

# STOP tekstmodel

## Inleiding en achtergrond

STOP is gericht op het coderen van (informatie in) overheidspublicaties. Deze publicaties zijn semi-gestructureerd; ze bevatten een combinatie van relatief ongestructureerde leesbare tekst en gestructureerde metadata.

Het coderen van de leesbare tekst van een publicatie is de kern van STOP. De opeenvolging van tekens, geordend in alinea's, tabellen, lijsten, etc. is de kapstok waaraan additionele informatie zoals metadata, verwijzingen naar afbeeldingen etc. kan worden opgehangen. Het tekstmodel beschrijft welke entiteiten relevant zijn bij het coderen van de tekst en de additionele informatie.

STOP is in principe syntax neutraal. Entiteiten als “hoofdstuk” of “artikel” kunnen op meerdere manieren vastgelegd worden. In de praktijk wordt XML gebruikt als bronformaat voor de publicaties. De structuur van een publicatie wordt beschreven met XML Schema en additionele business rules via Schematron. Het is onvermijdelijk dat de beperkingen en mogelijkheden van XML her en der impact hebben op STOP – dit wordt dan aangegeven.

STOP kent een aantal abstracte entiteiten die gebruikt worden voor classificatie van de niet—abstracte entiteiten:

- container
- block
- inline
- marker
- structuurelement
- metadata-uitspraak
- binaire

In de volgende hoofdstukken wordt nader beschreven welke rol de abstracte entiteiten vervullen, welke implementaties beschikbaar zijn, welke eigenschappen ze hebben, en hoe uitbreidingen kunnen plaatsvinden.

## Abstracte entiteiten

In dit hoofdstuk worden de abstracte entiteiten nader beschreven.

### Container

Een container is een groepering van 1 of meer containers en / of 1 of meer structuurelementen. De container bevat geen tekst of een default vormgeving. Containers zijn over het algemeen grote logische blokken binnen een publicatie. De inhoud van een container varieert sterk.

## Block

Een block is een container met tekst en kan daarnaast 1 of meer inline elementen en / of 1 of meer marker elementen en / of 1 of meer block elementen bevatten. Een block is daarmee een entiteit die fungeert als container op tekstniveau. Het equivalent van een block in HTML is het div element.

Voorbeelden van block entiteiten binnen STOP zijn alinea, tabel.

## Inline

Een inline bevat tekst en / of inline elementen. Het equivalent van een inline in HTML is het span element. Een voorbeeld van een inline is de aanduiding voor vet, italic, etc. Een inline kan ook gebruikt worden voor meer betekenisvolle aanduidingen zoals bouwhoogte.

## Marker

Een marker is een positie in de tekst waaraan informatie gekoppeld is. In tegenstelling tot een inline omvat een marker geen tekst. Een voorbeeld van een marker is een voetnootreferentie of een afbeelding.

## Structuurelement

Een structuurelement is een speciaal soort container waarmee de formele structuur van een regeling kan worden uitgedrukt. Het gaat om hoofdstuk, paragraaf, etc.

## Metadata-uitspraak

Een metadata-uitspraak is een combinatie van object, predicate en value. Een metadata container bevat 1 of meer metadata-uitspraken. Veelal is het object gelijk aan het de ouder van de metadata container waartoe de metadata-uitspraak behoort, maar dit is niet strikt noodzakelijk.

## Entiteiten

In dit hoofdstuk wordt een overzicht gegeven van de (niet-abstracte) STOP entiteiten. Er is gekozen voor een logische onderverdeling (in plaats van een groepering per abstracte entiteit). Specifieke details als cardinaliteit worden op deze plaats zoveel mogelijk vermeden.

## Documentniveau

Een officiële publicatie is een document dat door een bevoegd gezag wordt bekendgemaakt en daarmee gepubliceerd. De entiteit **officiële-publicatie** is een container die bestaat uit een metadata container en een container die het publicatieblad representeert (bijvoorbeeld **staatscourant**, **gemeentebblad**, etc.).

Sommige officiële publicaties bevatten wijzigingsinstructies die geconsolideerd moeten worden. Het resultaat van een dergelijke consolidatie is een zogeheten toestand. De entiteit **toestand** is een container die bestaat uit een metadata container (verplicht), een regeling container (verplicht)<sup>1</sup>.

