Using Patent2Net to Obtain Patent Data

In this article we look at using open source Patent2Net software for accessing the European Patent Office Open Patent Service (OPS). Patent2Net can be used either through a Windows Client (in two versions) or using Python (across platforms). We will focus on the using the Windows Client as this is the simplest method.

One important strength of Patent2Net is that it handles the parsing of the patent data that is returned from OPS for you and allows for downloads of summary tables in .csv format as well as sections of patent documents. Because the XML or JSON that is returned from the OPS is quite complex, it is considerably easier to use Patent2Net than to parse the data for yourself in say R or Python.

On the other hand, it is important to emphasise that the OPS service is not intended for bulk downloads of patent data and the [EPO Fair Use Charter](https://www.epo.org/searching/free/fair-use.html) sets a weekly limit of 2.5GB per week for free downloads. Queries per IP address are limited to 10 search related actions per minute and the EPO also requests that data retrieval by robots be scheduled between 19.00 and 07.00 hours GMT or at weekends to limit the pressure on the service.

### Downloading Patent2Net

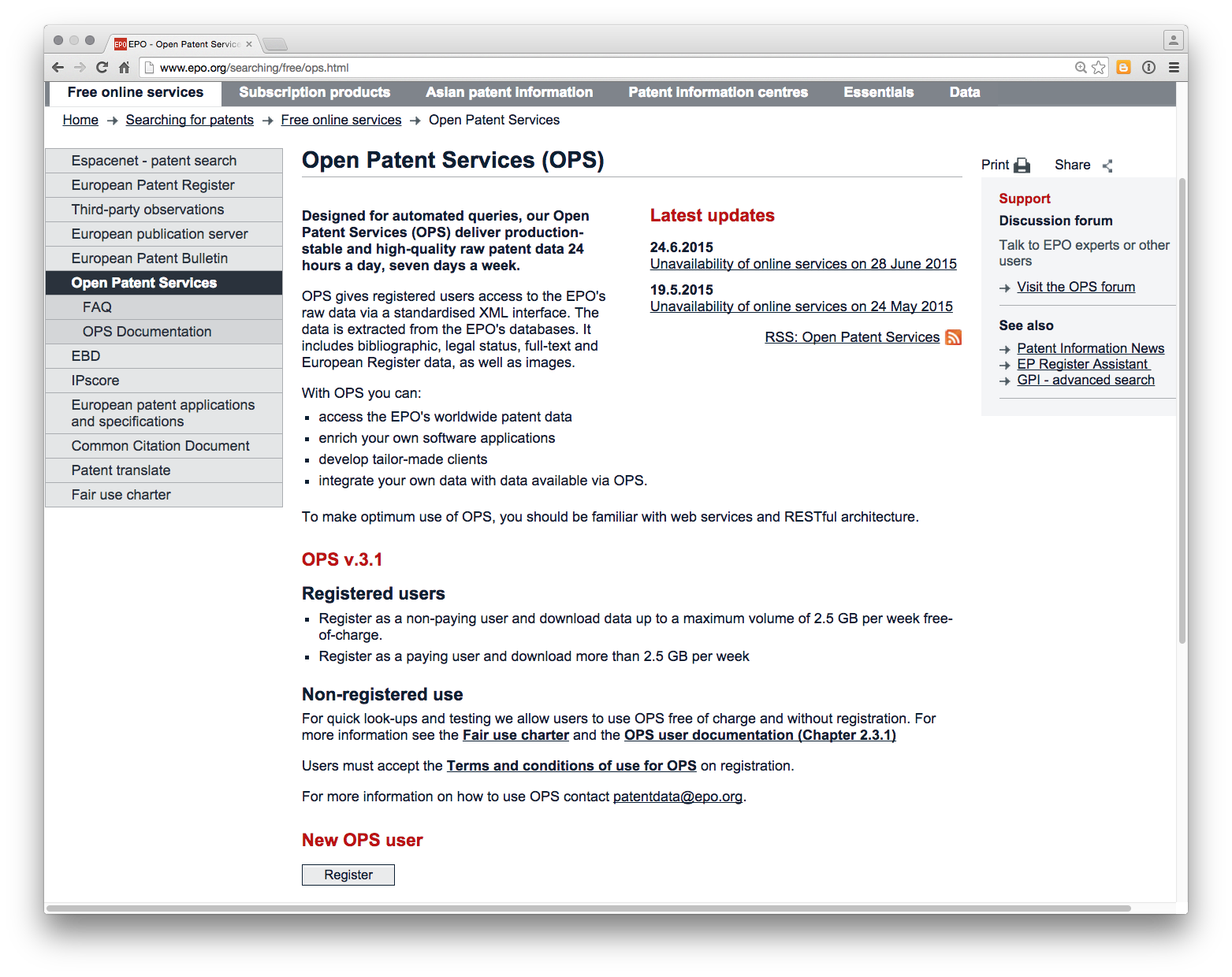
To install the Patent2Net Client in Windows (7 in this case) download the software from [here](http://patent2net.vlab4u.info/) by clicking on the word Crawler on the webpage or simply click [here](http://patent2net.vlab4u.info/P2N.zip) for a direct download. As an alternative visit the Github repository [here](https://github.com/Patent2net/Patent2Net) for the main Python installation and download a beta version of a Windows interface [here](https://github.com/Patent2net/Patent2Net/blob/master/ApplicationPatent2net.exe?raw=true). For licence information read this [file](https://github.com/Patent2net/Patent2Net/blob/master/Licence_CeCILL_V2.1-en.txt).

We will illustrate both approaches because the original Windows client helps us to understand the latest beta version.

