

Technische uitleg dataset SGD

Deze pagina geeft uitleg over en voorbeelden van de

- Beschrijvende metadata (dcx),
- Structurele metadata (mpeg21-didl),
- Full-text objecten (pdf, ocr-xml, alto-xml, jpg)
- Harvest-API o.b.v. OAI -PMH
- Zoek-API o.b.v. SRU

voor de set SGD (<http://www.kb.nl/banners-apis-en-meer/dataservices-apis/staten-generaal-digitaal>) die de Koninklijke Bibliotheek als open data aanbiedt.

Om snel aan de slag te kunnen met SGD bieden we een [Snelstart SGD \(pdf\)](#) aan. Ook kan het nuttig zijn de [Snelstart EDBO \(pdf\)](#) te raadplegen om meer inzicht te krijgen hoe SRU-queries binnen de KB-infrastructuur opgebouwd moeten worden.

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")

Beschrijvende Metadata

Ieder document is van een aantal beschrijvende metagegevens voorzien. Een deel van de metagegevens geldt voor ieder document. Een ander deel van de metagegevens is per soort document verschillend. De tabel hieronder beschrijft de velden

De volgende metagegevens gelden voor ieder document	
<dc:type>	Het soort document (documenttype). Dit kan zijn <ul style="list-style-type: none"> • Handelingen (kamerverslagen), • Kamerstuk (bijlagen), • Aanhangsel (kamervragen) of • Register
<dcterms:temporal>	Het vergaderjaar. Dit bestaat over het algemeen uit twee jaartallen, bijvoorbeeld 1990-1991. Bij registers: jaartal(len) waar het register betrekking op heeft
<dc:coverage>	Kamer / type vergadering <ul style="list-style-type: none"> • Eerste Kamer, • Tweede Kamer • Staten-Generaal • Openbare Commissie Vergadering (OCV) en de Uitgebreide (UCV).
<dc:rights>	De rechten op het object. Die liggen formeel bij de Tweede Kamer der Staten-Generaal, maar zijn vrijgegeven onder de CC0 1.0 Universal verklaring . Een uitzondering hierop zijn de foto's in de Handelingen. Het auteursrecht berust bij de fotografen.
<dcx:recordIdentifier>, <dc:identifier>, <record-id>	Unieke identifiers
<dc:description>	Een beschrijving
<dcterms:extent>	Het aantal pagina's in het document
Voor documenten van het type Handeling (kamerverslag) gelden de volgende metagegevens	
<ul style="list-style-type: none"> • Voorbeeldrecord http://services.kb.nl/mdo/oai?verb=GetRecord&identificer=SGD:sgd:19921993%3A0000869 	
<dc:date>	De datum: dit betreft de datum waarop de vergadering is gehouden.
<dcx:startpage> & <dcx:endpage>	De paginanummers: dit betreft de paginanummers uit het originele document.
<dcx:contributor role="speaker">	Sprekers: dit betreft de personen die tijdens een vergadering aan het woord zijn geweest
<dcx:parlementNumber>	Kamerstukken: dit betreft de kamerstukken die in een vergadering genoemd worden
Voor documenten van het type Kamerstuk (Bijlage) gelden de volgende metagegevens	
<ul style="list-style-type: none"> • Voorbeeldrecord : http://services.kb.nl/mdo/oai?verb=GetRecord&identificer=SGD:sgd:19831984%3A0000427 	
<sgd:kamerstuknummer>	Kamerstuknummer: dit betreft het hoofdnummer van het document zoals dat in het oorspronkelijke document is vermeld.
<sgd:ondernummer>	Ondernummer: dit betreft het ondernummer van het document zoals dat in

