

Uitvragen

Officiële Publicaties

Handleiding voor het uitvragen van de collectie Officiële Publicaties

Auteurs	Ilja Andreas
Bestandsnaam	Handleiding uitvragen Officiële Publicaties V1.1
Status	V1.1 definitief



Document informatie

Documenthistorie

Datum	Versie	Auteur	Opmerking
05-12-2012	0.1	Ilja Andreas	1 ^e opzet
19-12-2012	0.2	Ilja Andreas	Extra zoekelementen toegevoegd
20-12-2012	0.3	Ilja Andreas	Openstaande vragen verwerkt
27-12-2012	0.4	Ilja Andreas	Zoeken met SRU stuk aangepast
24-01-2013	0.5	Ilja Andreas	Bulk Uitlevering stuk aangepast
23-07-2013	0.51	Ilja Andreas	Definitief gemaakt
22-10-2013	1.0	Ilja Andreas	Voorbeeld toegevoegd en versie 1.0 van gemaakt
07-10-2014	1.1	Ilja Andreas	Metadata.xml en metadata_owms.xml toegevoegd.



Inhoudsopgave

Document informatie	2
Inhoudsopgave	3
1 Samenvatting & Inleiding	4
2 Gebruikte termen in dit document	5
3 Zoekfunctionaliteit middels de webservice	6
3.1 Opbouw van de zoekvraag	7
3.2 CQL query	9
3.2.1 Niet gevoelig voor hoofdletters, accenten, etc.	9
3.2.2 Operatoren	10
3.2.3 Complexere zoekvragen	10
3.2.4 Sorteren	11
3.3 Uitvragen via HTTP GET of HTTP POST	11
3.4 Het zoekresultaat: De opbouw van een XML response	12
3.5 Ophalen gegevens	15
3.6 Externe bijlagen	15
4 Bulk UitleverSysteem (BUS)	16
4.1 Bron	16
4.2 Tijdstip van vullen	16
4.3 Toegang	16
4.4 Mappenstructuur	17
4.5 Metabestand Changelog.xml	18
4.5.1 Toegevoegd	19
4.5.2 Veranderd	19
4.5.3 Verwijderd	19
4.6 Gebruik dagelijkse updates	20
5 Vragen of opmerkingen over dit document	21
Bijlage 1 – Mogelijke zoekvelden	22
Bijlage 2 – Voorbeeld metabestand changelog.xml	24

1 Samenvatting & Inleiding

Voor het uitvragen van de collectie Officiële Publicaties bestaan er een aantal opties. Er kan gebruik gemaakt worden van de website <https://zoek.officielebekendmakingen.nl> waarmee officiële bekendmakingen én parlementaire documenten doorzocht kunnen worden. Verder bestaat de mogelijkheid van hergebruik van de data van de collectie Officiële Publicaties. Hiermee kunnen Officiële Publicaties (inclusief parlementaire publicaties) beschikbaar gesteld worden aan derden voor eigen gebruik en/of voor verdere distributie, eventueel na verrijking van de content. Mogelijke toepassingen hiervan zijn bijvoorbeeld het tonen van de publicaties op gemeentelijke websites, het tonen van bekendmakingen op een kaart en het ontsluiten van data op mobiele telefoons of tablets.

Voor de mogelijkheid van hergebruik van data van de collectie Officiële Publicaties zijn er twee technische koppelingen beschikbaar. De eerste betreft een technische koppeling (ook wel webservice of API genoemd) waarmee eenvoudig kan worden gezocht in de data van de collectie. Deze webservice is onderdeel van de zoekfunctionaliteit die bekend is als 'de zoekdienst'. Er zijn met deze webservice complexe zoekvragen mogelijk als: “zoek alle publicaties die vandaag in de Staatscourant zijn gepubliceerd door de gemeente Assen met het woord 'Verkeersbesluit' in de titel EN het woord 'voetpad' in de tekst”.

De tweede technische koppeling biedt de mogelijkheid voor systematische afname van officiële publicaties. Dit is een voorziening gebaseerd op SFTP genaamd het Bulkuitleversysteem. Dit systeem biedt de mogelijkheid om officiële publicaties in grote aantallen tegelijkertijd te kunnen afnemen. Hiermee is het mogelijk om als afnemer periodiek (bijv. dagelijks) alles van de collectie Officiële Publicaties op te halen óf alleen de nieuw toegevoegde en gewijzigde publicaties sinds de vorige ophaalactie. De afnemer bepaald zelf welke selectie wordt opgehaald. Deze voorziening is bereikbaar via bestanden.officielebekendmakingen.nl.

