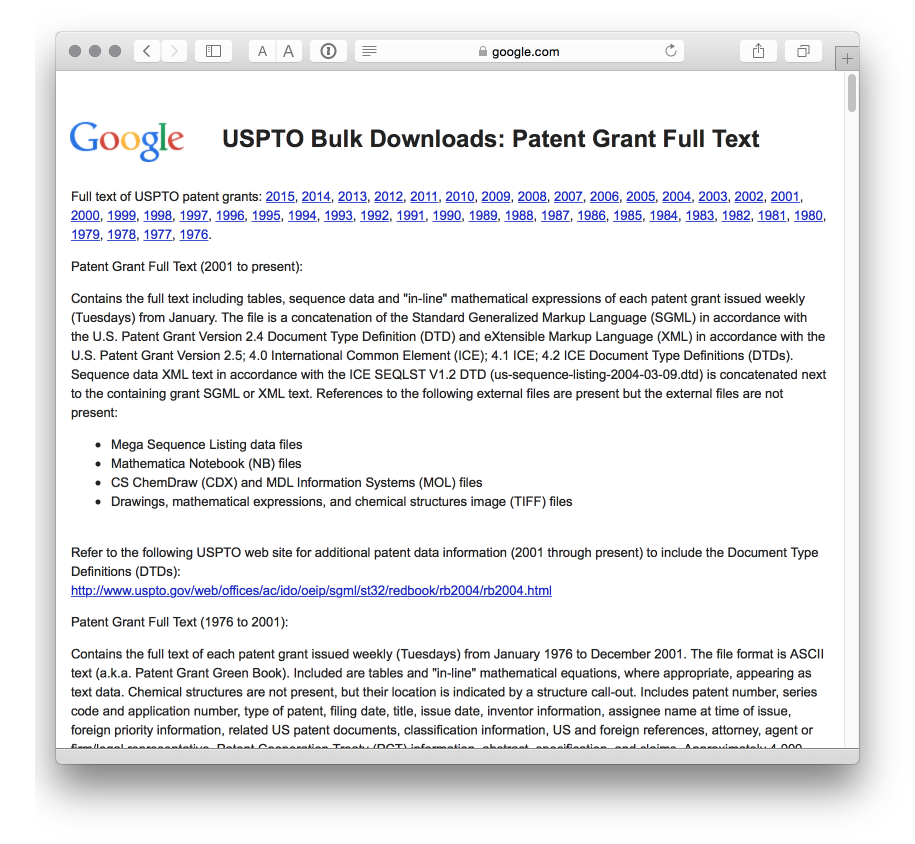
The [Google Patent Search API](https://developers.google.com/patent-search/terms) has been deprecated. Access through the Google Custom Search API with the API flag for patents [reported](http://stackoverflow.com/questions/15028166/python-module-for-searching-patent-databases-ie-uspto-or-epo) to be &tbm=pts with example code for using the API in Python. We will come back to this in another post.

**6.** [**Google USPTO Bulk download**](https://www.google.com/googlebooks/uspto.html)

The [USPTO patent databases](http://patft.uspto.gov) may be archaic but you can download the entire US collection from the [Google USPTO Bulk download service](https://www.google.com/googlebooks/uspto-patents.html).