	het oorspronkelijke document is vermeld.
<sgd:rijkswetnummer>	Rijkswetnummer: sommige Kamerstukken bevatten een Rijkswet, die van een afzonderlijk nummer zijn voorzien.
<dc:title>	Titel: dit betreft de titel zoals die bovenaan het oorspronkelijke document staat.
<sgd:signature>	Personen die het stuk ondertekend hebben
<dcx:startpage> & <dcx:endpage>	De paginanummers: dit betreft de paginanummers uit het originele document
Voor documenten van het type Kamervraag / Aanhangsel gelden de volgende metagegevens	
<ul style="list-style-type: none"> Voorbeeldrecord http://services.kb.nl/mdo/oai?verb=GetRecord&identificer=SGD:sgd:19911992%3A0000038 	
<sgd:aanhangselnummer>	Nummer aanhangsel: dit betreft het aanhangselnummer zoals dat in het oorspronkelijke document is vermeld
<dc:subject>	Onderwerp: dit betreft de kamervraag die in het Aanhangsel aan de orde is.
<sgd:questioner>	Vraagsteller: dit betreft de persoon die de kamervraag gesteld heeft. Er kunnen meerdere vraagstellers bij één Aanhangsel voorkomen.
<sgd:dateQuestion>	Datum vraag: dit betreft de datum waarop de kamervraag is gesteld.
<sgd:responder>	Beantwoorder: dit betreft de persoon die de kamervraag heeft beantwoord. Er kunnen meerdere beantwoorders bij één Aanhangsel voorkomen.
<sgd:dateResponse>	Datum antwoord: dit betreft de datum waarop de kamervraag is beantwoord.
<dcx:startpage> & <dcx:endpage>	De paginanummers: dit betreft de paginanummers uit het originele document.
Voor documenten van het type Register gelden de volgende metagegevens	
<ul style="list-style-type: none"> Voorbeeldrecord http://services.kb.nl/mdo/oai?verb=GetRecord&identificer=SGD:sgd:1967:0001424 	
<dc:title>	Titel van het register
<sgd:registerType>	Type register: personen of zaken (=onderwerpen)

Naast bovenstaande vier soorten objecten zijn er nog twee andere objecttypen :

- [kaarten](#) die zijn gepubliceerd in kamerstukken en
- lemma's uit [de persoons-en zaakregisters](#).

Deze zijn in aparte sets *sgd:kaarten* (1112 records), resp. *sgd:register*(38.294 records) ontsloten.

De tabel hieronder beschrijft de metadatavelden voor deze objecten.

Voor Kaarten gelden de volgende metagegevens	
<ul style="list-style-type: none"> Voorbeeld http://services.kb.nl/mdo/oai?verb=ListRecords&set=sgd:kaarten&metadataPrefix=dcx 	
<dc:title>	Titel van de kaart, zoals die (meestal) op/bij de kaart is vermeld

<dc:description>	Toelichting / achtergrondinformatie bij de kaart,
<dc:date>	Jaartal van de kaart
<dc:subject>	Thema('s) van de kaart. Voor overzicht mogelijke thema's zie http://statengeneraaldigitaal.nl/kaarten
<sgd:region>	Regio van de kaart. Voor overzicht mogelijke regio's zie http://statengeneraaldigitaal.nl/kaarten
<sgd:scale>	Schaal van de kaart
<dcterms:spatial>	Geografisch gebied van de kaart. Bounding box - 4 paren (x1;x2;y1;y2) Per paar : (breedte, lengte). Paren gescheiden door “;”
<sgd:mapcentre>	Geografische breedte & lengte van het middelpunt van de kaart
<dcx:recordidentifier>	Persistente link naar de gedigitaliseerde kaart. Let op : op het eind van de URL staat steeds “:0n:image” . Hier ontbreekt een nul , waardoor de link niet werkt. Er moet een extra 0 worden toegevoegd “:00n:image”

Voor elke **lemma uit een register** gelden de volgende metagegevens (de triviale zijn weggelaten)

- Voorbeeld <http://services.kb.nl/mdo/oai?verb=ListRecords&set=sgd:register&metadataPrefix=dcx>

<dc:identifier>	Persistente URL naar lemma in XML-formaat
<dc:title>	Titel v/h lemma (persoonsnaam of onderwerp)
<dcterms:isPartOf >	Persistente URL naar structureel metadatabestand (mpeg21-didl) van het register waar lemma deel van uit maakt
<sgd:startDate> & <sgd:endDate>	Begin en einddatum v/h lemma
<sgd:registerType>	Type register waar lemma deel van uit maakt: persoons- of zaakregister
<sgd:registerTitle>	Titel v/h register waar lemma deel van uit maakt. Voor alle mogelijke titels zie http://statengeneraaldigitaal.nl/registers
<sgd:registerUri>	Zie bij <dcterms:isPartOf >