In dit document zullen de twee mogelijkheden voor het hergebruik van data van de collectie Officiële Publicaties verder in detail uitgewerkt worden.

2 Gebruikte termen in dit document

Voordat we overgaan op een toelichting op de mogelijkheden van het uitvragen van de collectie Officiële Publicaties, is het noodzakelijk om een aantal termen te definiëren.

→ **Leeswijzer:** Lees deze termen eerst door voor het verdere document te lezen.

BUS	Bulk UitleverSysteem. Dit systeem biedt de mogelijkheid om in een bulk de data van de collectie Officiële Publicaties te ontsluiten voor hergebruik.
CQL	Contextual Query Language. Taal waarmee zoektermen kunnen worden gedefinieerd (vergelijkbaar met SQL). De relevante delen van deze taal worden beschreven in dit document; zie http://www.loc.gov/standards/sru/specs/cql.html voor extra achtergrondinformatie.
Ftp	FTP staat voor File Transfer Protocol. Dit is een protocol dat uitwisseling van bestanden tussen computers vergemakkelijkt. Het standaardiseert een aantal handelingen, die tussen besturingssystemen vaak verschillen.
OEP	Afkorting van de collectie Officiële Publicaties.
SRU	Search and Retrieval by URL. Internationale open standaard voor webservices waarmee gezocht kan worden in gestructureerde data. De relevante delen van deze standaard worden beschreven in dit document; zie http://www.loc.gov/standards/sru/ voor extra achtergrondinformatie.
Veldnaam	Een inhoudelijk veld waarop gezocht kan worden, bijvoorbeeld "dcterms:title" voor de titel van een publicatie. Nota bene: in SRU wordt dit een 'index' genoemd, maar omdat dit in de database wereld vaak een andere betekenis heeft, wordt in dit document alleen 'veldnaam' gebruikt.
Webservice	Technisch mechanisme om gegevens uit te wisselen. Ook wel koppeling, koppelvlak, interface of API (Application Programming Interface) genoemd (hoewel deze termen niet geheel hetzelfde zijn, is dit voor dit document niet van belang).
Zoekdienst	De zoekmachine achter overheid.nl die wordt gebruikt door zowel de user interface van overheid.nl als door de in dit document beschreven webservice.
Zie ook	Overige termen worden als bekend verondersteld voor de doelgroep van dit document. Zie de volgende pagina's op wikipedia voor meer informatie: XML ; URL; client ; record ; query ;

3 Zoekfunctionaliteit middels de webservice

Het is mogelijk om de collectie Officiële Publicaties uit te vragen door gebruik te maken van een webservice. Hiermee kan eenvoudig worden gezocht in de data van de collectie. Deze webservice is onderdeel van de zoekfunctionaliteit die bekend is als 'de zoekdienst' en biedt de mogelijkheid om complexe zoekvragen te stellen als: “zoek alle publicaties die vandaag in de Staatscourant zijn gepubliceerd door de gemeente Assen met het woord 'Verkeersbesluit' in de titel EN het woord 'voetpad' in de tekst”. Het onderstaande schema geeft in hoofdlijnen weer hoe de webservice werkt.



1. **De client stelt een zoekvraag in de vorm van een URL met parameters.** Hierbij is het mogelijk om te zoeken op meerdere inhoudelijke veldnamen. Dit wordt verder beschreven in hoofdstuk 3.2.

De volgende zoekvraag zoekt bijvoorbeeld naar publicaties met het woord “algemene” in de titel, met een maximum van 10 resultaten:

<https://zoek.officielebekendmakingen.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=cvdr&startRecord=1&maximumRecords=10&query=title=algemene>

2. **De webservice retourneert een XML response met daarin één van de twee mogelijke berichten:**

2a. XML met zoekresultaten. Dit zoekresultaat kan worden gebruikt om te presenteren aan een gebruiker of het kan worden gebruikt voor verdere verwerking. Dit wordt verder beschreven in hoofdstuk 3.4.

2b. XML met een foutmelding.

De webservice is gebaseerd op de **internationale open standaard** Search & Retrieval by URL (SRU). Het is mogelijk om **real time** te zoeken met de webservice. Op het gebruik van deze webservice is een **fair use policy** van kracht. Voor meer detail informatie over het opstellen van de zoekvraag (stap 1), de xml response (stap 2) en de foutafhandeling (stap 2b) is er een technische handleiding beschikbaar die is bedoeld voor ontwikkelaars van partijen die de data van Overheid.nl willen ontsluiten: Handleiding 'Bevragen van de zoekdienst via SRU' (in PDF-formaat).