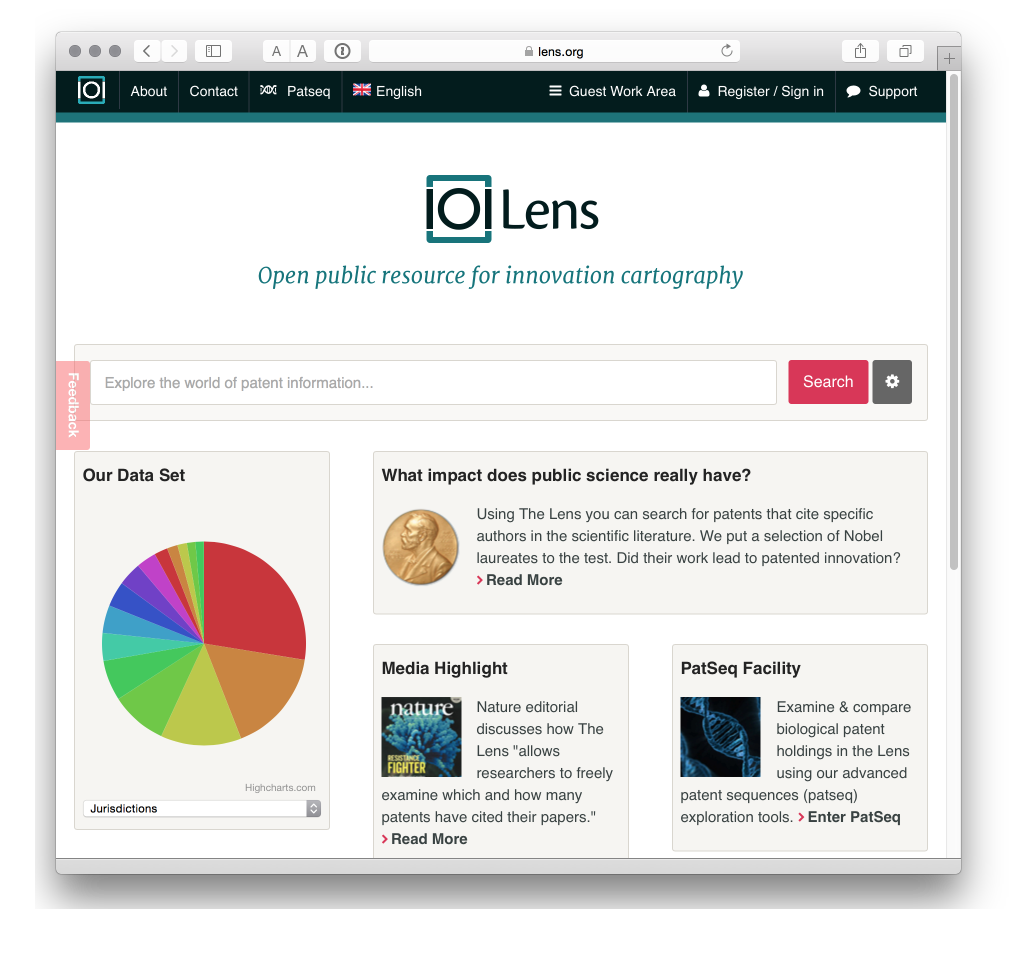


It is a fantastic service, and an example to patent offices everywhere on what can be done with patent data. If you have a good broadband connection and the hard drive space, it is quite good fun to suddenly have access to millions of patent records. We used the service to text mine the collection for millions of biological species names as reported [here](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078737).



**7.** [**The Lens**](https://www.lens.org/lens/)

Previously known as the Patent Lens this is a well designed site with quite a few visualisation options and access to sequence data. It is possible to share data but not, as far as I can work out, to export it. That seriously limits this site for patent analysis purposes unless you rely on their internal tools.



