



KB

Koninklijke Bibliotheek

Quickstart for searching & harvesting ANP

Searching via (J)SRU

To harvest documents that have specific characteristics, queries may be performed using the SRU protocol. "SRU" stands for "Search and retrieval via URL". More information on this standard can be found at <http://www.loc.gov/standards/sru/>

JSRU is a Java implementation of the SRU protocol at the KB. The base URL for searching via SRU is: <http://jsru.kb.nl/sru/sru>

An example of a query is:

<http://jsru.kb.nl/sru/sru?version=1.2&maximumRecords=50&operation=searchRetrieve&startRecord=1&recordSchema=dc&x-collection=ANP&x-facets=&query=%28date%20within%20%2201-01-1940%2031-12-1990%22%29&x-fields=content>

This results in descriptions of the documents with date between 01-01-1940 and 31-12-1990. The "x-fields=content" parameter gives the content of the documents as well (the OCR).

In the results the metadata of the documents can be found, e.g.

```
- <srw:records>
- <srw:record>
  <srw:recordPacking>xml</srw:recordPacking>
  <srw:recordSchema>info:srw/schema/1/dc-v1.1</srw:recordSchema>
- <srw:recordData>
  <dc:identifier>http://resolver.kb.nl/resolve?urn=anp:1946:05:06:10:mpeg21</dc:identifier>
  <dc:type>Text</dc:type>
  <content>Lombaers Op last van don directeur-generaal voor bijzondere rechtspleging is Arnold Meyer van het
    bewaringskamp te Tilburg ove gebracht naar het huis van bewaring in Den Bosch. De behandeling van
    zijn zaak voor het bijzonder gerechtshof kK van Den Bosch kan binnenkort worden tegemoetgezien.(
    anp)</content>
  <dc:date>1946/05/06 00:00:00</dc:date>
  <dc:title>ANP Nieuwsbericht - 06-05-1946 - 10 </dc:title>
  <dc:language>dut</dc:language>
</srw:recordData>
<srw:recordPosition>1</srw:recordPosition>
</srw:record>
```

To obtain a description of the entire document, use the value of the dc:identifier, which is a resolver URL (e.g. <http://resolver.kb.nl/resolve?urn=anp:1946:05:06:10:mpeg21>), to retrieve the MPEG21-DIDL file of the document.

Mpeg21-DIDL is a standard to describe the structure of compound digital objects (more on this standard: <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>).

All files that belong to the document (scans, pdf, fulltext etc.) are listed as component elements in the mpeg21 file.

To gain more insight in how SRU-queries for other datasets within the KB-infrastructure can be constructed, it can also be useful to read the [Quickstart SGD \(pdf - in Dutch\)](#) and the [Quickstart EDBO \(pdf - in Dutch\)](#)

Harvesting using the OAI-PMH protocol

Harvesting activities should be performed outside Dutch office hours.

Documents can also be harvested using the OAI/PMH protocol. "OAI/PMH" stands for "Open Archives Initiative Protocol for metadata Harvesting". OAI/PMH is typically used for incrementally harvesting entire collections. General information about the protocol may be retrieved at: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

The OAI base-URL for harvesting from KB collections is: is <http://services.kb.nl/mdo/oai>. For example, to retrieve a list of identifiers use the ListIdentifiers command with set=anp, use the following url: <http://services.kb.nl/mdo/oai?verb=ListIdentifiers&set=anp&metadataPrefix=didl>

To get the next list of URLs add the resumptiontoken in the response as parameter in the URL. With the "from" parameter you can add a datestamp to selectively harvest records which were added to the collection since a particular date.

Individual metadata records can be obtained by using these URLs in a GetRecord command: <http://services.kb.nl/mdo/oai?verb=GetRecord&identifier=> To get the next list of URLs add the resumptiontoken in the response as parameter in the URL. With the "from" parameter you can add a datestamp to selectively harvest records which were added to the collection since a particular date.

Individual metadata records can be obtained by using these URLs in a GetRecord command: <http://services.kb.nl/mdo/oai?verb=GetRecord&identifier=anp:anp:1937:10:01:2:mpeg21&metadataPrefix=didl>

Getting the files

To get to the files (fulltext xml, images) use the URLs that are listed in the MPEG21 files, eg

- <http://resolver.kb.nl/resolve?urn=anp:1937:10:01:2:mpeg21:image>
- <http://resolver.kb.nl/resolve?urn=anp:1937:10:01:2:mpeg21:ocr>

Conditions for reuse & licensing

1. ANP has made the objects (JPGs, OCRs, ALTOs) in this set available under a CC-BY-NC 3.0 license. See <http://creativecommons.org/licenses/by-nc/3.0/deed.en> for more details.



Please use the following format when quoting the name, source and license of the object:

*Algemeen Nederlands Persbureau (ANP). Source: <persistent URL to JPG>
– CC-BY-NC - <http://creativecommons.org/licenses/by-nc/3.0/deed.en>*

example of persistent URL to JPG <http://resolver.kb.nl/resolve?urn=anp:1938:10:01:2:mpeg21:image>

For commercial reuse of the objects you'll need written permission from ANP in advance. Please contact business@anp.nl

2. The Koninklijke Bibliotheek has waived the copyright on the **metadata** in this set. These files are therefore available under the [CC0 1.0 Universal declaration](http://creativecommons.org/licenses/by-nc/3.0/deed.en).



Questions or remarks about the ANP set?

Please send a mail to dataservices@kb.nl

Last update : 18-12-2012