3.1 Opbouw van de zoekvraag



Om een zoekvraag te stellen, dient er een URL opgebouwd te worden die alle zoekcriteria bevat. Een voorbeeld hiervan is onderstaande zoekvraag om de eerste 10 publicaties van het type “Wet” op te vragen van het “Ministerie van Veiligheid en Justitie”

```
https://zoek.officielebekendmakingen.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=oep&query=%28type=%22Wet%22%20and%20creator=%22Ministerie%20van%20Veiligheid%20en%20Justitie%22%29
```

Door op de bovenstaande URL te klikken kan worden bekeken hoe een XML response eruit ziet. De URL voor het stellen van een zoekvraag bestaat uit een basis en een aantal parameters. De basis URL betreft:

```
https://zoek.officielebekendmakingen.nl/sru/Search
```

Met deze URL wordt het SRU koppelvlak aangesproken. De belangrijkste parameters in dit koppelvlak zijn:

Parameter	Verplicht	Vaste waarde	Toelichting
version	Ja	1.2	Versie van de SRU standaard
operation	Ja	searchRetrieve	Geeft aan dat de actie een zoekvraag betreft
x-connection	Ja	oep	Naam van de gegevensverzameling waarin gezocht wordt

Parameter	Verplicht	Vaste waarde	Toelichting
startRecord= &maximumRecords=	Nee	-	<p>Deze parameters worden doorgaans in combinatie gebruikt om een beperkte set resultaten op te vragen, bijvoorbeeld resultaat 11-20 van de in totaal 600 zoekresultaten.</p> <p>Met startRecord wordt aangegeven vanaf welk record de zoekresultaten moeten worden teruggegeven. Indien deze parameter niet meegegeven wordt, dan krijgt deze standaard de waarde 1.</p> <p>Met maximumRecords wordt aangegeven hoeveel zoekresultaten er totaal moeten worden teruggegeven. Indien deze parameter niet meegegeven wordt, dan krijgt deze standaard de waarde 10. Het maximale aantal zoekresultaten voor een query ligt op 4020.</p>
query	Ja	-	<p>De daadwerkelijke zoekvraag in de vorm van een 'CQL query' (Contextuel Query Language), bijvoorbeeld "creator=Amsterdam and title=Algemene". Zie de volgende paragraaf voor een toelichting.</p>

Er zit een maximum aan het aantal zoekresultaten dat teruggegeven wordt (maximaal 4200). Indien er meerdere zoekresultaten zijn dan kunnen deze allemaal opgehaald worden door dezelfde zoekvraag te herhalen waarbij het startrecord net zo lang opgehoogd wordt totdat alle zoekresultaten teruggegeven zijn.

Bijvoorbeeld:

1. Stel een zoekvraag aan het SRU koppelvlak met startRecord=1 en maximumRecords=100.
2. Indien nog niet alle resultaten zijn opgehaald (Er zijn meer dan 100 zoekresultaten), verander dan de startRecord-parameter naar 101 en stel de vraag opnieuw:
3. Als vervolgens er weer meer dan 100 zoekresultaten zijn stel dan de zoekvraag opnieuw waarbij de startRecord-parameter aangepast is naar 201.
4. Herhaal dit totdat alle records zijn opgehaald waarbij dus elke keer de startRecord-parameter wordt aangepast naar "startRecord oud + maximumRecords".

3.2 CQL query

Voor het uitvragen van de collectie Officiële Publicaties kunnen zoekvragen worden gesteld in CQL (Contextual Query Language). De zoekcriteria hebben betrekking op de veldnamen van de collectie Officiële Publicaties. De CQL query staat in de URL na query=. In de CQL query kan volstaan worden met de naam zonder namespace prefix, dus bijvoorbeeld `title=Algemene` is gelijk aan `dcterms.title=Algemene`.

Voor het uitvragen van de collectie Officiële Publicaties kan er gezocht worden op een aantal veldnamen (zie bijlage 1). Het is bijvoorbeeld mogelijk om alle publicaties van een bepaalde publicerende organisatie op te zoeken door te zoeken op “creator”. Om dus bijvoorbeeld alle publicaties van de gemeente Utrecht op te zoeken kan de volgende zoekquery gebruikt worden: <https://zoek.officielebekendmakingen.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=oep&query=creator=Utrecht>.

De zoekvelden betreffen generieke zoekvelden en niet publicatie specifieke zoekvelden (zoals OP-ID). Als het gewenst is om te zoeken naar één of twee specifieke publicaties, dan dient het Raadpleegsysteem op de website <https://zoek.officielebekendmakingen.nl> gebruikt te worden. De query interface is hier niet voor bedoeld.