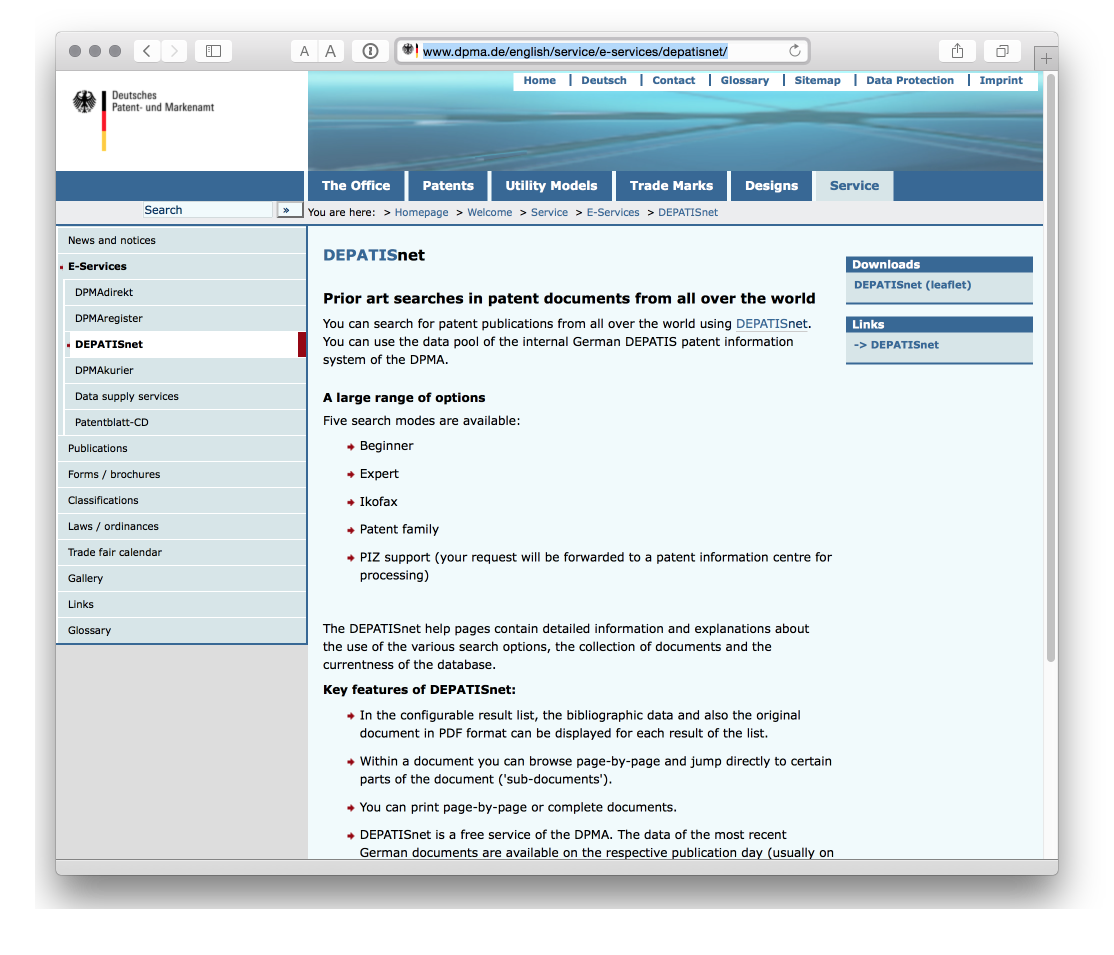
**8.** [**Free Patents Online**](http://www.freepatentsonline.com)

Sign up for a free account for enhanced access and to save and download data. It has been around quite a while now and while the download options are limited we really rather like it.



**9.** [**DEPATISnet**](http://www.dpma.de/english/service/e-services/depatisnet/)

We are not covering national databases. However, the patent database of the German Patent and Trademark Office struck us as potentially very useful. It allows for searches in English and German and has extensive coverage of international patent data, including the China, EP, US and PCT collections. The coverage details are [here](https://depatisnet.dpma.de/DepatisNet/depatisnet?action=datenbestand). Worth experimenting with.



**10.** [**OECD Patent Databases**](http://www.oecd.org/sti/inno/oecdpatentdatabases.htm)

One that is more for patent statisticians. The OECD has invested a lot of effort into developing patent indicators and resources including citations, the Harmonised Applicants names database [HAN database](http://www.oecd.org/sti/inno/43846611.pdf), mapping through the [REGPAT database](http://www.oecd.org/sti/inno/40794372.pdf) among other resources that are available free of charge.