Een publicatieblad container (zoals **staatscourant**) bevat een **bladaanduiding** container, een **regeling** container en een **nawerk** container. De regeling is in dit geval de “hoofdttekst” van de

---

<sup>1</sup> Omdat een toestand niet in een publicatieblad wordt gepubliceerd, is er hier geen noodzaak voor een publicatieblad entiteit.

bekendmaking, terwijl de optionele nawerk container ruimte biedt voor nota toelichting, bijlage en dergelijke.

## Regelingtekst

Een **regeling** bestaat uit<sup>2</sup> een opschrift, aanhef, lichaam, slotformulier, ondertekening, en eventuele bijlagen. De volgorde van deze entiteiten is relevant.

Het **opschrift** is een omschrijving (titel) van de regeling. Het komt overeen met wat voorheen werd aangeduid als "intitulé". Alhoewel het opschrift veelal bestaat uit een enkele string, is er modelmatig in STOP een opschrift container, die bestaat uit een (1) alinea blok, en pas daarbinnen tekst.

Een **aanhef** bevat inleidende tekst over de regeling, waaronder de considerans. Aanhef is een container, die een of meer block(s) bevat (alinea of lijst, geen tabel).

Een **lichaam** is een container die de hoofdtekst van de regeling bevat. Dit komt overeen met regeling-tekst in voorlopers van STOP. Een lichaam bevat of een tekst volgens artikelstructuur (met optioneel groepering via hoofdstuk, paragraaf, etc.) of een vrije tekst (eventueel geordend via een divisiestructuur).

Het **slotformulier** bevat afsluitende (procedurele) tekst over de regeling. De **ondertekening** bevat zaken datum en plaats van publicatie, en rol plus naam van de ondertekenaar(s). Beide voorgaande entiteiten worden gecombineerd tot een entiteit **regeling-sluiting**. Dit is een container.

In voorgangers van STOP werd de inhoud van regeling-sluiting zo betekenisvol mogelijk gecodeerd. In de praktijk bleek dit een te overmatig niveau van detail. In STOP wordt daarom gekozen voor generieke alinea entiteiten met een optionele rol eigenschap die dan wel moet gevuld wordt met voorgeschreven waarden (zoals "slotformulier").

De entiteit **bijlage** is een container die 1 of meer blocks bevat. Dit kunnen tabellen, afbeeldingen, of tekst zijn. Let op: er zijn twee soorten bijlagen binnen STOP. Binnen regelingtekst gaat het om bijlage(n) die formeel tot de regeling behoren (en dus ook op hetzelfde niveau bindend zijn).

Er bestaat een verschil tussen **officiële-publicatie** en **toestand** voor wat betreft de inhoud van regeling. Bij **toestand** is aanhef en regeling-sluiting niet relevant, en binnen het lichaam is informatie over was-wordt niet relevant (die is immers dan reeds gebruikt bij de totstandkoming van de toestand). Binnen **officiële-publicatie** is was-wordt informatie juist wel relevant voor die publicaties die als basis voor een toestand dienen.

## Tabellen

Een tabel is een block. Hierdoor kunnen tabellen (als siblings) gecombineerd worden met alinea's, lijsten en afbeeldingen binnen een container.

STOP ondersteunt het CALS model voor tabellen<sup>3</sup>. Omwille van de interoperabiliteit is gekozen voor het [CALS Exchange Table Model](https://en.wikipedia.org/wiki/CALS_Table_Model).

---

<sup>2</sup> Zie [Aanwijzingen voor de regelgeving, aanwijzing 94](#)

<sup>3</sup> Zie [https://en.wikipedia.org/wiki/CALS\\_Table\\_Model](https://en.wikipedia.org/wiki/CALS_Table_Model)

In afwijking op voorlopers wordt binnen STOP ook het HTML tabel model ondersteund. Het betreft hier de implementatie zoals deze is opgenomen binnen XHTML 1.0 Strict.

## Lijsten

Een lijst is geclassificeerd als een block. Deze keuze is met name gebaseerd op het feit hoe lijsten in de tekst worden gebruikt. Gezien de recursieve aard van lijsten en het feit dat de tekst nooit op lijstniveau zit (maar altijd binnen lijst-items) was een keuze voor container immers ook verdedigbaar geweest.

