A Content Negotiation Update

Martin Fenner

Blog post published April 28, 2017 in *DataCite Blog - DataCite*
https://doi.org/10.5438/0000-01qj

Abstract

While it is a best practice for DOIs (expressed as URL) to send the user to the landing page for that resource [@https://doi.org/10.7717/peerj-cs.1; @https://10.1101/097196], sometimes we want something else: **metadata** , e.g. to generate a citation, or to go to the **content** itself. The easiest way to do that is to use DOI content negotiation.

Keywords

Information Systems (Subfield), DOI, Metadata, THOR

Recommended citation

Fenner, M. (2017, April 28). A Content Negotiation Update. *DataCite Blog - DataCite*. https://doi.org/10.5438/0000-01qj

Copyright

Copyright © Martin Fenner 2017. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

While it is a best practice for DOIs (expressed as URL) to send the user to the landing page for that resource , sometimes we want something else: **metadata**, e.g. to generate a citation, or to go to the **content** itself. The easiest way to do that is to use DOI content negotiation.

In this blog post we will give an introduction to DOI content negotation, describe some of the issues we identified with our current implementation, and announce a major service update addressing these issues and launching in two weeks.

DOI Content Negotiation uses the HTTP **Accept** header together with the DOI expressed as URL. This is best explained by an example, using a recent post from this blog:

```
curl -LH "Accept: application/x-bibtex" https://doi.org/10.5438/0000-0C2G
```

This command returns BibTeX metadata for this DOI:

```
@misc{https://doi.org/10.5438/0000-0C2G,
  doi = {10.5438/0000-0C2G},
  url = {https://doi.org/10.5438/0000-0C2G},
  author = {Cruse, Patricia},
  publisher = {DataCite},
  title = {The OI Project gets underway planning an Open Organization Identifier Registry}
  year = {2017}
}
```

Or maybe you want to go directly to the content, in this case the JATS XML for the blog post :

```
curl -LH "Accept: application/xml" https://doi.org/10.5438/0000-0C2G
```

Sometimes it is easier to use a normal link instead, e.g. in a web browser where sending HTTP headers is a bit involved.

```
curl -L https://data.datacite.org/application/x-bibtex/10.5438/0000-0C2G
```

For this to work you have to go directly to the DataCite content negotiation service at **data.datacite.org** instead of using the **doi.org** DOI proxy.

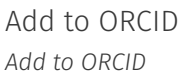
The most popular use of DOI content negotiation is citation formatting in any of the 1000s of citation styles provided by the Citation Style Language (CSL) project. For this we combine DOI content negotiation, which generates a JSON file from the metadata in a format that CSL understands (Citeproc JSON), with our DOI citation formatting service, which generates the formatted citation:

```
curl -LH "Accept: text/x-bibliography; style=apa" https://doi.org/10.5438/0000-0C2G
```

This command returns a citation formatted in APA style:

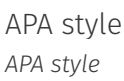
```
Cruse, P. (2017). The OI Project gets underway planning an Open Organization Identifier Re
```

There is an easier way to get this information from DataCite if you don't feel comfortable using the command line. First find the DOI you want to cite in DataCite Search, e.g. https://search.datacite.org/works/10.5438/0000-0C2G in this case.



Add to ORCID

Then click on the **Cite** button and select on of several popular citation styles. We again pick the APA style:



APA style

This is the same formatted citation we saw earlier, as DataCite Search is using the content negotiation and citation formatting services in the background. These services have been around for several years, and we work together with other DOI registration agencies on this. The citation service in DataCite Search for example is based on Crossref open source code originally written for Crossref Metadata Search.

Upcoming Changes

There is still room for improvement of the content negotiation service. As part of work in the EC-funded THOR project we took a closer look at what you can and can't do with DOI content negotiation. We identified a number of gaps, and will address them by launching an updated content negotiation service in May.

Many content types only supported with required metadata

The current DataCite content negotiation only supports the required metadata (identifier, creator, title, publisher, publicationYear, resourceTypeGeneral) for most content types. This means that even for something as straightforward as BibTeX DataCite is not converting the supported optional metadata such as `description`, `subject/keywords`, `version` or `license`. And for rich metadata such as RDF, a long list of properties is not converted.

Limited person name parsing

Person names are very difficult to parse, in particular when not using the optional properties `givenName` and `familyName` introduced in DataCite Schema 4.0 last year . But proper parsing is important for citations, as different citation styles use different formatting rules for names. And BibTex expects a particular format that is hard to support without proper person name parsing.

Better support for consistent metadata across DOI registration agencies

Although content negotiation is supported by several DOI registration agencies, there is only limited support for common metadata beyond the basic citation metadata. This makes it harder than it should it to combine metadata records from different DOI registration agencies. The

current RDF support is limited, and no DOI registration agency is offering content negotiation that converts metadata from another registration agency.

Support for additional content types

The list of supported content types could be much longer, from emerging metadata standards such as schema.org or scholix to community-specific standards such as DDI for social sciences, DATS for the life sciences , or codemeta for software.

Unrecognized content types raise an error

When the DataCite DOI content negotiation encounters an unrecognized content type, it returns a **406 Not Acceptable** error. This means that it is impossible to implement content negotiation downstream at the data center, as all content negotiation requests will be redirected to DataCite. Data centers can register their custom content types in the DataCite MDS using the `media` API endpoint, but this feature is unfortunately not widely used.

DataCite has rewritten most of the code of the content negotiation service, and has extracted out the metadata conversion into a standalone library that can also be used locally via the command line. The new service is currently in final testing at https://data.test.datacite.org, and we plan to launch this into production ion May 9th. We have worked on all the limitations listed above, and some of the highlights include:

- extensive support for schema.org/JSON-LD as a common metadata standard independent of a particular DOI registration agency
- extensive RDF support as XML or Turtle via schema.org metadata
- tested with DOIs from DataCite and Crossref, supports conversion of Crossref metadata to DataCite metadata
- BibTeX and RIS files have proper file extension (.bib and .ris) for easier import into other applications.
- new content type `codemeta`, other content types will be added in the coming months based on user feedback
- available as open source software via Github and Docker Hub.

There is one breaking change that users should be aware of: the content type `text/html` will no longer be supported, as the DOI proxy always forwards requests with this content type to the URL registered in the handle system. Users who want to see a HTML representation of the DOI metadata should go to DataCite Search, i.e. https://search.datacite.org/works/10.5438/0000-0C2G instead of https://data.datacite.org/10.5438/0000-0C2G. We will be improving the information shown for a single DOI in DataCite Search in the coming months, and appreciate user feedback.

References

The post A Content Negotiation Update appeared first on DataCite.