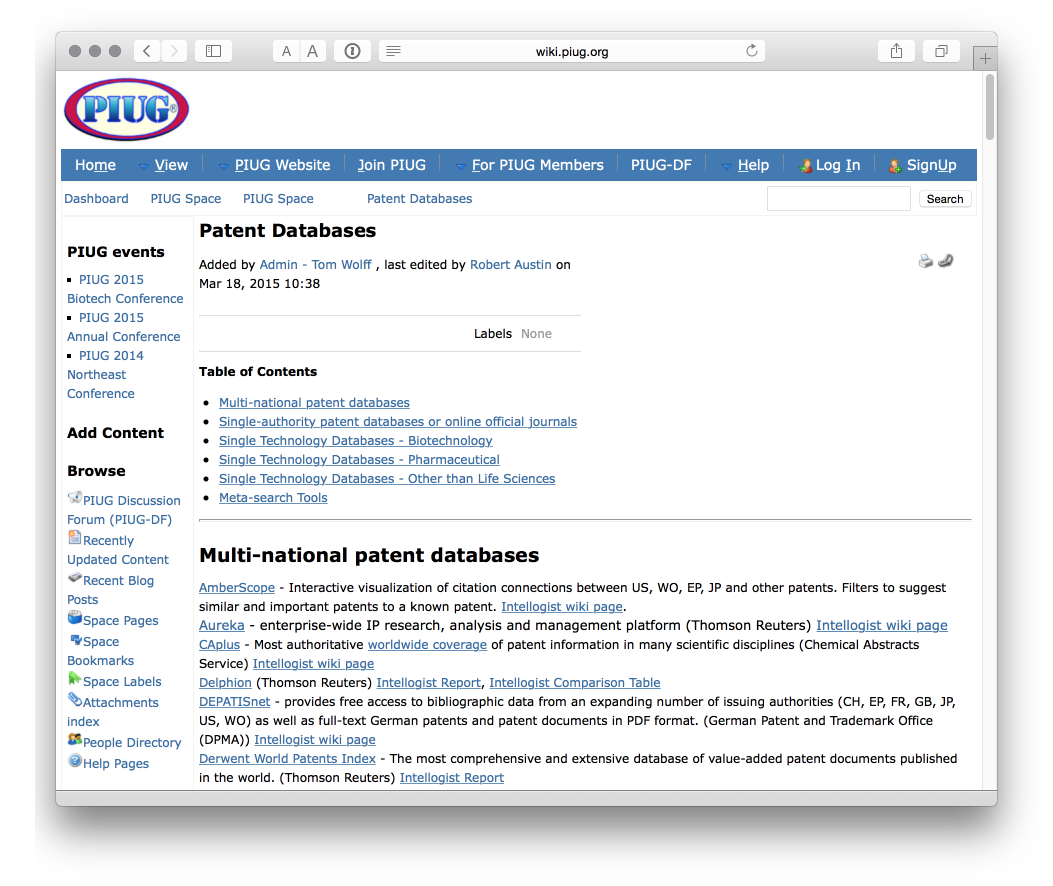


Along the same lines the US National Bureau of Economic Research [NBER US Patent Citations Data File](http://www.nber.org/patents/) is an important resource.

**11. Other data sources**

A number of companies provide access to patent data, typically with tiered access depending on your needs and budget. Examples include [Thomson Innovation](https://www.thomsoninnovation.com/login), [Questel Orbit](https://www.questel.com/index.php/en/), [STN](http://www.stn-international.de/index.php?id=123), and [PatBase](https://www.patbase.com/login.asp). We will not be focusing on these services but we will look at the use of data tools to work with data from services such as Thomson Innovation.

For more information on free and commercial data providers try the excellent [Patent Information User Group](http://www.piug.org) and its list of [Patent Databases](http://wiki.piug.org/display/PIUG/Patent+Databases) from Tom Wolff and Robert Austin.



Also worth mentioning is the Landon IP [Intellogist](http://www.intellogist.com/wiki/Main_Page) blog which maintains [Search System Reports](http://www.intellogist.com/wiki/Category:Intellogist_Reports)