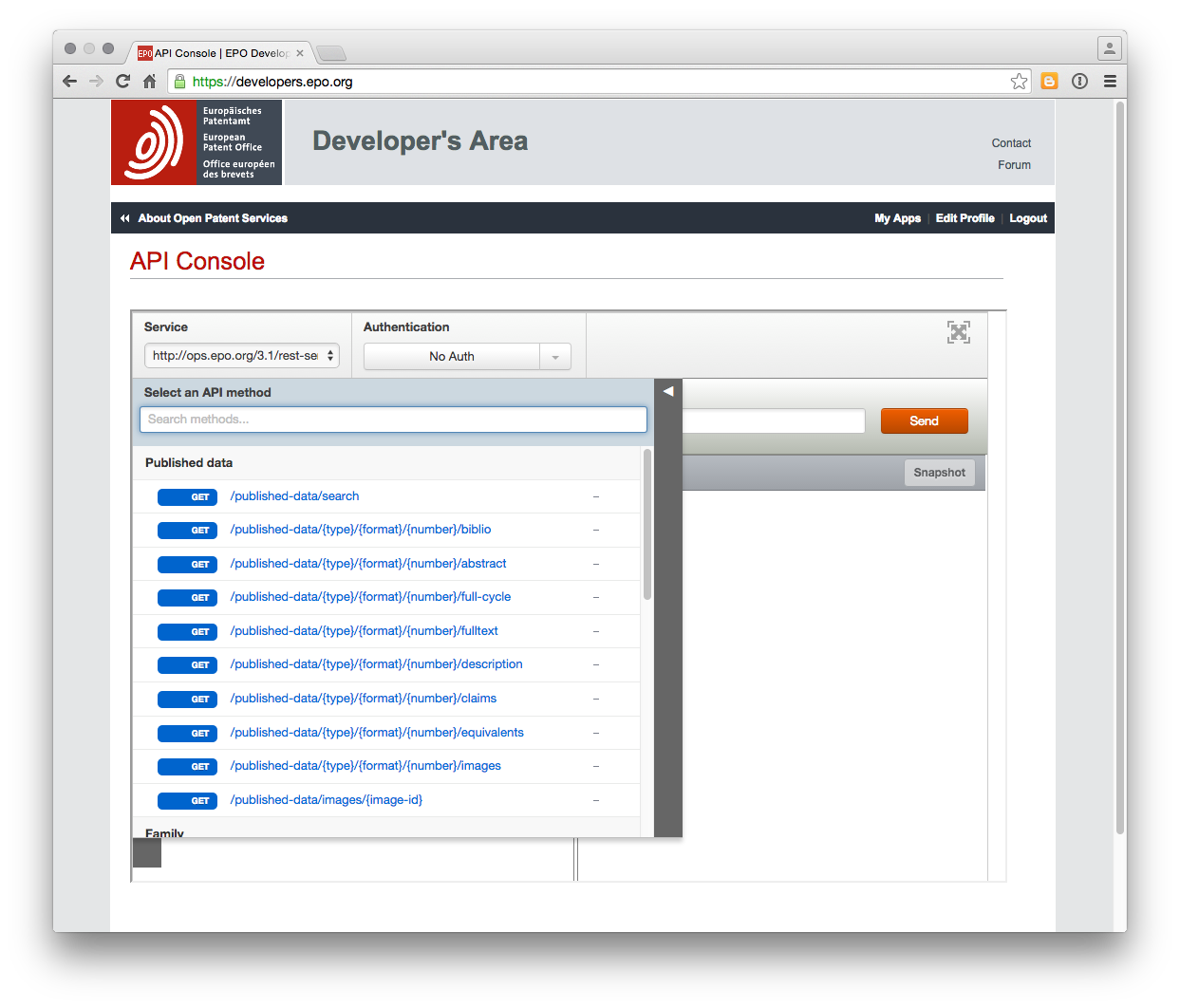
### Installing Patent2Net

We will repeat the instructions that are available in this file [here](http://patent2net.vlab4u.info/readme.txt) but add screenshots to make it as simple as possible.

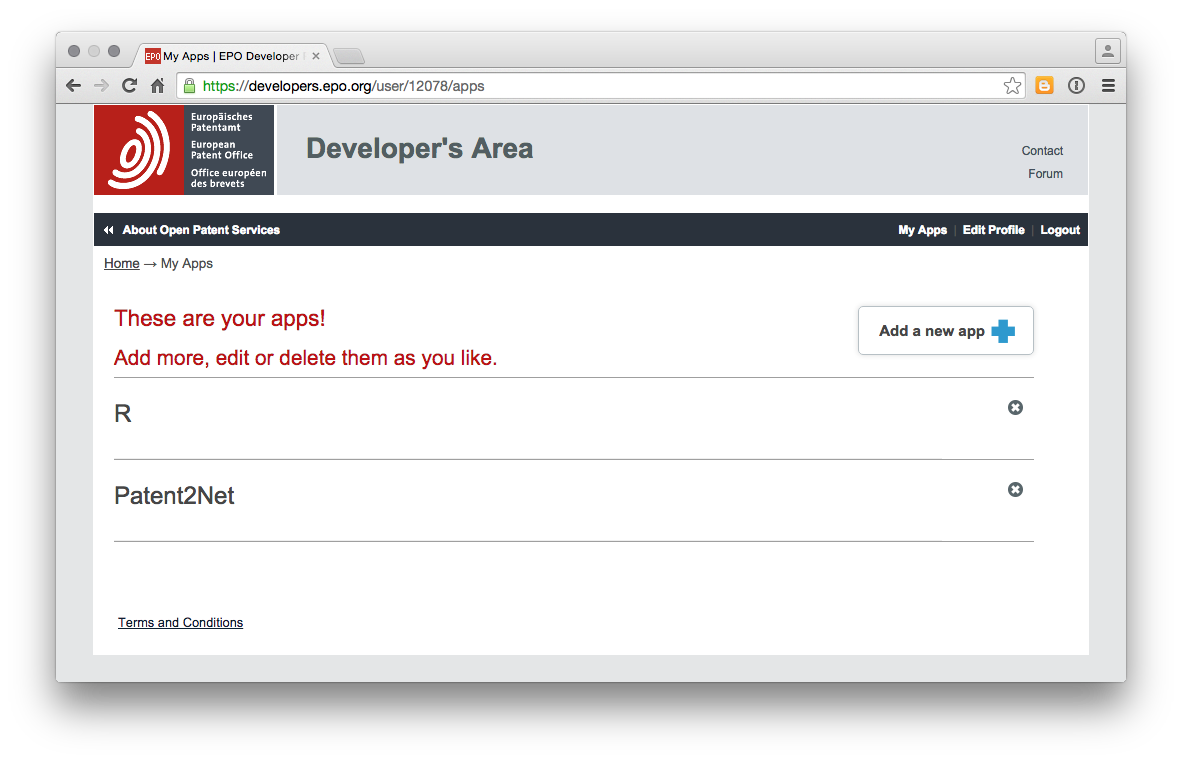
1. Unzip the P2N.zip file where you would like to locate the file (the Desktop is fine).
2. Register with OPS [here](http://www.epo.org/searching/free/ops.html) by selecting New OPS user.