Structurele metadata & objecten

De beschrijvende metadata van een object met *identifier sgd:18561857:0000174* is op te vragen via <http://services.kb.nl/mdo/oai?verb=GetRecord&identifier=SGD:sgd:18561857:0000174>

De bijbehorende bestanden worden als volgt verkregen

- Webpagina – beeldweergave pagina 4 van document <http://statengeneraaldigitaal.nl/document?id=sgd:18561857:0000174&pagina=4>
- Webpagina – tekstweergave pagina 4 van document <http://statengeneraaldigitaal.nl/document/tekst?id=sgd:18561857:0000174&pagina=4>
- Structurele metadata (mpeg21-didl) <http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174&role=xml> of, equivalent, via <http://services.kb.nl/mdo/oai?verb=GetRecord&identifier=SGD:sgd:mpeg21:18561857:0000174&metadataPrefix=didl>
- JPG van pagina 4 van document <http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174:0004:image>

- Thumbnail van pagina 4 van het document (15% van oorspronkelijke grootte)
<http://imageviewer.kb.nl/ImagingService/imagingService?&id=sgd:mpeg21:18561857:000174:0004:image&&zoom=0.15>
- PDF hele document
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174>
- OCR/TXT van pagina 4 van document
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174:0004:OCR>
- ALTO woordcoördinaten van pagina 4 van document
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174:0004:wordPositions>
- Volgorde van sprekers tijdens de vergadering
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:18561857:0000174:contributors>

Daarnaast:

- Lemma - item in personenregister (1967)
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:1967:0001424:L0003:lemma>
- Lemma - item in zaakregister (1972)
<http://resolver.kb.nl/resolve?urn=sgd:mpeg21:1972:0001695:L0124:lemma>

De URLs die beginnen met *http://resolver.kb.nl* zijn persistent en moeten bij voorkeur gebruikt worden in verwijzingen.

Metadata harvesten via de download-API (OAI-PMH)

De SGD-dataset heeft een download-API o.b.v. het [OAI-PMH-protocol](#). Hiermee kun je zowel de beschrijvende als de structurele metadata harvesten m.b.v. een OAI-harvester

1. De OAI-PMH-syntax is hoofdlettergevoelig!
2. Beschrijving KB OAI-repository : <http://services.kb.nl/mdo/oai?verb=Identify>
3. Voor het harvesten van de *beschrijvende metadata*:
<http://services.kb.nl/mdo/oai?verb=ListRecords&set=SGD&metadataPrefix=dcx> (let op: SGD dus met hoofdletters!)
4. Voor het harvesten van de *beschrijvingen van de register-lemma's*:
<http://services.kb.nl/mdo/oai?verb=ListRecords&set=sgd:register&metadataPrefix=dcx> (sgd:register met kleine letters)
5. Voor het harvesten van de *beschrijvingen van de kaarten*
<http://services.kb.nl/mdo/oai?verb=ListRecords&set=sgd:kaarten&metadataPrefix=dcx>
Let op het veld <dcx:recordidentifier>: hier staat 00n:image, maar hier moet een **extra nul** bij (000n:image) om de url werkend te maken.
6. *ResumptionToken*: met de laatste XML-tag in deze respons (de ResumptionToken) - SGD!2011-01-28T15:08:08.209Z!!dcx!0 - kan de volgende batch records opgevraagd worden:
<http://services.kb.nl/mdo/oai?verb=ListRecords&resumptionToken=SGD!2011-01-28T15:08:08.209Z!!dcx!0> (metadataPrefix=dcx hoeft dan niet meer in de query string aangegeven te worden.) Op die manier kan iteratief de hele set beschrijvende metadata overgehaald worden.