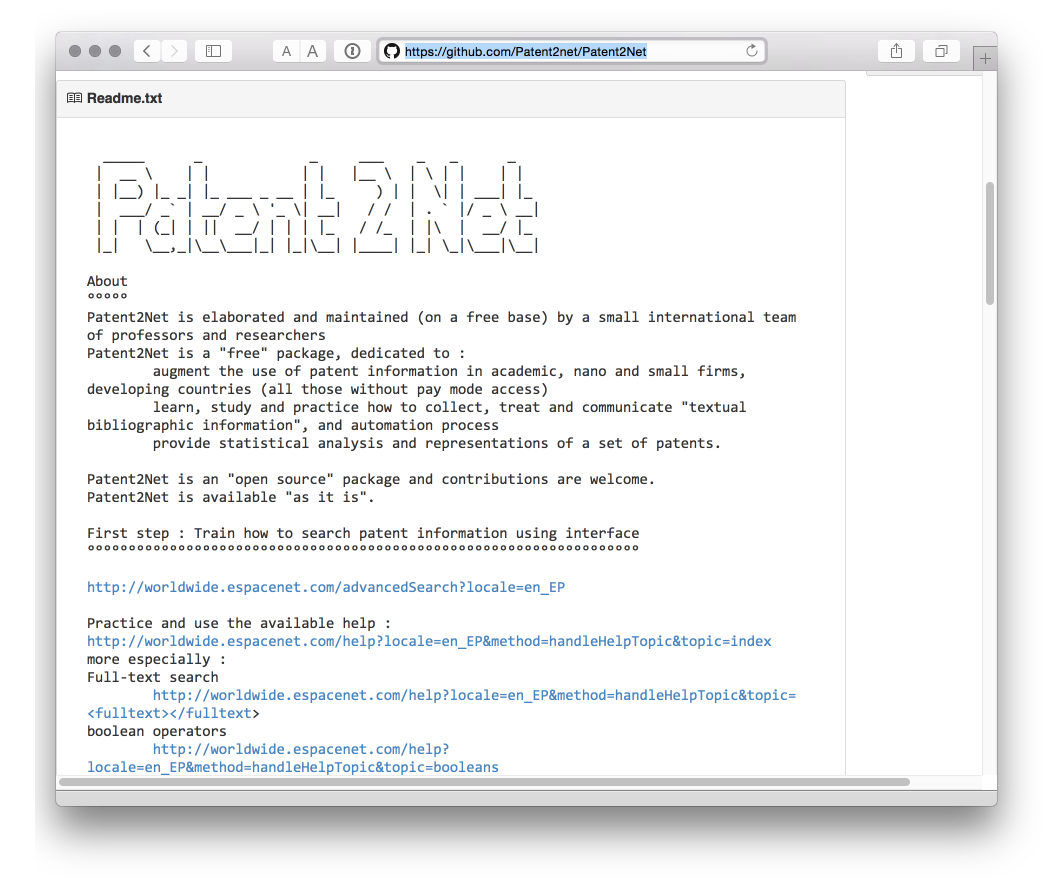


### Tools for Accessing Patent Data

In closing this article we will highlight a couple of tools for accessing patent data, typically using APIs and Python. We will come back to this later and are working to try this approach in R.

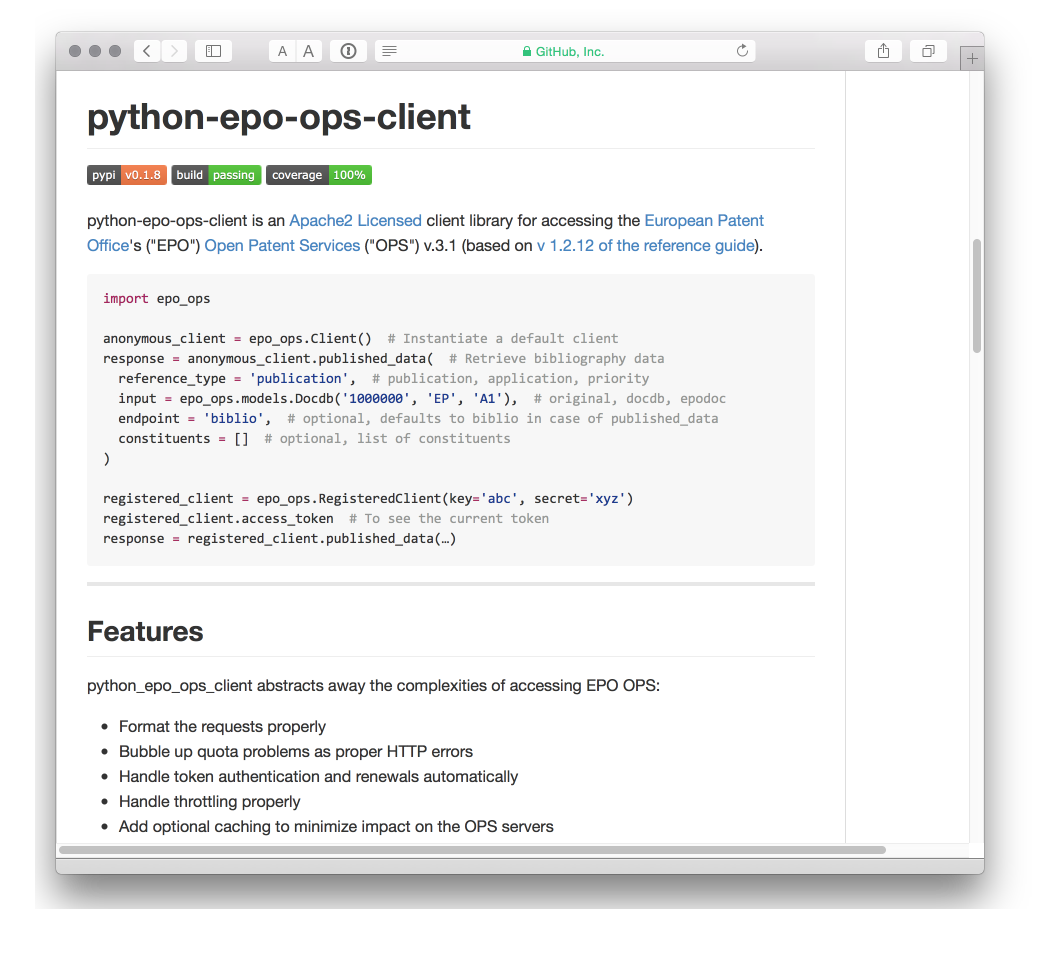
**12.** [**Patent2Net**](https://github.com/Patent2net/Patent2Net) **in Python**

A tool to access and process the data from the European Patent Office OPS service.