1. Once you have registered you need to create a new App called Patent2Net. First go to My Apps.

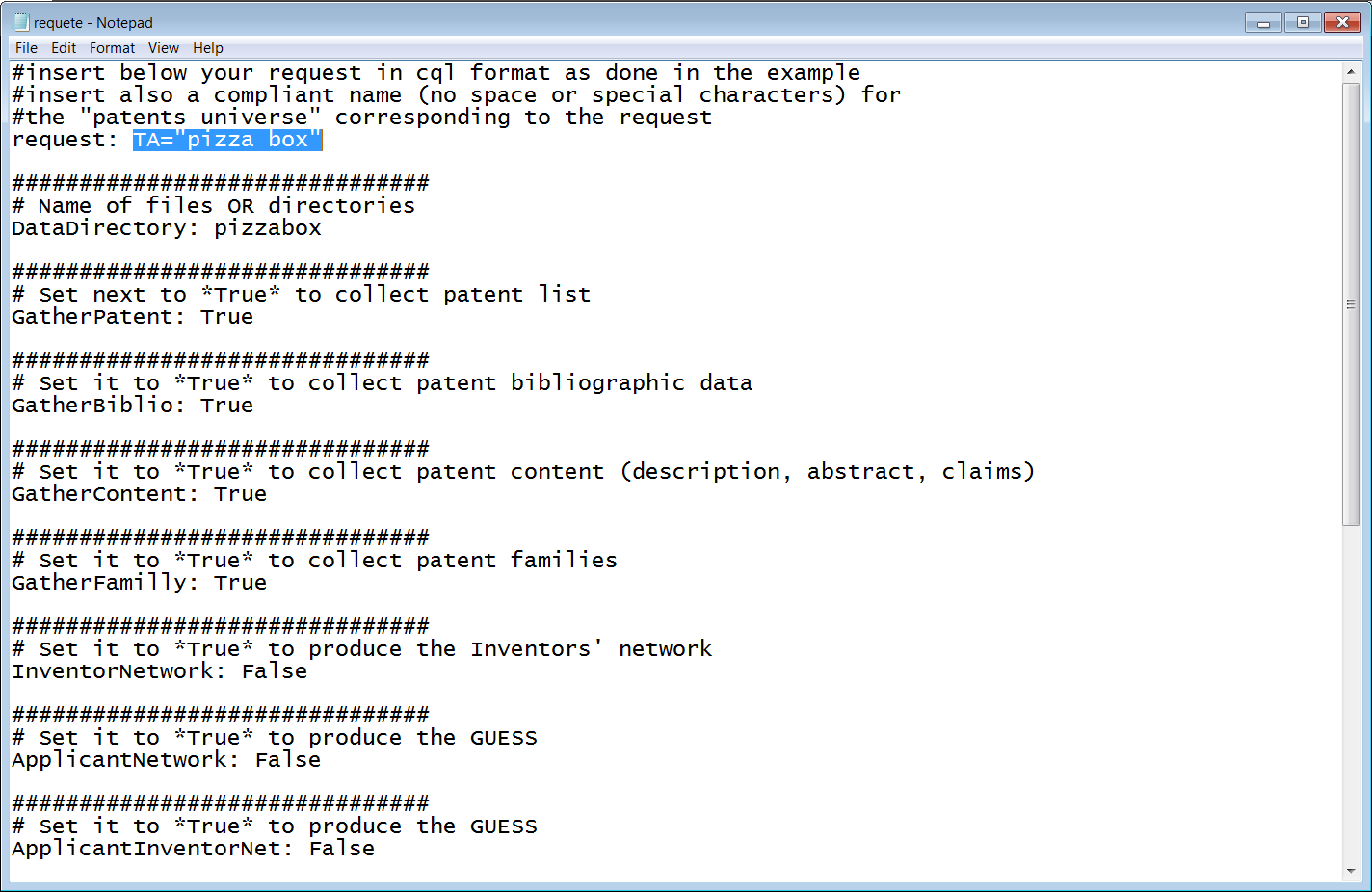


Then click on the large Add a new app icon and add an app called Patent2Net



1. Open the P2N folder wherever you have stored it and find the file called "requet.cql"

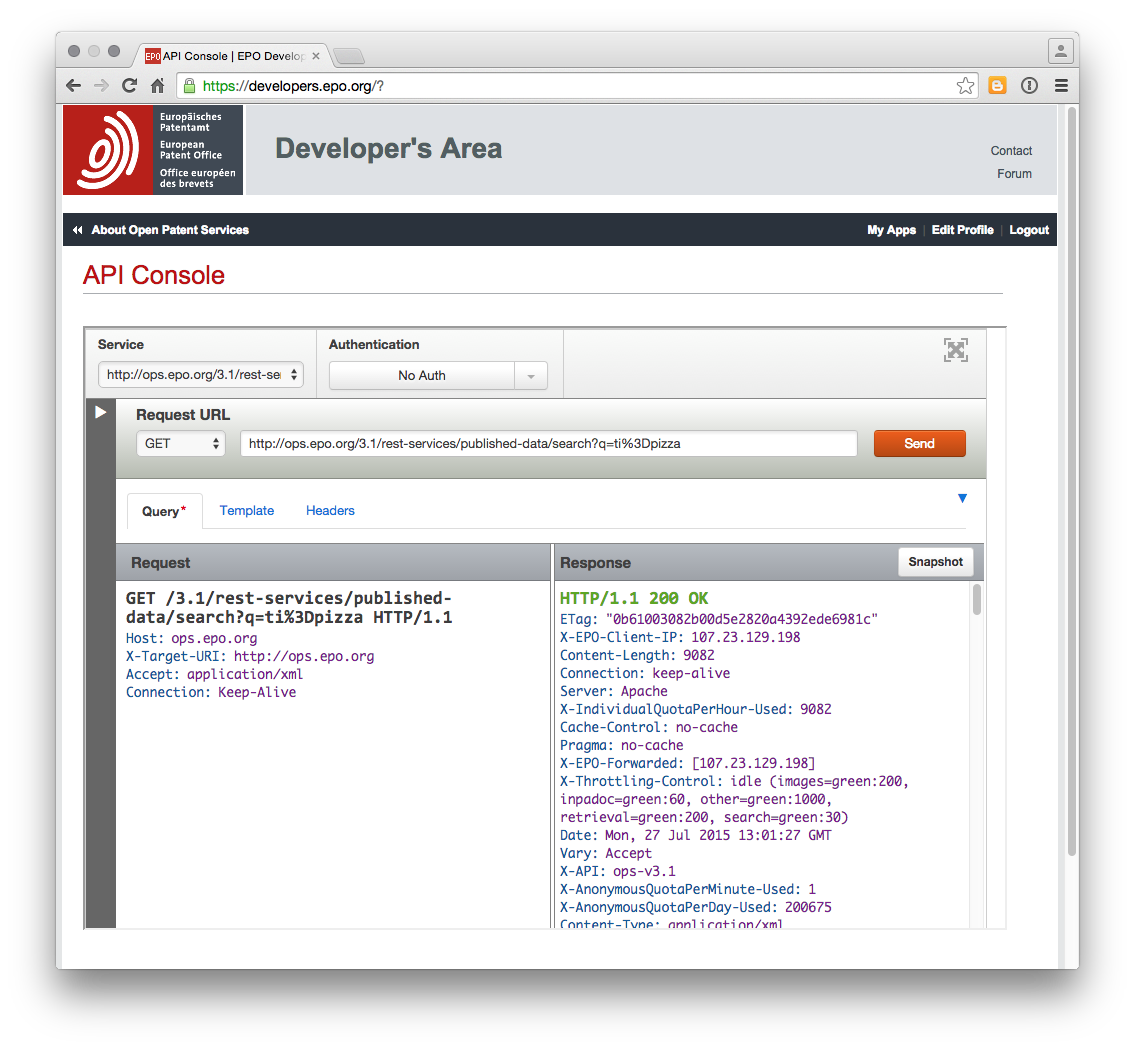
Change the request to something that you want. In this case we have used a title and abstract search TA="pizza box" because we love pizza boxes, particularly musical ones.



The default folder for downloads is entered in the DataDirectory field and is called Maroc3. In this case we have changed it to pizza box. Change this if you want to. Note that upon completion the data will be stored in a folder called DONNES.