7. *Structurele metadata* : kan op vergelijkbare manier geharvest worden door metadataPrefix=didl te kiezen :
<http://services.kb.nl/mdo/oai?verb=ListRecords&set=SGD&metadataPrefix=didl>
8. *Identifiers hele set*: via
<http://services.kb.nl/mdo/oai?verb=ListIdentifiers&set=SGD&metadataPrefix=dcx> (of metadataPrefix=didl voor structurele metadata) kan een lijst van object-identifiers worden opgevraagd.
9. *Identifiers kaarten* :
<http://services.kb.nl/mdo/oai?verb=ListIdentifiers&set=sgd:kaarten&metadataPrefix=dcx>
10. *Identifiers register-lemma's*:
<http://services.kb.nl/mdo/oai?verb=ListIdentifiers&set=sgd:register&metadataPrefix=dcx>

Dataset doorzoeken via de zoek-API (SRU)

De SGD-dataset heeft ook een zoek-API o.b.v. het SRU-protocol. Hiermee kun je de beschrijvende metadata doorzoeken. SRU is gebaseerd op CQL.

Enkele voorbeelden van SRU-queries voor SGD zijn:

- De SRU-syntax is hoofdlettergevoelig!
- Informatie over de KB metadata index: <http://jsru.kb.nl/sru/sru?operation=explain>
- *Eenvoudige zoekvraag*:
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=Wiegel&recordSchema=dcx> geeft de beschrijvingen terug van de eerste 20 publicaties waar 'Wiegel' in de full-text voor komt.
 Let op: dit is niet gelijk aan "Geeft de eerste 20 records waar 'Wiegel' in voor komt. (metadatozoeken vs full-text zoeken)
- Als we die *full-text (OCR)* in de respons willen zien, voegen we "&x-fields=content" aan de request toe en veranderen het recordSchema naar dc
<http://jsru.kb.nl/sru/SGD?version=1.1&operation=searchRetrieve&recordSchema=dc&query=Wiegel&x-fields=content>
- Als we i.p.v. de hele full-text alleen de *abstract* willen zien, laten we "&x-fields=content" weg:
<http://jsru.kb.nl/sru/SGD?version=1.1&operation=searchRetrieve&recordSchema=dc&query=Wiegel>
- *RecordSchema*: indien er geen expliciet recordSchema meegegeven wordt, dan wordt standaard recordSchema=dc aangenomen (zie explain record).
- *Maximum aantal records* : Standaard worden 20 records teruggegeven. Als je er maximaal 50 wil opvragen, geef je maximumRecords=50 als argument mee:
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=Wiegel&maximumRecords=50> . De aanroep
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=Wiegel&maximumRecords=1> geeft dus maar 1 Wiegel-record terug
- *Beginrecord*: Als je records 4 t/m 10 uit de Wiegel-resultaatset wilt tonen, gebruik je de startRecords-parameter, samen met de maximumRecords parameter als volgt:
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=Wiegel&startRecord=4&maximumRecords=7>

- *Samengestelde zoekvraag I* : Als we willen zoeken op 'De Minister van Binnenlandse Zaken' gaat dit als volgt:
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=%22De%20Minister%20van%20Binnenlandse%20Zaken%22&maximumRecords=10&recordSchema=dcx&x-fields=content>
- *Samengestelde zoekvraag II* : Als we willen zoeken op 'De Minister van Binnenlandse Zaken' in de full-text van alleen Kamerstukken (Bijlagen) die ondertekend zijn door "E. van Thijn" , en maximaal 10 records willen terugkrijgen, gaat dit als volgt:
<http://jsru.kb.nl/sru/SGD?&operation=searchRetrieve&version=1.1&query=%22De%20Minister%20van%20Binnenlandse%20Zaken%22%20AND%20sgd.signature=%22E.%20van%20Thijn%22%20AND%20dc.type=Kamerstuk&&maximumRecords=10&recordSchema=dcx>

Zie tevens [Snelstart SGD \(pdf\)](#) en de [Snelstart EDBO \(pdf\)](#) voor meer uitleg over & meer voorbeelden van SRU-queries binnen de KB-infrastructuur

Vragen of opmerkingen over de SGD-set?

Mail naar dataservices@kb.nl

Laatste update : 21-11-2012