De zoekresultaten kunnen oplopend of aflopend gesorteerd worden op publicatiedatum (available) door achter de zoek query het volgende toe te voegen: “ sortby available/sort.descending” of “ sortby available/sort.ascending” e.g.
 “<https://zoek.officielebekendmakingen.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=oep&startRecord=1&maximumRecords=10&query=keyword=verleende%20sortby%20available/sort.descending>”

3.2.1 Niet gevoelig voor hoofdletters, accenten, etc.

Het maakt niet uit of een zoekvraag met of zonder accenten, koppeltekens, diakrieten, hoofdletters of kleine letters gesteld wordt. Ofwel: Een zoekopdracht voor 'cafe' vindt ook 'café', 'CAFÉ' en 'Café'.

3.2.2 Operatoren

De volgende operatoren etc. kunnen worden gebruikt in CQL :

Relations

=	gedeeltelijke match, tekst komt voor in het betreffende veld
==	volledige match, tekst komt overeen met de waarde in het betreffende veld
<, >, <=, >=	vergelijking, van toepassing op datum-velden
adj	resultaten bevatten de de opgegeven zoektermen in exact de volgorde als opgegeven: een zogenaamde "exact phrase search"
all	resultaten bevatten de opgegeven zoektermen in willekeurige volgorde
any	resultaten bevatten in ieder geval een van de opgegeven zoektermen

Booleans

and	resultaten moeten aan beide criteria wordt voldaan
or	resultaten moeten aan minstens één van beide criteria voldoen
not	resultaten moeten níet aan het criterium achter 'not' voldoen

Wildcards

*	wildcard
?	wildcard voor 1 willekeurig karakter in een string

Let op: het gebruik van aanhalingstekens betekent hier dus geen *exact search phrase* zoals dat voor andere zoekmachines vaak wel het geval is. Hiervoor wordt de 'adj' (adjacent) relatie gebruikt.

Zie <http://www.loc.gov/standards/sru/resources/cql-context-set-v1-2.html> voor een uitvoeriger beschrijving.

3.2.3 Complexere zoekvragen

Full-text zoeken

Om full-text te zoeken, kan de volgende zoekopdracht gebruikt worden:

...&query=keyword=fiets

Zoekopdracht op meerdere veldnamen

Om verkeerde interpretatie van queries te voorkomen wordt aanbevolen criteria te groeperen met behulp van haakjes. Waarden kunnen het best tussen dubbele quotes (" ") in de query worden opgenomen:

...&query=(title="Algemene Plaatselijke Verordening") and (dcterms.creator=Utrecht)

3.2.4 Sorteren

De zoekresultaten worden standaard gesorteerd op relevantie. Deze wordt bepaald op basis van het voorkomen van zoektermen in metadata en/of de bodytekst.

Voor het sorteren op een specifieke veldnaam kan de parameter sortBy gebruikt worden:

...&query=keyword=fiets sortBy dcterms.modified/sort.descending

Sortby kan niet op alle veldnamen gebruikt worden. Voor de collectie Officiële Publicaties kan dit gebruikt worden met alle velden behalve met Titel en Alternatieve titel.

3.3 Uitvragen via HTTP GET of HTTP POST

Het is mogelijk om de webservice via HTTP GET of HTTP POST uit te vragen. SOAP wordt niet ondersteund. Zie <http://www.loc.gov/standards/sru/specs/transport.html> voor meer informatie.

3.4 Het zoekresultaat: De opbouw van een XML response

De webservice geeft een zoekresultaat terug in XML conform de SRU specificaties. De opbouw van dit XML wordt in dit hoofdstuk beschreven.



Indien de SRU query niet begrepen wordt door de zoekdienst, of als er een fout optreedt, dan volgt een standaard SRU foutbericht. Dit wordt in detail beschreven in de Handleiding 'Bevragen van de zoekdienst via SRU'.

Het XML zoekresultaat bevat ieder afzonderlijk gevonden record (resultaat), extra informatie per record en extra informatie over de hele resultaatset. Ten behoeve van de leesbaarheid zijn de namespaces in de voorbeelden weggelaten. De generieke SRU elementen vallen in de namespace "http://www.loc.gov/zing/srw/". In hoofdlijnen ziet een zoekresultaat er als volgt uit (merk op: de regelnummers staan niet in het resultaat: deze worden gebruikt om naar te verwijzen):

```

1. <searchRetrieveResponse>
2.   <version>1.2</version>
3.   <numberOfRecords>14</numberOfRecords>
4.   <records>
5.     <record>...</record>
6.     <record>...</record>
7.     ...
8.   </records>
9.   <nextRecordPosition>11</nextRecordPosition>
10.  <extraResponseData>
11.  </extraResponseData>
12. </searchRetrieveResponse>
  
```

Alle elementen in bovenstaand overzicht zijn standaard SRU elementen.