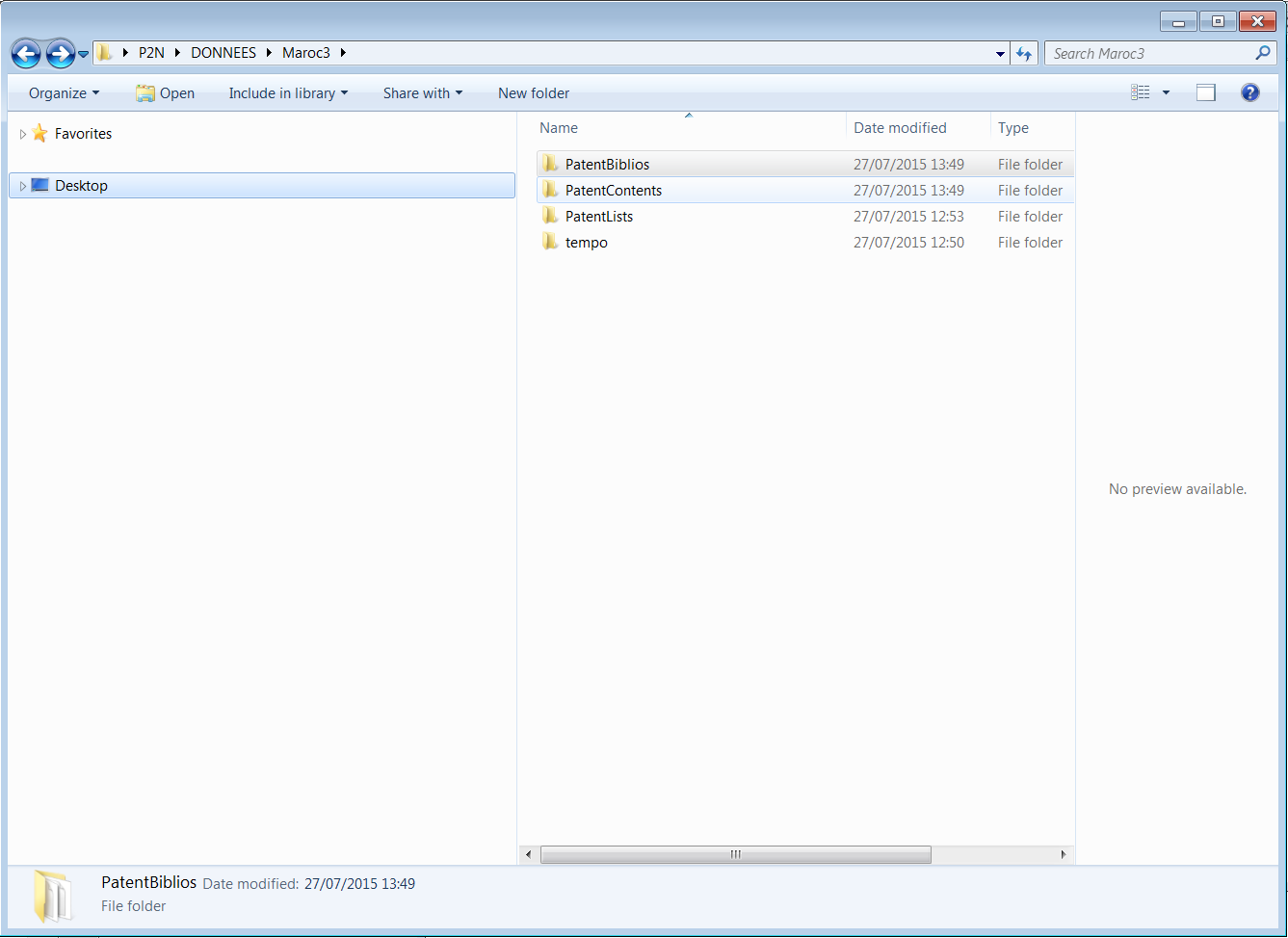
Leave the other settings as is but note that we could change from TRUE to FALSE etc as needed. The new beta interface simplifies this and is recommended (see below).

For other cql queries see the OPS Documentation [here](http://documents.epo.org/projects/babylon/eponet.nsf/0/7AF8F1D2B36F3056C1257C04002E0AD6/$File/OPS_v3.1_documentation_version_1.2.14_en.pdf) and try the Developers testing area [here](https://developers.epo.org/?). The Developers testing area is the best place to try out different types of queries and to assess the results.

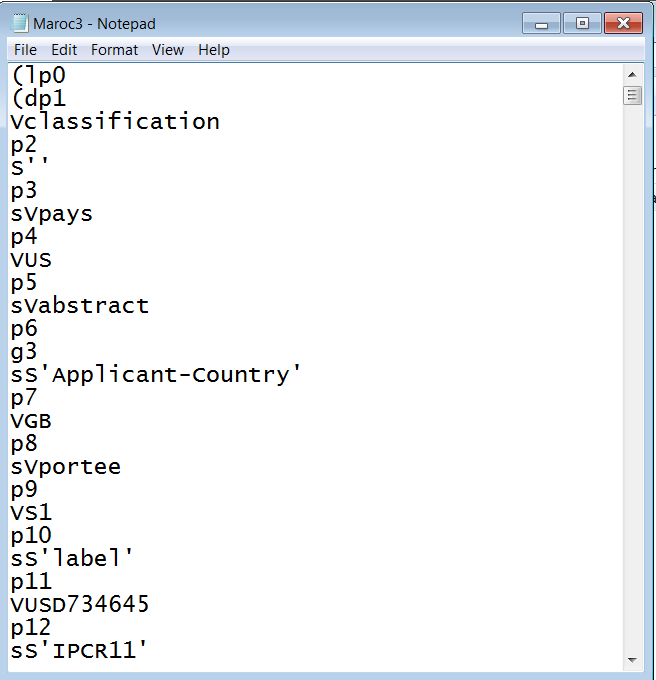


When testing Patent2Net we recommend using quite a restrictive query to begin with. For example, "pizza box" returns 203 results in the Titles and Abstracts while pizza returns +3000 and will attempt to download them. We managed to download 2,000 abstracts in this way. However, we would suggest using smaller queries until you have a better understanding of the amount of data that is downloaded relative to the weekly limit to avoid being blocked by the OPS system.

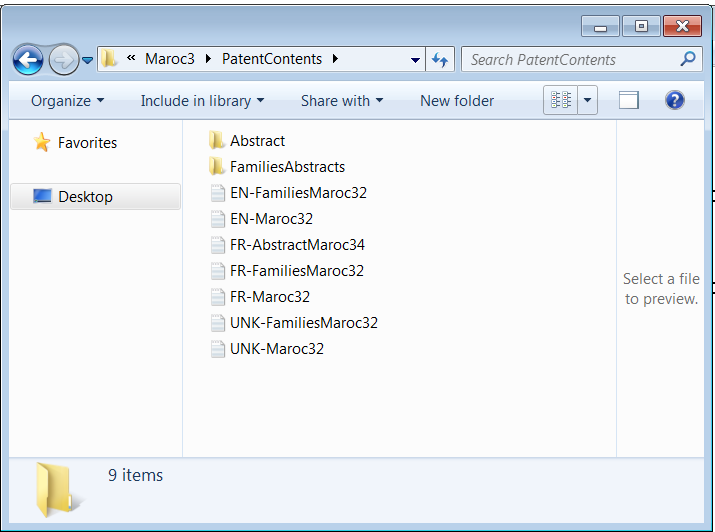
When we ran a query on pizza we ended up with a set of folders as follows.