Een **definitielijst** is een specifieke vorm van een lijst die gericht is op het coderen van een lijst met definities. Een definitielijst bevat 1 of meer **definitie-item** entiteiten. Deze bestaan uit een volgnummer, een term en een definitie.

## Artikelen en groepering daarvan in hoofdstuk, paragraaf, etc.

Een **artikel** is een bepaling in een regelgevend document, bijvoorbeeld een wet, een regeling of een verdrag. Het artikel is bedoeld om een concrete regel te geven op het 'laagste' niveau van de regeling<sup>4</sup>.

Indien het aantal artikelen toeneemt dan kan omwille van de toegankelijkheid gekozen worden voor een groepering van artikelen. In [aanwijzing 97](#) van de [Aanwijzingen voor de regelgeving](#) wordt hier nader op ingegaan:

2 Bij een verdeling op één niveau worden de onderdelen "hoofdstuk" of "paragraaf" genoemd.

3 Bij een verdeling op twee niveaus worden de onderdelen van het eerste niveau "hoofdstuk" en de onderdelen van het tweede niveau "paragraaf" genoemd.

4 Bij een verdeling op meer dan twee niveaus worden de onderdelen in volgorde van omvang "deel", "hoofdstuk", "titel", "afdeling" en "paragraaf" genoemd, met dien verstande dat in ieder geval de aanduidingen "hoofdstuk" en "paragraaf" worden gebruikt.

In lijn met bovenstaande aanwijzing zijn er binnen STOP entiteiten aangemaakt voor deel, hoofdstuk, titel, afdeling, en paragraaf gemaakt. Via business rules wordt het toegestane gebruik ("1 niveau dan hoofdstuk of paragraaf") afgedwongen. In de praktijk wordt hier overigens nog wel eens van afgeweken (vandaar dat dit via business rules wordt afgedwongen, en niet via de definitie van de entiteiten zelf).

Zowel **artikel** als de groeperingen **hoofdstuk**, etc. zijn geclassificeerd als structurelement. Dit laat onverlet dat er een essentieel verschil qua semantiek tussen beide is. Het artikel is de kern: hier staan de bepalingen (de tekst) en dit is de eenheid van wijziging. De groeperingen zijn optioneel, en worden geacht louter redactionele tekst te bevatten (bijvoorbeeld "vervallen").

Een artikel bevat een kop en daarnaast een keuze tussen vrije tekst (in de vorm van block(s)) of een of meer **lid** entiteiten. De vrije tekst kan bestaan uit alle soorten blocks. Een **lid** is genummerd en bevat tekst<sup>5</sup>. Aangezien vrije tekst binnen een artikel ook een lijst kan bevatten – en de lijstitems ook genummerd zijn – kan verwarring ontstaan met leden. In de praktijk is een opsomming veelal

<sup>4</sup> Zie [https://nl.wikipedia.org/wiki/Artikel\\_\(recht\)](https://nl.wikipedia.org/wiki/Artikel_(recht)).

<sup>5</sup> De aanwijzingen voor de regelgeving zijn aanmerkelijk beperkter dan de praktijk: <http://wetten.overheid.nl/jci1.3:c:BWBR0005730&hoofdstuk=3&paragraaf=3.5&aanwijzing=99&z=2011-05-11&g=2011-05-11>. STOP laat meer toe dan 1 alinea, maar kan via business rules tot conformiteit met aanwijzing 97 komen.

gecombineerd met 1 of meer alinea's. In dat geval moet het om vrije tekst gaan, omdat buiten de leden geen alinea's kunnen staan.

## Werkingsgebied

Werkingsgebied betreft het (ruimtelijk) gebied waarop (een onderdeel) van een publicatie betrekking heeft. Iedere publicatie heeft een werkingsgebied. Indien niet aangegeven komt het werkingsgebied overeen met het gebied waarover het bevoegd gezag dat verantwoordelijk is voor de publicatie iets te zeggen heeft<sup>6</sup>.

De waarde van een werkingsgebied is een set van 1 of meer geografische namen (uit een te benoemen waardenlijst, bijvoorbeeld "Alkmaar") of GML 3.2 structuren<sup>7</sup>.