Regel	Omschrijving
1	<searchRetrieveResponse> is het root element.
2	Het element <version> heeft betrekking op de versie van de SRU standaard.
3 & 9	Het element <numberOfRecords> geeft het totaal aantal gevonden resultaten weer. Dit staat los van het aantal records in het XML bericht dat wordt bepaald door gebruik van 'maximumRecords' in de query. Wordt in de zoekvraag bijvoorbeeld <code>maximumRecords=10</code> gespecificeerd, en zijn er 14 resultaten, dan bevat <numberOfRecords> de waarde 14. Het aantal records in het XML bericht wordt niet expliciet opgenomen, maar kan bepaald worden door het aantal <record> elementen te tellen. Indien van toepassing wordt het element <nextRecordPosition> (regel 9) opgenomen, waarmee wordt verwezen naar het eerstvolgende record in de resultaten set. In het voorbeeld zou dat 11 zijn.
4	Het element <records> is het moederelement van alle <record> elementen.
5-7	Ieder <record> bevat 1 zoekresultaat. Dit element wordt in de volgende paragraaf toegelicht.
10	Het element <extraResponseData> bevat extra informatie op het niveau van de hele resultaatset. Hieronder vallen onder meer facetten. Dit element wordt verder beschreven in de technische Handleiding 'Bevragen van de zoekdienst via SRU' (in PDF-formaat).

Elk record element bevat informatie over het gevonden resultaat. Deze bestaat uit een aantal generieke SRU-elementen (in het overzicht hieronder dikgedrukt) en een aantal specifiek voor de zoekdienst.

```

1. <record>
2.   <recordSchema>http://standaarden.overheid.nl/sru/
   </recordSchema>
3.   <recordPacking>xml</recordPacking>
4.   <recordData>
5.     <gzd>
6.       <originalData>
7.         ...
8.       </originalData>
9.       <enrichedData>
10.      ...
11.    </enrichedData>
12.    </gzd>
13.  </recordData>
14.  <recordPosition>1</recordPosition>
15. </record>

```

Regel	Omschrijving
1	Ieder <record> bevat een gevonden resultaat.
2	Ieder <record> voldoet aan een <recordschema>.
3	De <recordPacking> is voor de zoekdienst altijd 'XML' en niet 'string'.
4	Onder het element <recordData> staat de metadata van het record.
5	Het <gzd> element staat voor Gemeenschappelijke Zoekdienst. Dit is het root element voor de metadata.
6 – 8	<p>Onder <originalData> staat de data in OWMS-formaat zoals deze ook in de collectie Officiële Publicaties staat. Het element <originalData> bestaat uit de elementen <owmskern> (met daarin verplichte metadatavelden volgens OWMS), <owmsmantel> (met optionele metadatavelden volgens OWMS) en een collectiespecifiek element.</p> <p>De volgende gegevens worden meegegeven in de owmskern:</p> <ul style="list-style-type: none"> - dcterms:title: de naam van de publicatie - dcterms:identifier: het unieke id van de publicatie - dcterms:type: het soort publicatie - dcterms:creator: de partij die de eindverantwoordelijkheid draagt voor de publicatie - dcterms:spatial: - dcterms:temporal: <p>De volgende gegevens worden meegegeven in de owmsmantel:</p> <ul style="list-style-type: none"> - dcterms:alternative: de alternatieve titel van de publicatie - dcterms:available: de publicatiedatum - dcterms:issued: - dcterms:source: - dcterms:subject: - dcterms:organisationtype: het type publicerende organisatie - dcterms:publicationname: de publicatiesoort waarin de publicatie is opgenomen - dcterms:jaargang: de jaargang van de publicatiesoort waarin de publicatie is opgenomen - dcterms:publicationissue: het nummer van de publicatiesoort waarin de publicatie is opgenomen - overheidop:idgeometrie: - overheidvb:referentienummer: - overheidvb:typeverkeersbesluit: - overheidvb:weggebruiker: - overheidvb:wegcategorie:
9–11	<p>In het element <enrichedData> staan veldnamen die niet als element in de broncollectie staan, maar daar wel uit afgeleid kunnen worden.</p> <p>De volgende gegevens worden hierin meegegeven:</p> <ul style="list-style-type: none"> - locationURI: het pad naar de betreffende publicatie op de ftp server

Regel	Omschrijving
14	In het element <recordPosition> wordt de positie van het record binnen de totale resultaten set getoond.

