

## Snelstart zoeken & harvesten SGD

### Zoeken naar 'drees' in SGD:

<http://jsru.kb.nl/sru/?operation=searchRetrieve&x-fields=snippet&x-highlight=content&x-facets=facets%3Atemporal%2Ccoverage%2Ctype&query=drees&startRecord=1&maximumRecords=1&x-collection=SGD>

#### Parameters:

- operation=searchRetrieve  
(altijd hetzelfde voor zoeken)
- x-fields=snippet  
(text snippet van full text aanzetten)
- x-highlight=content  
(text snippet op content)
- x-facets=facets:temporal,coverage,type  
(geeft deze drie facetten terug)
- query=drees  
(de query)  
CQL query-syntax (zie: <http://www.loc.gov/standards/sru/specs/cql.html>)
- startRecord=1  
(eerste hit)
- maximumRecords=1  
(records per resultatenpagina)
- x-collection=SGD  
(zoeken in staten generaal digitaal)

### Zoeken naar alles in SGD met datum tussen 1 jan 1940 en 31 december 1945:

[- query=\(date within "201-01-1940 31-12-1945"\)](#)

### Zoeken naar documenten type 'Handelingen':

[- query=\(type = "Handelingen"\)](#)

#### Gefacetteerd zoeken:

```
<facets:facet>
  <facets:name>coverage</facets:name>
  <facets:values>
    <facets:value frequency="25">eerste kamer</facets:value>
    <facets:value frequency="78">tweede kamer</facets:value>
    <facets:value frequency="7">tweede kamer ocv / ucw</facets:value>
  </facets:values>
</facets:facet>
```

[- query=\(coverage+exact+"eerste+kamer"\)+and+drees](#)

### Interpreteren zoekresultaat:

Volg de url onder **dc:identificier** voor structurele metadata van het record:

```
<srw:recordData>
  <snippet>
heer Biesheuvel (A.R.P.) in verband met bepaalde uitlatingen van de directeur van de
rijksbegroting. (Ingezonden 10 juni 1968.) 1. Heeft de Minister kennis genomen van de rede,
die de directeur van de rijksbegroting, de heer W. Drees jr., heeft uit
  </snippet>
  <dcterms:isPartOf>Staten-Generaal digitaal</dcterms:isPartOf>
  <dcx:recordIdentifier>sgd:19671968:0001488</dcx:recordIdentifier>
  <dc:identificier>
http://services.kb.nl/mdo/oai?verb=GetRecord&identificier=SGD:sgd:mpeg21:19671968:0001488&metadataPrefix=didl
  </dc:identificier>
  <dc:type>Aanhangsel</dc:type>
  <dc:title>Aanhangsel Tweede Kamer 1967-1968 nummer 753</dc:title>
  <dc:description>Aanhangsel Tweede Kamer 1967-1968 nummer 753</dc:description>
  <dcterms:extent>2</dcterms:extent>
</srw:recordData>
```

### Ophalen structuurmetatadata:

In zoekresultaat **dc:identificier**

<http://services.kb.nl/mdo/oai?verb=GetRecord&identificier=SGD:sgd:mpeg21:19671968:0001488&metadataPrefix=didl>

Bestaat uit componenten (didl:Components)

#### PDF hele document downloaden (ref in didl:Resource):

```
<didl:Component dc:identificier="sgd:mpeg21:19671968:0001488:PDF">
  <didl:Descriptor>
    <didl:Statement mimeType="text/plain" dc:type="role">PDF</didl:Statement>
  </didl:Descriptor>
  <didl:Resource
    mimeType="application/pdf"
    ref="http://ressourcessgd.kb.nl/SGD/19671968/PDF/SGD\_19671968\_0001228.pdf/"/>
</didl:Component>
```

En items (didl:Item) voor elke pagina:

#### Pagina 1:

```
<didl:Item dc:identificier="sgd:mpeg21:19671968:0001488:0001">
  <didl:Descriptor>
    <didl:Statement mimeType="text/plain" dc:type="role">page 1</didl:Statement>
  </didl:Descriptor>
  (...)
</didl:Item>
```

#### OCR tekst voor 1e pagina downloaden:

```
<didl:Component dc:identificier="sgd:mpeg21:19671968:0001488:0001:OCR">
  <didl:Descriptor>
    <didl:Statement mimeType="text/plain" dc:type="role">OCR</didl:Statement>
```

```
</didl:Descriptor>
<didl:Resource
  mimeType="text/xml"
```

```
ref="http://ressourcesgd.kb.nl/SGD/19671968/FULLTEXT/SGD\_19671968\_tekst\_0007787.xml"/>
</didl:Component>
```

De afbeelding van de 1e pagina:

```
<didl:Component dc:identificer="sgd:mpeg21:19671968:0001488:0001:image">
  <didl:Descriptor>
    <didl:Statement mimeType="text/plain" dc:type="role">image</didl:Statement>
  </didl:Descriptor>
  <didl:Resource
    mimeType="image/jpeg"
    ref="http://ressourcesgd.kb.nl/SGD/19671968/JPEG/SGD\_19671968\_0007787.jpg"/>
</didl:Component>
```

Locaties van woorden op de afbeelding:

```
<didl:Component dc:identificer="sgd:mpeg21:19671968:0001488:0001:wordPositions">
  <didl:Descriptor>
    <didl:Statement mimeType="text/plain" dc:type="role">wordPositions</didl:Statement>
  </didl:Descriptor>
  <didl:Resource
    mimeType="text/xml"
```

```
ref="http://ressourcesgd.kb.nl/SGD/19671968/COORDINATEN/SGD\_19671968\_coordinaten\_0007787.xml"/>
</didl:Component>
```

## Harvesten (alle metadata binnenhalen)

Via OAI/PMH:

<http://services.kb.nl/mdo/oai?verb=ListRecords&set=SGD&metadataPrefix=dcx>

Dit geeft de eerste 400 records. Helemaal onderaan:

```
<resumptionToken>
  SGD!2011-01-28T15:08:08.209Z!!dcx!0
</resumptionToken>
</ListRecords>
```

Voor de volgende 400 records:

<http://services.kb.nl/mdo/oai?verb=ListRecords&set=SGD&metadataPrefix=dcx&resumptionToken=SGD!2011-01-28T15:08:08.209Z!!dcx!0>

Totdat er geen resumptionToken meer wordt teruggegeven.

## Voorwaarden hergebruik & bronvermelding SGD

- De **objecten** in deze set (PDFs' JPGs, OCR's, ALTO's) zijn beschikbaar onder de [CCO 1.0 Universal verklaring](#).

- De Tweede Kamer der Staten-Generaal en de Koninklijke Bibliotheek hebben afstand gedaan van het auteursrecht op de beschrijvende & structurele **metadata**. Deze bestanden zijn dus ook beschikbaar onder de [CCO 1.0 Universal verklaring](#).

Door de CC0-vermelding die op zowel de objecten als metadata zit, is het niet noodzakelijk om aan bron- & naamsvermelding te doen. Dit neemt niet weg dat de KB en de 2e Kamer dit wel bijzonder op prijs stellen. Gebruik hierbij dan het volgende format:

*Tweede Kamer der Staten-Generaal & Koninklijke Bibliotheek, Den Haag. Bron:*

*www.statengeneraaldigitaal.nl – [link naar object]*

(voor de link naar het object : zie voorbeelden van (persistente) URLs onder “Techniek”)

### **Vragen of opmerkingen over de SGD-set?**

Mail naar [dataservices@kb.nl](mailto:dataservices@kb.nl)

*Laatste update : 18-10-2012*