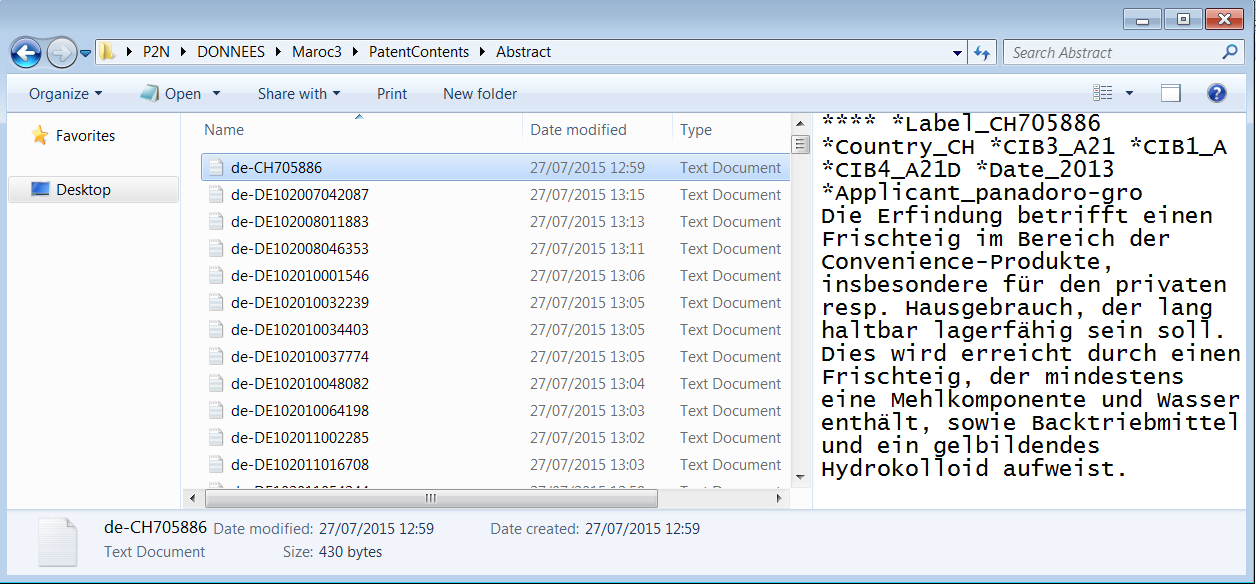
The patent biblios folder contained two files with a basic list that was difficult to interpret that looked like this.



A Families file was also created. It was a little difficult to interpret this data. In the PatentContents folder we saw these results (note that we retained the Maroc3 name for our pizza results when testing).



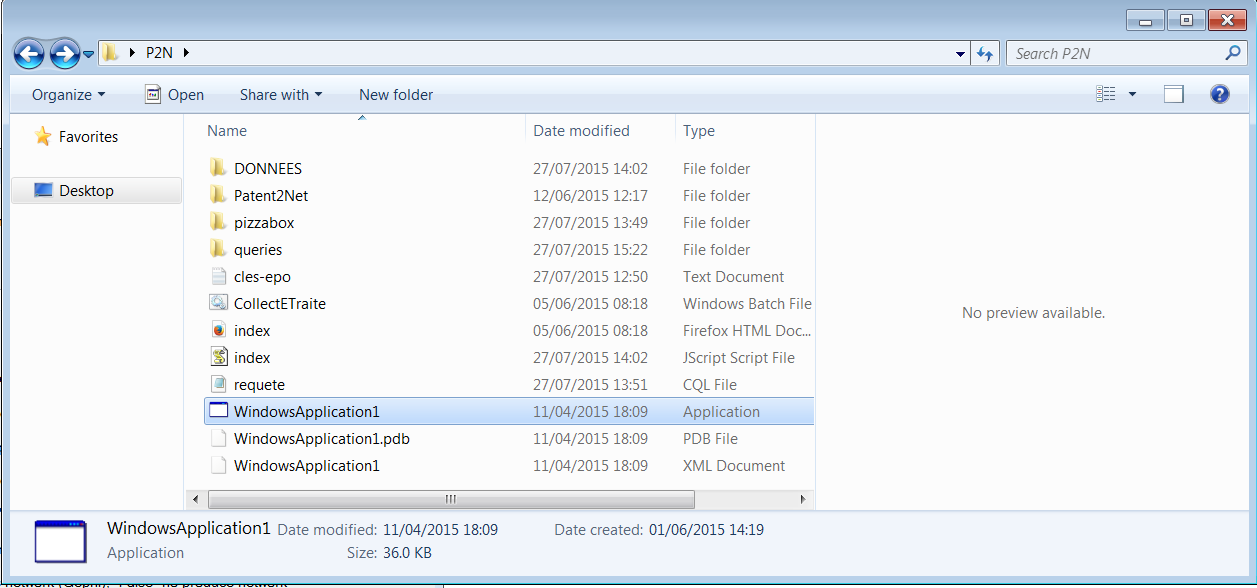
The abstracts folder contained the abstracts and some basic identifier details



What is interesting here is that relatively quickly we were able to obtain a significant number of abstracts (bearing in mind the constraints noted above) although we were less successful with the biblio information and piecing the data together would take time.