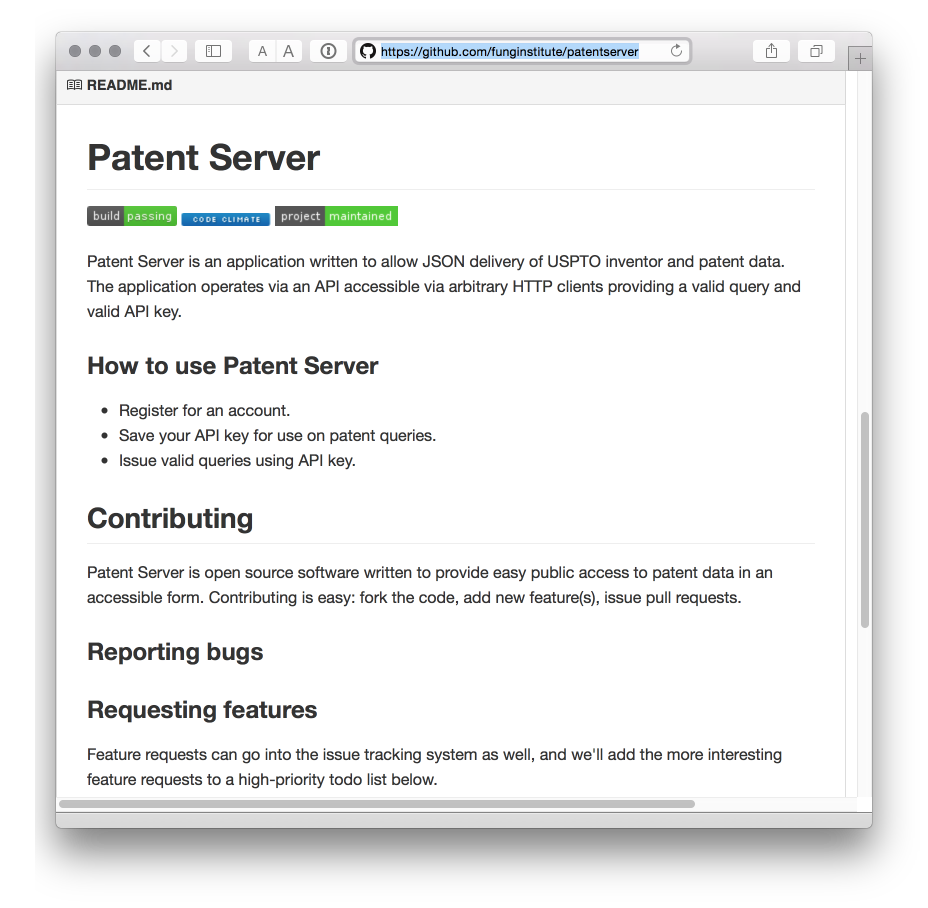
**13.** [**Python EPO OPS Client**](https://github.com/55minutes/python-epo-ops-client) **by Gsong**

A Python client for OPS access developed by Gsong and freely available on GitHub. Used in Patent2Net above.



**14.** [**Fung Institute Patent Server**](https://github.com/funginstitute/patentserver) **for USPTO data in JSON**

Researchers at the Fung Institute have also been active in developing open source resources for accessing and working with patent data. We highlight patentserver but it is worth checking out other resources in the repository such as [patentprocessor](https://github.com/funginstitute), a set of Python scripts for processing USPTO bulk download data.



## Round Up

That should be more than enough to get started with patent data and hopefully adds some useful pieces of information for long term patent professionals. Please feel welcome to add comments and suggestions on important free resources.

### Learn More

* Visit the [Github project page](http://poldham.github.io/)
* Access materials in the repository directly and add your own [here](https://github.com/poldham/opensource-patent-analytics)