# Text and Data Mining using the Crossref TDM Service

Technical Help Documentation

# How to get started downloading articles for TDM

This overview shows how to access content via the Crossref TDM service using [curl](https://curl.haxx.se/) as an example. It does not aim to cover the wide variety of TDM software applications available. The same principles will apply in the majority of cases.

Wiley uses the Crossref TDM service to enable bulk access to content for TDM purposes. This ensures that users are made aware of the terms and conditions under which content is made available. Our policy page [**here**](http://olabout.wiley.com/WileyCDA/Section/id-829771.html) provides more detail on this. In order to use the TDM application you must have a click-through client token which you get from Crossref. To get a token, you will need to log in to the Crossref service using your ORCID identifier.

In summary, the process is:

1. Obtain an ORCID
2. Accept the Wiley click-through license at Crossref
3. Obtain a Click-through client token from Crossref
4. Identify relevant articles
5. Download articles.

## Obtain an ORCID

An ORCID (Open Research and Contributor ID) is a non-proprietary code that uniquely identifies a researcher/author. More information can be obtained from <http://orcid.org/content/about-orcid>.

To get an ORCID, you must register at <https://orcid.org/register>, providing your name and email address. This is the form:

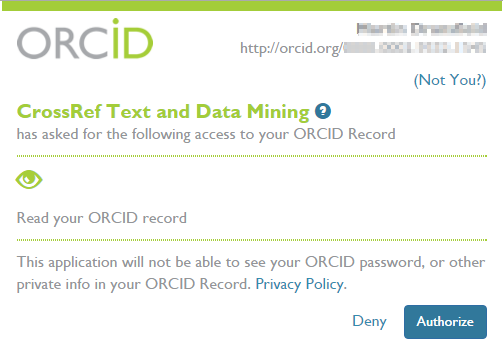


## Obtain a click-through client token

Once you have registered for an ORCID, go to [https://apps.Crossref.org/clickthrough/researchers](https://apps.crossref.org/clickthrough/researchers). You will see a page that looks like this:

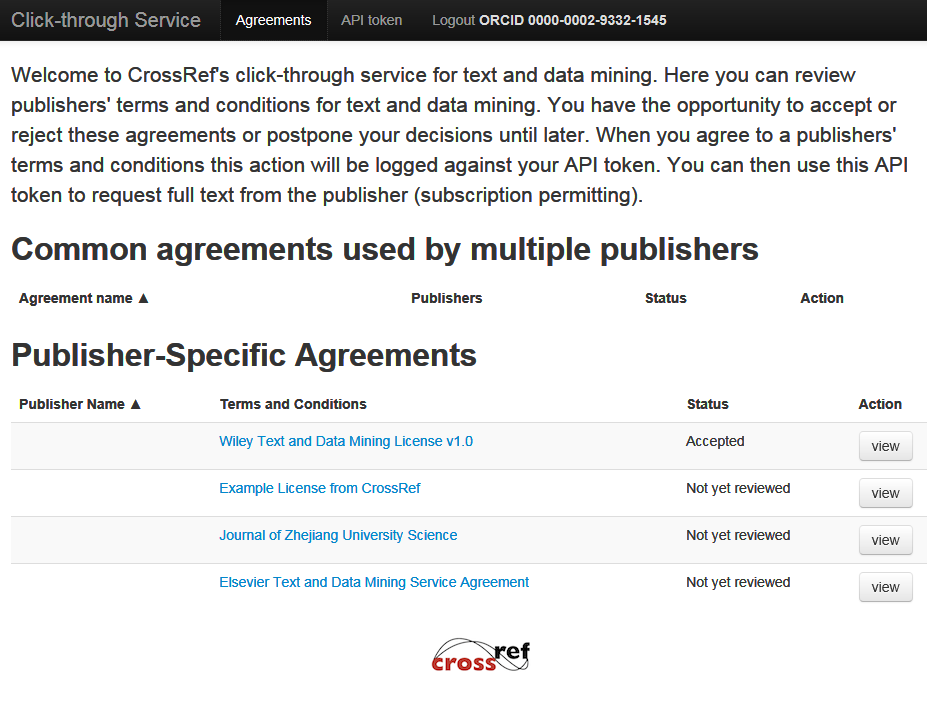


Click the “Log in with ORCID” button and you will be presented with a pop-up like this:

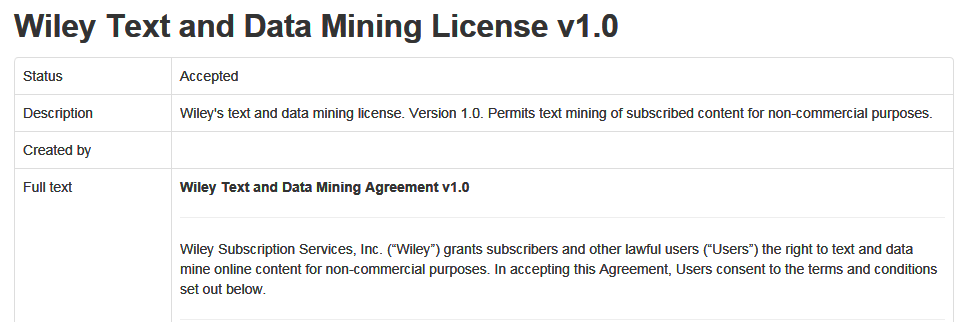


If you’re not logged in at ORCID, you will see a different page asking you to sign in or register.

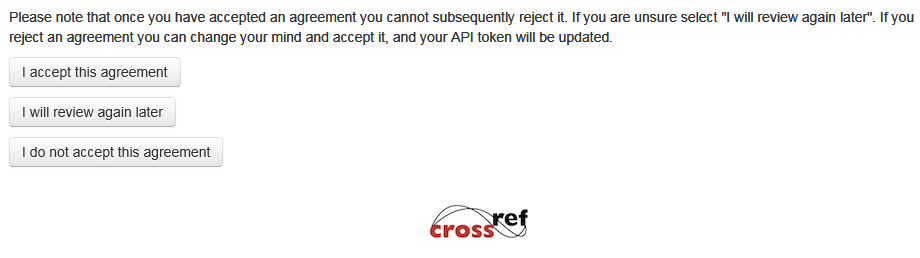
Once, you’re authenticated, you will see this page:



As you can see, this example shows that the Wiley Text and Data Mining License has been viewed and accepted by this user. Your view will initially show it as “Not yet reviewed” as the others shown here. To proceed, you need to click the “view” the Wiley license. The license looks like this:



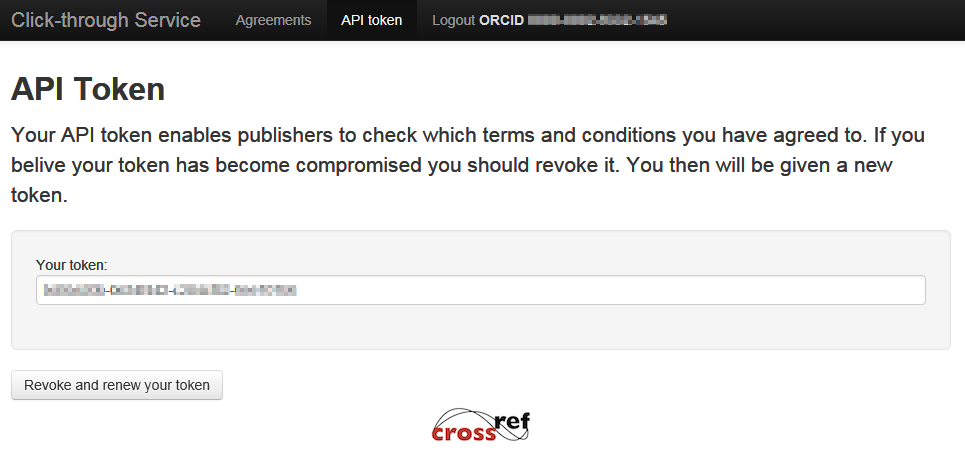
At this point, you need to read the complete license and decide whether you’re willing to accept the terms and conditions the license imposes. The bottom of the license page looks like this:



Assuming you are willing to accept the license, click the “I accept this agreement” button and the browser will reload the original list page in the form shown above (with the Status field showing as “Accepted”.)