## The Beta Windows User Interface

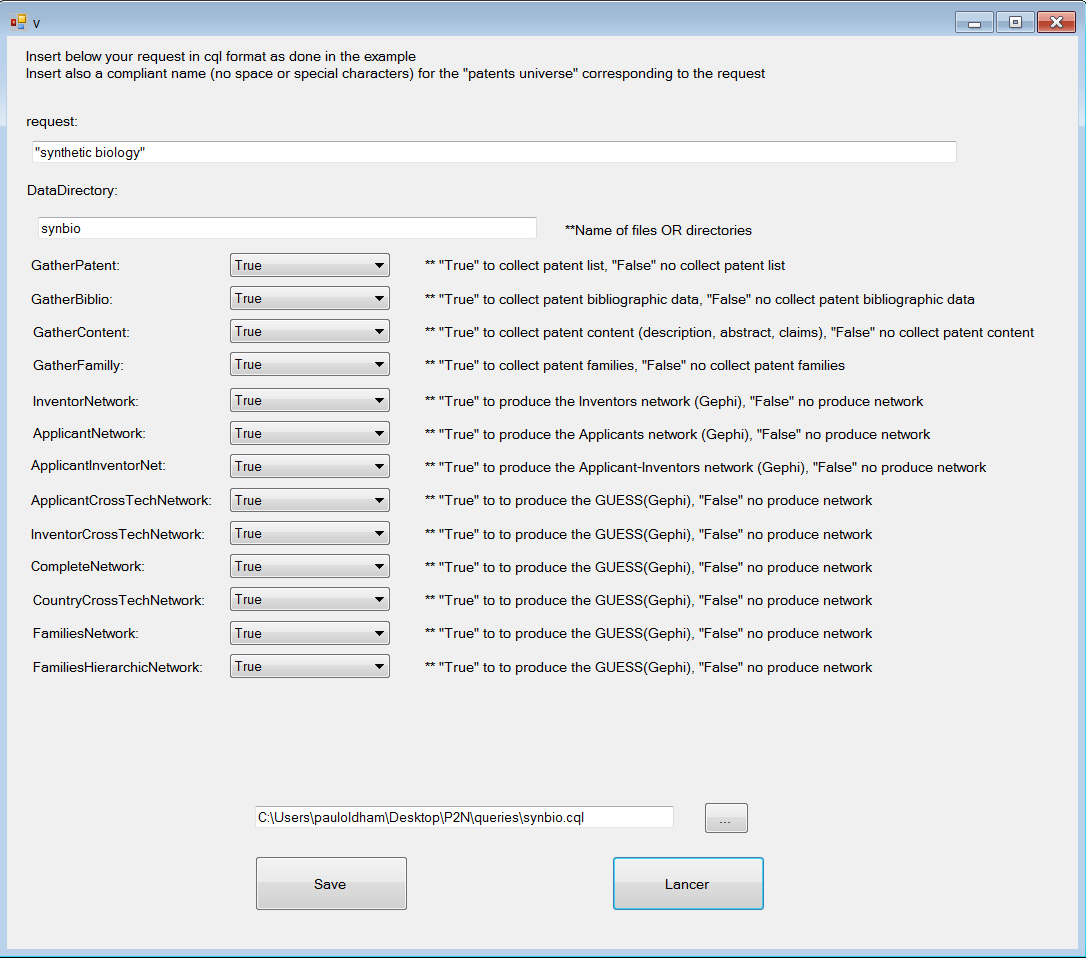
Following this experiment with the original interface we switch over to the new beta interface. This can be opened by double clicking WindowsApplication1 below



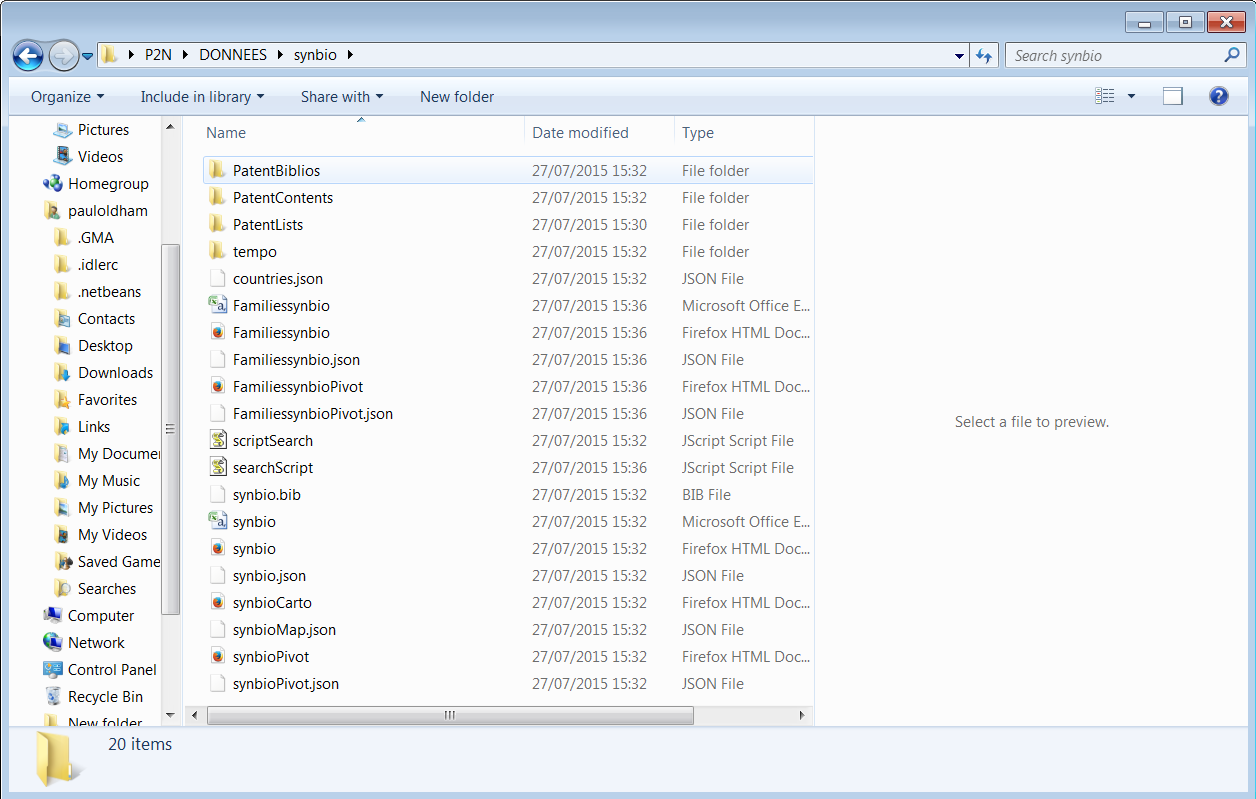
We will then see the following interface.