3.5 Ophalen gegevens

In de zoekresultaten staat het element locationURI. Met behulp van deze locatie kan de betreffende publicatie opgehaald worden op de ftp server bestanden.officiëlebekendmakingen.nl. Voor deze ftp server wordt gebruik gemaakt van passive FTP waarop via anonymous FTP ingelogd kan worden. Voor meer informatie hierover zie hoofdstuk 4.

3.6 Externe bijlagen

Het kan voorkomen dat een publicatie één of meerdere externe bijlagen bevat. Indien deze aanwezig zijn, is voor elke externe bijlage een element OVERHEIDop.externeBijlage aanwezig met daarin de identifier van de bijlage. Het pad naar de externe bijlage kan als volgt opgebouwd worden: <https://{domain}/{area}/{work}/currentItem>. Hierbij geldt dat:

- Domain = de naam van het domein waar de publicatie repository is geregistreerd. Dit betreft altijd de waarde: "repository.officiële-overheidspublicaties.nl".
- Area = de naam van folder van de externe bijlage waarnaar verwezen wordt. Dit betreft altijd de waarde: "externebijlagen".
- Work = de identifier van de betreffende externe bijlage.
- CurrentItem: hiermee wordt verwezen naar het actuele item, zoals dat binnen de repository is geadministreerd volgens de metadata.

Met behulp van dit pad kunnen de betreffende externe bijlagen opgehaald worden uit de repository.

4 Bulk UitleverSysteem (BUS)

Voor de systematische afname van officiële publicaties is een voorziening ingericht gebaseerd op SFTP genaamd het Bulkuitleversysteem. Deze voorziening is bereikbaar via bestanden.officielebekendmakingen.nl. Het bulk uitleversysteem biedt de mogelijkheid om officiële publicaties in grote aantallen tegelijkertijd te kunnen afnemen. Hiermee is het mogelijk om periodiek (bijv. dagelijks) alles van de collectie Officiële Publicaties op te halen óf alleen de nieuw toegevoegde en gewijzigde publicaties sinds de vorige ophaalactie. Er kan zelf bepaald worden welke selectie wordt opgehaald.

4.1 Bron

Het Bulkuitleversysteem wordt dagelijks gevoed vanuit het publicatiemechanisme dat ook www.officielebekendmakingen.nl en zoek.officielebekendmakingen.nl voedt. Alle publicaties die op zoek.officielebekendmakingen.nl staan, behalve de privacygevoelige publicaties, zijn beschikbaar op het Bulkuitleversysteem.

4.2 Tijdstip van vullen

Het vullen van het Bulkuitleversysteem gebeurt in het publicatieproces. Dit betekent dat regulier gezien het Bulkuitleversysteem gevuld wordt met alle officiële bekendmakingen tussen 9 uur en 10 uur 's ochtends. In uitzonderlijke gevallen wordt een bekendmaking ook later op de dag pas geplaatst.

Parlementaire documenten worden gedurende de gehele dag bijgeplaatst, in eerste instantie in onopgemaakte vorm die later vervangen zullen worden door de opgemaakte stukken.

Elke wijziging leidt tot het bijwerken van een metabestand in XML. Dit bestand ([changelog.xml](#)) zegt welke bestanden op welk tijdstip zijn aangepast. Meer informatie hierover volgt in paragraaf 4.5.

4.3 Toegang

Het Bulkuitleversysteem is vrij toegankelijk voor afnemers. Om gebruik te maken van deze optie dient er ingelogd te worden op de ftp server bestanden.officielebekendmakingen.nl. Er wordt gebruik gemaakt van passive FTP waarop via anonymous FTP ingelogd kan worden.

Het Bulkuitleversysteem is versleuteld via het SFTP-protocol (SSH File Transfer Protocol). Deze versleuteling kent twee doelen: allereerst zorgt deze versleuteling middels een certificaat ervoor dat de afnemer met zekerheid kan vaststellen dat het hier inderdaad om bestanden.officiëlebekendmakingen.nl gaat. Ten tweede is het niet mogelijk voor derden om te traceren welke officiële bekendmakingen worden afgenomen.