## Obtain a Click-through client token from Crossref

From the Click-through service main page, click the API token link; you will end up on a page that looks like this:



The actual token should be kept confidential, so it’s only shown pixelated here. You need to pass your token as the value of a header in any TDM request you make.

## Identify articles you’re interested in analysing

You will need to identify the DOIs of relevant articles in order to access them using Crossref. This process is beyond the scope of this document.

## Download articles

Downloading articles for analysis breaks down into two steps.

### Finding the download URL

You should find the download URLs for a given DOI by checking with Crossref. There is a useful method described in the Crossref TDM documentation ([http://tdmsupport.Crossref.org/researchers/](http://tdmsupport.crossref.org/researchers/), 1 – Fetch the Metadata).

### Downloading an article

Obviously, there are lots of options available for downloading, ranging from freely-available tools through to custom-written downloaders. We will illustrate using [curl](https://curl.haxx.se/), a commonly-used tool.

The basic procedure is to make a request for the download URL of the article, passing your Clickthrough Client token as the value of the CR-Clickthrough-Client-Token header. As an example, we’ll download the following article:

Halikiopoulou, D. and Vlandas, T. (2015), The Rise of the Far Right in Debtor and Creditor European Countries: The Case of European Parliament Elections. The Political Quarterly, 86: 279–288. doi:10.1111/1467-923X.12168

Query Crossref metadata for the download link shows that the download URL for this article is: https://api.wiley.com/onlinelibrary/tdm/v1/articles/10.1111%2F1467-923X.12168.

The command line to use for curl is:

curl -L -H "CR-Clickthrough-Client-Token: xxxxxxxx-xxxxxxxx-xxxxxxxx-xxxxxxxx" \  
-D 12168-headers.txt \  
https://api.wiley.com/onlinelibrary/tdm/v1/articles/10.1111/1467-923X.12168 \  
-o 12168.pdf

The options used are:

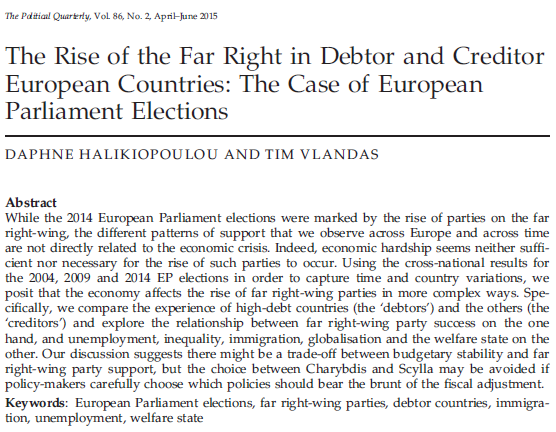
-L Follow redirects (this is important because our TDM application will redirect you on a successful request).

-H Add the CR-Clickthrough-Client-Token header to the request.

-D Write the response headers to a file (useful for debugging purposes)

-o Write the PDF to a file

Assuming you have a license to this article, the PDF should be downloaded and stored in the output file:



The Response headers file should look something like this:

|  |
| --- |
| HTTP/1.1 302 Redirect  Content-Type: application/atom+xml;charset=utf-8  Date: Tue, 08 Nov 2016 15:09:09 GMT  CR-TDM-Rate-Limit: 60  CR-TDM-Rate-Limit-Remaining: 56  CR-TDM-Rate-Limit-Reset: 1478617681657  Location: http://onlinelibrary.wiley.com/store/10.1111/1467-923X.12168/asset/poqu12168.pdf?v=1&t=iv9mtxan&s=aea530ef5ade929f1cbf9ed89ce1ae2205200f01  Connection: keep-alive  HTTP/1.1 200 OK  Date: Tue, 08 Nov 2016 15:09:08 GMT  Server: Apache  Last-Modified: Tue, 19 May 2015 08:50:38 GMT  ETag: "1e0d70-2c89c-5166b678a9380"  Accept-Ranges: bytes  Content-Length: 182428  Content-Type: application/pdf |

