

## Open datasets van de KB - Staten-Generaal Digitaal

De Koninklijke Bibliotheek (KB) wordt steeds vaker benaderd door partijen, projecten en ontwikkelaars die (een deel) van de data die de KB beheert willen hergebruiken. Met deze vorm van dienstverlening, Dataservices genoemd, wil de KB aan deze vraag tegemoet komen. Hiermee staat de KB als 'grondstofleverancier' aan de basis van een waardeketen.

De KB biedt de [dataset Staten-Generaal Digitaal \(SGD\)](#) aan als open (cultuur)data. Enkele karakteristieken van deze dataset:

- Alle Kamerverslagen, Kamerstukken en Kamervragen van de Eerste Kamer en de Tweede Kamer in de periode 1814 tot 1995. [Meer informatie over de dataset ...](#)
- De bestanden en objecten (PDF, JPG, OCR, ALTO) zijn beschikbaar zonder rechtenbeperkingen (CC0-licentie en Publiek Domein Verklaring)
- De metadata is te [harvesten](#) via OAI-PMH en te doorzoeken via [SRU](#). De objecten zijn via persistente links gekoppeld aan de metadata. [Meer over de techniek \(PDF\)](#).



### Hergebruik & toepassingen open dataset SGD

Mede door het openstellen van de SGD-dataset zijn verschillende initiatieven ontstaan. Eén daarvan is het onderzoeksprogramma Political Mashup en de afgeleide onderzoeksprojecten *Ngram-viewer* en *Polidocs*.

#### Political Mashup - wetenschappelijk onderzoek naar politieke partijen

[Political Mashup](#) is een onderzoeksprogramma waarin politicologen, historici, informatici en tekst-wetenschappers samenwerken om politieke data inzichtelijker en makkelijker toegankelijk te maken. Naast de SGD-set wordt gebruik gemaakt van data van onder meer partij- en verkiezingsprogramma's, politieke biografieën, websites van politieke partijen, krantenberichten en berichten op weblogs of Twitter van politici of politiek geïnteresseerden. Al deze gegevens worden in een rijk geannoteerde en inmiddels [in de prijzen gevallen](#) dataset in XML-formaat bijeengebracht en opgeslagen.

*"Elke burger wordt door Political Mashup in staat gesteld zowel voor het verre verleden, als voor de dag van gisteren, het politieke proces beter te doorgronden en ook zijn gekozen volksvertegenwoordigers effectief te controleren. Het bevordert daarmee niet alleen de betrouwbaarheid (en herhaalbaarheid) van het politiek-historisch onderzoek, maar stelt de burger ook in staat echt democratisch burger te zijn."*  
[prof. dr. W.W. Mijnhardt, juryvoorzitter van de Dataprijs humaniora en sociale wetenschappen](#)

#### Voordelen

Het digitaal bijeenbrengen, opslaan en het semantisch kunnen doorzoeken van grote hoeveelheden politiek gerelateerde data kent enorme voordelen. De belangrijkste op een rij:

- Er kunnen sneller, eenvoudiger en nauwkeuriger dan voorheen grote hoeveelheden politiek materiaal worden doorzocht en geanalyseerd.
- Door de grote variëteit aan politiek materiaal - van officiële politieke documenten tot digitale bronnen zoals blogs en nieuwsberichten - kunnen complexere onderzoeksvragen sneller en beter worden beantwoord.
- De digitale zoekresultaten kunnen snel worden gevisualiseerd. Bijvoorbeeld als grafiek, graaf of tagcloud. Zonder de zoekresultaten volledig uit te pluizen is al een snelle eerste indruk mogelijk.



*Semantisch zoeken levert enorme voordelen op. Vergelijk bijv. eens de zoekresultaten van "Zendtijd voor Politieke Partijen" op de Mashup met die op [StatengeneraalDigitaal.nl](http://StatengeneraalDigitaal.nl). Political Mashup voorziet in rijke, geannoteerde, gestructureerde en verbonden tekst terwijl StatengeneraalDigitaal 'slechts' platte tekst met metadata toont.*