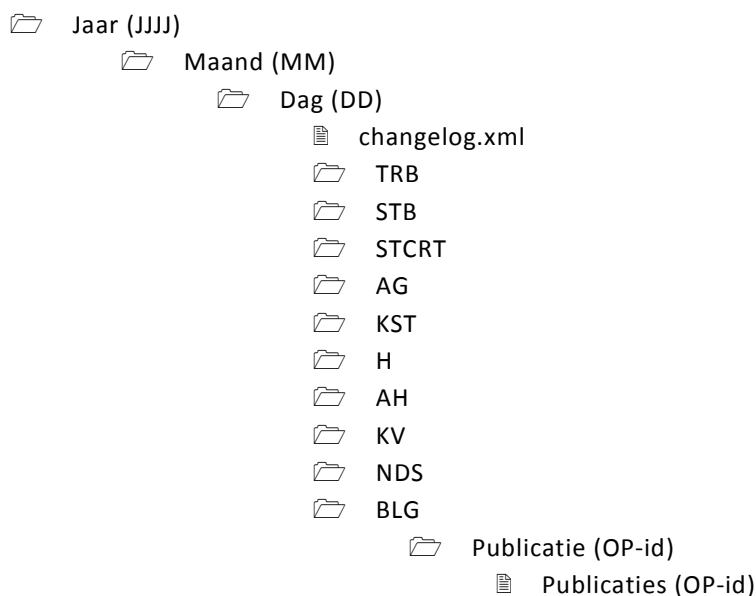
4.4 Mappenstructuur

Nadat er ingelogd is, wordt de root directory gepresenteerd. Vervolgens is het mogelijk om alle bestanden rechtstreeks te benaderen en te downloaden door de juiste FTP URL aan te roepen.

De inhoud van het BUS is opgebouwd volgens een standaardconventie. Deze conventie omvat zowel de opbouw van de mappenstructuur alsook een metabestand om de wijzigingen per dag aan te geven.

In het Bulkuitleversysteem worden bestanden geplaatst in een structuur gebaseerd op publicatiedatum en de publicatiebladen (Tractatenblad, Staatsblad, Staatscourant, Kamerstukken, Handelingen, etc.). De bestanden zullen per identifier in een aparte map opgenomen worden. Per identifier zijn alle bestandsformaten opgenomen als [identifier].[formaat]. De PDF, XML en ODT spreken hierbij voor zich. De XHTML is opgenomen in een zipbestand met daarin de afbeeldingen die voorkomen in het bestand.

De bestandsstructuur waarin de publicaties beschikbaar zijn is als volgt:



De bestandsstructuur is onderverdeeld op datum, eerst de jaren, dan de maanden en de dagen. Daarna is er per dag een map voor iedere publicatiesoort:

- TRB = Tractatenblad
- STB = Staatsblad
- STCRT = Staatscourant
- AG = Agenda's

- KST = Kamerstukken
- H = Handelingen
- AH = Kamervragen met antwoorden (Aanhangsels)
- KV = Kamervragen zonder antwoorden
- NDS = Niet-dossierstukken
- BLG = Bijlagen

Een voorbeeld is:

/2012/11/06/stb/stb-2012-538/

In deze directory staan de volgende bestanden:

stb-2012-538.pdf

stb-2012-538.xml

stb-2012-538.html.zip

stb-2012-538.odf

metadata_owms.xml

metadata.xml

Het systeem logt van elke transactie welke bestanden zijn gedownload, het aantal bytes dat is gestuurd, de tijd die het gekost heeft en het IP-nummer. Het loggen vindt plaats om twee redenen:

- het verzamelen van statistische gegevens om het gebruik van het BUS dienst te analyseren en eventueel aan te passen;
- het opsporen van eventuele fouten bij gemelde incidenten.

Het is mogelijk dat het systeem de toegang weigert omdat de server te druk bezet is. Het systeem weigert dan de login met een melding hiervan. In dit geval wordt aangeraden om het op een ander tijdstip opnieuw te proberen.

4.5 Metabestand Changelog.xml

Bij elke dag wordt een changelog.xml metabestand bijgehouden. Dit bestand beschrijft welke wijzigingen op de betreffende dag zijn gepubliceerd. Het beschrijft dus per dag alle toevoegingen van nieuwe bestanden, maar ook aanpassingen (of verwijderingen) die op deze dag hebben plaatsgevonden van bestanden die in het verleden geplaatst zijn. Dit bestand is altijd aanwezig en heeft altijd dezelfde naam. De meest recent behandelde publicaties staan bovenaan.

Het changelog.xml bestand bevat de volgende informatie:

- datum en tijd laatste aanpassing changelog.xml bestand
- per publicatie:
 - locatie verschijningsvorm (zowel XML als PDF, HTML en ODT; als die aanwezig is)
 - actie m.b.t. de publicatie

Bij elke aanpassing van het changelog.xml bestand wordt de datum en tijd aangepast. Dit geldt ook voor changelog.xml bestanden die in een map van een datum uit het verleden staan. Hierdoor kan altijd bepaald worden of de meest recente versie van het changelog.xml bestand is geraadpleegd.

.....