Points to note about the response headers are:

1. Our TDM application always responds initially with a redirect to a server dedicated to the serving of binary resources.
2. We implement rate-limiting as described in the Crossref documentation ([http://tdmsupport.Crossref.org/researchers/](http://tdmsupport.crossref.org/researchers/), 3 – Fetching the full text).

Unsuccessful requests should receive one of the errors which are explained in the following section.

# Errors

|  |  |  |
| --- | --- | --- |
| Status | Message | Remedial action |
| 400 | No CR-Clickthrough-Client-Token | All requests to the TDM application are expected to conform to the Crossref TDM protocol, documented at [http://clickthroughsupport.Crossref.org/click-through-service-for-researchers/](http://clickthroughsupport.crossref.org/click-through-service-for-researchers/). The request which received this error did not contain a CR‑Clickthrough‑Client‑Token header. |
| 403 | An error has occurred. It appears that you have not accepted Wiley’s license for text and data mining. You need to accept the Wiley license using the Crossref click-through service and obtain a valid Client API token. For more information, follow instructions for researchers: http://clickthroughsupport.Crossref.org/ | Our TDM application requires the acceptance of supplementary terms and conditions at Crossref’s clickthrough service. |
| An error has occurred. It appears that you or your institution/organisation does not have access to the content that you have requested (e.g. through an existing subscription). Please check that you are requesting content for which you have full-text access. If your institution/organisation subscribes to the content, you must make the request from their network. For more information, follow the instructions for researchers: http://clickthroughsupport.Crossref.org/ | You must have access to the content you wish to download, whether through an institutional license, or because the content is open access. |
| 404 | {  "fault": {  "faultstring": "Classification failed for host: api.wiley.com url: /onlinelibrary/tdm/v1/article/…",  "detail": {  "code": "CLASSIFICATION\_FAILED"  }  }  } | Our TDM application did not recognise a part of the URI requested. This message means that there was something wrong with the path part of the URI before the DOI. Our TDM application responds to requests to URIs of the form:  <http://api.wiley.com/onlinelibrary/tdm/v1/articles/>...  In the example message, the error is “/article” rather than “/articles”. Please correct and try again. If you get this response to the correct URI, please contact Customer Services. |
| An error has occurred. The DOI you have requested cannot be found. The DOI may not exist, the DOI may not be available through the Crossref click-through service or it may belong to a publisher other than Wiley. Please check that the DOI can be found in Wiley Online Library and try again. For more information, follow instructions for researchers: http://clickthroughsupport.Crossref.org/ | Our TDM application did not receive the DOI in the request. Usually this happens as a result of an error when constructing a TDM URI rather than obtaining it from Crossref. If you constructed this URI, you can test whether the DOI is valid using [http://dx.doi.org/{DOI}](http://dx.doi.org/%7bDOI%7d). If it is valid, or you obtained the URI from Crossref, please report to Customer Services. |
| 429 | Too many requests | Our TDM application implements rate-limiting in the way described in [http://tdmsupport.Crossref.org/researchers/](http://tdmsupport.crossref.org/researchers/) (Section 3 – Fetching the full text). You should see the value of the CR‑TDM‑Rate‑Limit‑Remaining header in the response have a value of zero (0) when this error occurs. To continue downloading you will have to wait until the time indicated in the CR‑TDM‑Rate‑Limit‑Reset header. If you receive this response and the value of CR‑TDM‑Rate‑Limit‑Remaining is not zero please send more details to Customer Services. |