### Semantisch zoeken

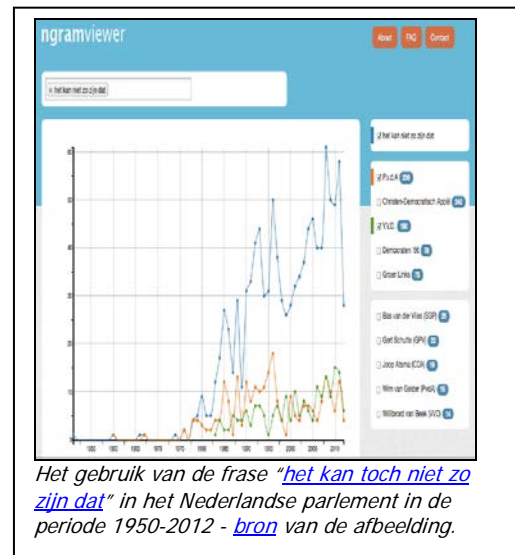
De data is semantisch doorzoekbaar waardoor betekenisvollere zoekresultaten worden gegenereerd. Bijvoorbeeld welke politici regelmatig in de Kamer op elkaar reageren, wie een onderwerp voor het eerst geagendeerd heeft, welke beloften tijdens de verkiezingen over een thema zijn gedaan of het publiek reageerde op een bepaalde politieke gebeurtenis.

### Politieke ngram-viewer

Een 'spin-off' van Political Mashup is de politieke *ngram-viewer*. Geïnspireerd door de ngram-viewer van [Google Books](http://Google Books), maakt de Political Mashup ngram-viewer het mogelijk het gebruik van (maximaal 7) woorden of woordgroepen (meerdere opeenvolgende woorden) in het Nederlandse parlement door de tijd heen te visualiseren a.d.h.v. een tijdslijn met de frequentie van die woorden/groepen. De ngram-viewer beschikt over filters, waardoor het vergelijken van het gebruik van verschillende termen door partijen of zelfs individuele parlementsleden mogelijk is.

De viewer geeft verder inzicht in de *agenda setting* (welke partij zet een onderwerp of thema het eerste op de agenda en welke partijen volgen) en het gebruik van *framing*. Denk bij dit laatste aan het gebruik van bepaalde woorden (bijvoorbeeld *kopvot*, *Jan Modaal* of *jip-en-janneketaal*) door politieke partijen waarbij een onderscheid tussen partijen en zelfs gebruik door individuele personen gemaakt kan worden.

De ngram-viewer is ontwikkeld door het bedrijfje [Dispectu](http://Dispectu) dat is opgericht door twee voormalige informaticastudenten die betrokken waren bij Political Mashup. Het bedrijf structureert, analyseert en visualiseert grote datasets door er interactieve & slimme (mobiele) toepassingen mee te bouwen.



### Polidocs

De KB staat (mede) aan het begin van een waardeketen: de KB heeft de SGD-set als open data onder CC0-licentie beschikbaar gemaakt, het team rond Political Mashup heeft deze ruwe data gekopieerd, gestructureerd, geannoteerd en verbonden met andere bronnen. Op basis van deze verrijkte open data hebben studenten & wetenschappers *value-adding* applicaties gebouwd, waaronder PoliDocs.

[PoliDocs](http://PoliDocs) is een zoekmachine waarmee gezocht kan worden in moties, kamervragen en uitspraken van leden van de Eerste Kamer en de Tweede Kamer uit de periode 1984-2008. De zoekmachine is ontwikkeld als proof-of-concept door een groep van de opleiding Informatiekunde aan de UvA; ze wilden aantonen dat het mogelijk is om een zoekmachine te maken voor de Nederlandse kamerstukken die dezelfde functionaliteit biedt als <http://www.theyworkforyou.com> uit het Verenigd Koninkrijk en voldoet aan 15 uitgangspunten voor een [Open Democratie](http://Open Democratie).

### Dataservices van de KB

De KB beheert grote hoeveelheden data, waarop diensten worden geleverd. Met de dienst [Dataservices](http://Dataservices) sluit de KB aan bij een internationale trend richting linked (open) data. Door gestructureerde data zo te publiceren dat data uit verschillende bronnen gecombineerd kan worden, kan waarde worden toegevoegd.

De beschikbare KB-datasets en de laatste ontwikkelingen en toepassingen staan op de webpagina [Dataservices / API's](http://Dataservices / API's). Voor meer informatie kunt u mailen naar [dataservices@kb.nl](mailto:dataservices@kb.nl).