OAI-PMH Service Updates

Richard Hallett

Blog post published November 25, 2019 in *DataCite Blog - DataCite*
https://doi.org/10.5438/ppth-pz62

Abstract

Our OAI-PMH service is one of the common ways we offer to harvest our public metadata, and we are launching a new version this Wednesday. This technology refresh allows us to continue supporting the OAI-PMH service. For the most part, there is no functional change, we adhere to the OAI-PMH standards and have attempted to keep the service as backward compatible as possible.

Keywords

Information Systems (Subfield), Infrastructure

Recommended citation

Hallett, R. (2019, November 25). OAI-PMH Service Updates. *DataCite Blog - DataCite*. https://doi.org/10.5438/ppth-pz62

Copyright

Copyright © Richard Hallett 2019. 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.

Our OAI-PMH service is one of the common ways we offer to harvest our public metadata, and we are launching a new version this Wednesday. This technology refresh allows us to continue supporting the OAI-PMH service. For the most part, there is no functional change, we adhere to the OAI-PMH standards and have attempted to keep the service as backward compatible as possible. The main change in the new service is that it uses our REST API instead of directly integrating with our Solr search index. The REST API uses our newer Elasticsearch-based search index under the hood, and with OAI-PMH being the last service still depending on Solr, this allows us to retire our Solr search index, completing the transition to Elasticsearch that started in 2018.

As with all our projects, the code is open source and available via http://github.com/datacite/viringo. It is a Python-based Flask application using an existing OAI-PMH library with a few tweaks for using it against our services.

Please note that there are a few minor differences:

Direct Solr queries no longer supported

Originally we supported using the Solr query syntax through OAI-PMH, this allowed various custom searches. However, as we are moving away from Solr this will no longer be possible. Custom queries are still possible and instead, we will support the same syntax that is used by the REST API, as documented at https://support.datacite.org/docs/api-queries. The way of sending the queries will remain the same in the format of a base64-encoded query.

OAI Identifiers have changed

The identifier that was returned in the attribute of a record used to be an internal database identifier, this could then be used to retrieve the record. However, this identifier was never a persistent identifier nor did it have any connection to the metadata. In the new service, we are using the DOI itself as the identifier and this now aligns better with our other APIs.

Resumption tokens are a different format

Due to the fact that we are no longer using Solr, but instead using our REST API built on Elasticsearch, we had to change the format of the resumption tokens. Any old resumption tokens will no longer work and fresh ones should be obtained.

In summary, we hope this change will allow us to continue supporting OAI-PMH and bring it in line with all the other work we are doing to enable metadata harvesting and discoverability. The documentation for the OAI-PMH service is available at https://support.datacite.org/docs/datacite-oai-pmh, feel free to contact us at mailto:support@datacite.org if you have questions, comments, or concerns.

The post OAI-PMH Service Updates appeared first on DataCite.