We entered the details of the query and then the directory to save the results. We also specified a queries folder to save the query.

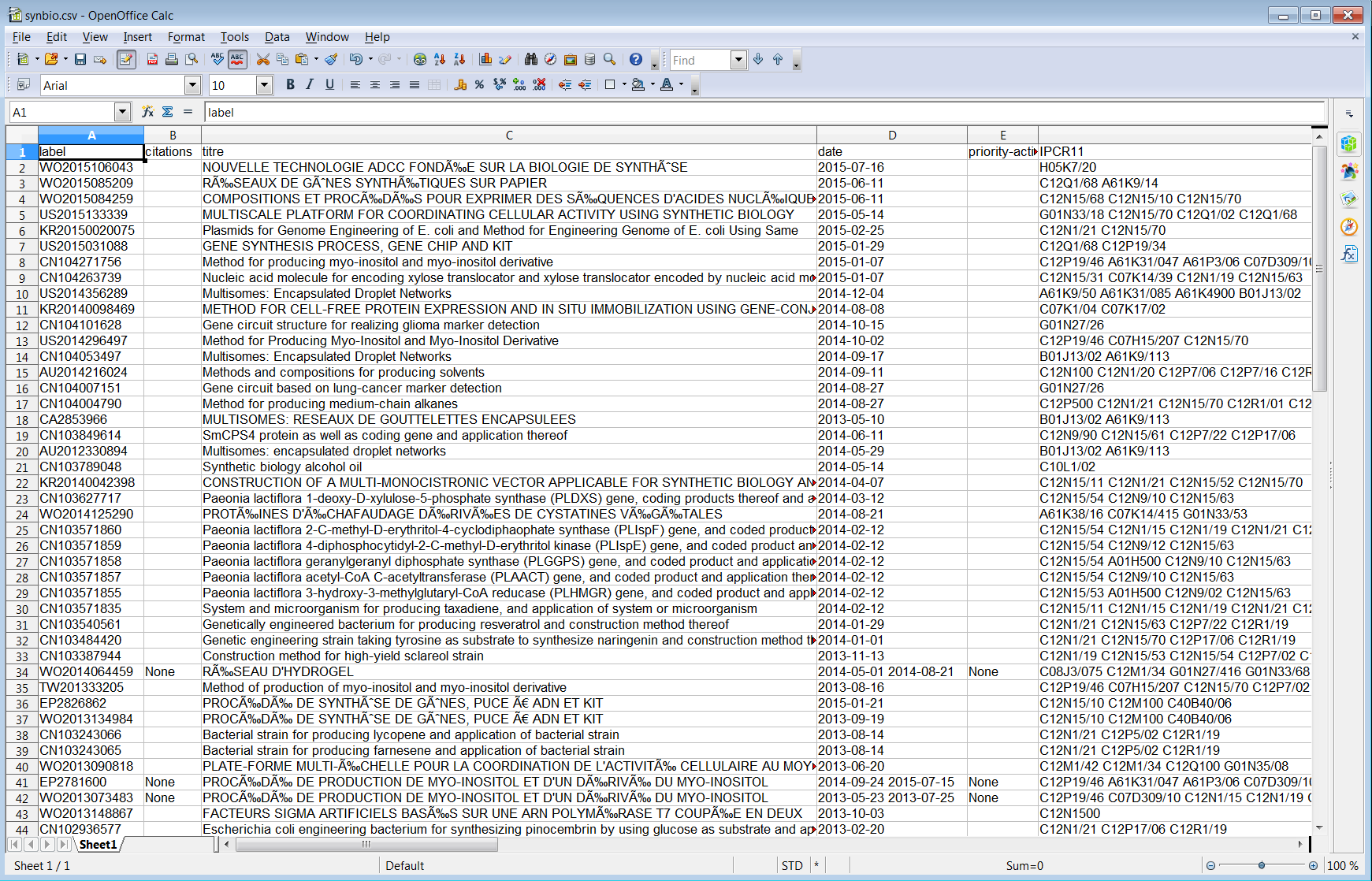
We ran in to problems that the interface was using the original pizza query as we had specified in the requete(request) .cql file rather than our new query. We found that we needed to enter the query into that file rather than our saved file (although that may originate from an error on our part)



The query then runs in the terminal and displays a range of messages (including error and problem messages). We should then see a folder in DONNES with the following contents



The data includes a .csv file containing a summary of results that can be opened in Excel or Open Office as below.



We also have access to a Families table (as Familiessynbio in this case). Note here that while the delimiter in the first table is a semi-colon, in the Families case it appears to be mixed and is likely to require further work to clean up.

The results folder also includes JSON files containing the data and links to an online javascript Pivot table library (see the online links when previewing a file).

In theory, although not in practice in this case, the interface will generate Gephi network files, although we did not see any files generated with a gexf or gephi extention.

### Round Up

In this article we have introduced the open source Patent2Net interface for retrieving patent data from the European Patent Office Open Patent Service. Patent2Net is particularly valuable because it takes the pain out of parsing the results from the OPS service and is particularly useful for gathering family and related data in a readily usable way.

The beta Windows interface for Patent2Net enjoys the advantage the no programming knowledge is required. However, bear in mind that the data is raw and will need to be cleaned in tools such as Open Refine, Python R or other tools.