Werkingsgebied is default gekoppeld aan het documentniveau, maar kan ook binnen het document op structurelementen<sup>8</sup> worden gespecificeerd. Hiervoor geldt dan als eis dat de hiërarchie van de tekst wordt gevolgd; een onderliggend werkingsgebied moet gelijk zijn aan of binnen een bovenliggend werkingsgebied liggen.

Een implicatie van de opname van werkingsgebied in STOP is dat de XML serialisatie moet kunnen omgaan met GML fragmenten<sup>9</sup>. Overigens is een publicatie op basis van STOP geen op zichzelf staand GML document (dat in een viewer te openen is), maar het is wel denkbaar dat dit via een standaard XSLT gegenereerd kan worden.

<nog bepalen: hoe omgaan met label?, toevoegen aan STOD>

## Metadata

In het tekstmodel is de lopende tekst – gebundeld als publicatie – de kern. Op diverse niveau's van de publicatie kan het vanuit verschillende doelstellingen nodig zijn om informatie toe te voegen die niet zichtbaar is in de lopende tekst. Deze informatie is metadata. Iedere metadata uitspraak volgt het patroon <object> <predicate> <value>.

Het object bevat een identifier van de entiteit waarover de uitspraak gaat. Bij serialisatie als XML is het object is veelal impliciet omdat het uit de plaatsing van de metadata uitspraak volgt. Indien metadata uitspraken fysiek ontkoppeld worden van de publicatie dan moet object expliciet gemaakt worden.

Het predicate is een string of een URI die aangeeft welke relatie tussen object en value bestaat.

De value van een metadata uitspraak is string, een (ongeordende) set of een (geordende) lijst

---

6 In voorlopers van STOP was het idee van werkingsgebied niet expliciet benoemd. Door in deze gevallen uit te gaan van (het gebied van) bevoegd gezag kan hiermee omgegaan worden.

7 Surface; Polygon; CompositeSurface; OrientableSurface.

8 Nog bepalen: ook andere entiteiten (bijvoorbeeld nota van toelichting)? Ook nog bepalen: werkingsgebied op onderdelen van artikel (lid; lijstonderdeel; volzin)

9 In de <http://www.opengis.net/gml/3.2> namespace. In voorgangers van STOP werd geen gebruik gemaakt van XML namespaces, maar in de STOP XML serialisatie zal minimaal 1 namespace gebruikt moeten worden.

In voorlopers van STOP werd dit (in XML) gerealiseerd via een metadata element met daarin 1 of meer (als set) meta elementen, ieder met name, content en optioneel scheme attribuut. Het probleem hiermee is dat samenhangende informatie niet goed uitgedrukt kon worden<sup>10</sup>. Voor STOP is een iets geavanceerder model gekozen.

Er is binnen STOP een metadata container. Deze bestaat uit 1 of meer metadata-uitspraken. Het object van een metadata-uitspraak is default de ouder van de metadata container. Implicatie van de keuze om metadata een container te laten zijn, is dat deze niet op het niveau van block geplaatst kan worden. Indien een metadata-uitspraak over een block of inline gaat dan moet dit expliciet als object benoemd worden in de metadata uitspraak (via een referentie naar het id).

## Consolidatie van metadata en werkingsgebied

Hoe maak je dit consolideerbaar?

- Identificatie moet mogelijk zijn.
  - Er is niet noodzakelijkerwijze een tekst, dus hoe presenteert je het feit dat er 1 of meer werkingsgebieden aan een structuurelement hangen?
- Granulariteit?
  - featureMember
  - UitbeiGML XML 1-op-1 vervangen
  -

## Nota-toelichting

Bij document als geheel; structuurelement; artikel

Hoe maak je dit consolideerbaar?

- Identificatie moet mogelijk zijn.
- Scenario's:
  - nieuw artikel – dan ook mogelijk nieuwe toelichting gekoppeld aan artikel
  - gewijzigde toelichting zonder wijziging in artikel

---

<sup>10</sup> Bijvoorbeeld wanneer een publicatie over meerdere locaties gaat waarbij iedere locatie weer bestaat uit een naam en een andersoortige identifier. In dat geval is het moeilijk om zonder hacks “paren” van naam en identifier te maken.