In het changelog.xml bestand kunnen de volgende acties voorkomen:

- toegevoegd
- veranderd
- verwijderd

In bijlage 2 is als voorbeeld een stuk van een changelog meegestuurd.

4.5.1 Toegevoegd

Alle publicaties die voor het eerst worden gepubliceerd krijgen als actie “toegevoegd”. Als de publicatie in een oudere map is gezet, dan wordt dit vermeld in zowel het changelog.xml bestand van de map van vandaag als in het changelog.xml bestand die in de oudere map staat.

4.5.2 Veranderd

Het kan voorkomen dat door aanpassingen, in bijvoorbeeld de metadata, een reeds gepubliceerde publicatie opnieuw wordt aangeboden. In dit geval wordt in de map van de originele publicatie de aangepaste verschijningsvormen vervangen en wordt in zowel de map van vandaag als de map van de originele publicatie, in het `changelog.xml` bestand de locatie van de originele publicatie en de actie “veranderd” opgenomen.

Voorbeeld:

Op 3 juli 2008 wordt een wijziging voor *kst-27624-r1677-3.xml* gedateerd 12 januari 2006 ontvangen.

Het BUS overschrijft in de map */2006/01/12/kst/st-27624-r1677-3/* de oude *kst-27624-r1677-3.xml* met de nieuwe *kst-27624-r1677-3.xml*. In de map */2008/07/03/* wordt een changelog.xml aangemaakt als er nog geen is, anders wordt de bestaande aangevuld (bovenaan), met als actie “veranderd” en de locatie. Tevens wordt in de map */2006/01/12/* het changelog.xml bestand aangevuld (bovenaan), met de locatie en als actie “veranderd”.

4.5.3 Verwijderd

Er zijn publicaties die kunnen worden gedepubliceerd en dus worden verwijderd van het BUS. In dit geval wordt de juiste map verwijderd en wordt in zowel de map van vandaag als in de map waaruit het bestand is verwijderd, in het changelog.xml bestand de locatie van de originele publicatie en de actie “verwijderd” opgenomen.

4.6 Gebruik dagelijkse updates

Een manier om het Bulkuitleversysteem te gebruiken is als volgt:

1. Log in op de sftp-faciliteit bestanden.officielebekendmakingen.nl
2. Pak de changelog.xml van de voorgaande dag (/jaar/maand/dag/changelog.xml)
3. Werk alle bestanden bij die in het changelog.xml genoemd worden.
4. Herhaal stappen 1 t/m 3 iedere dag.

Met deze werkwijze wordt gegarandeerd dat er geen content gemist wordt. Hierbij is de versheid van de bestanden wel maximaal een dag later. Het is natuurlijk ook mogelijk om het huidige changelog na 10.00 uur op te halen, maar dan moet er een vergelijking gemaakt worden tussen deze versie en de versie die aan het einde van de dag op het Bulkuitleversysteem staat, aangezien er gedurende de dag wijzigingen hebben kunnen plaatsvinden.

5 Vragen of opmerkingen over dit document

Indien u vragen heeft over dit document, dan kunt u contact opnemen met de KOOP servicedesk via het emailadres: servicedesk@koop.overheid.nl.

Bijlage 1 – Mogelijke zoekvelden

In onderstaande tabel staan alle velden waarop binnen OEP gezocht kan worden.

Functionele naam	Parameter in CQL query	Opmerkingen
-	keyword	Zoeken op vrije tekst in alles
Titel	title	Zoeken op vrije tekst in de titel Hier kan niet op gesorteerd worden
Alternatieve titel	alternative	Zoeken op vrije tekst in de citeertitel of redactionele titel Hier kan niet op gesorteerd worden
Rubriek of documenttype	type	Toegestane waarden: OVERHEIDop.Staatsblad OVERHEIDop.Staatscourant OVERHEIDop.Tractatenblad OVERHEIDop.Parlementair
Publicatiesoort, bijvoorbeeld Staatsblad of Staatscourant	publicationName	OVERHEIDop.publicationName
Type publicerende organisatie bijvoorbeeld Gemeente	organisationType	
De publicerende instantie, bijvoorbeeld Ministerie van Buitenlandse Zaken	creator	Toegestane waarden zijn alle waarden uit de volgende waardelijsten (zie http://standaarden.koop.asp4all.nl/owms/3.5/doc/waardelijsten/): OVERHEID.Adviescollege OVERHEID.Deelgemeente OVERHEID.Gemeente OVERHEID.HoogCollegeVanStaat OVERHEID.Ministerie OVERHEID.OpenbaarLichaamVoorBedrijfe nBeroep OVERHEID.Politiekorps OVERHEID.Provincie OVERHEID.Rechterlijkemacht OVERHEID.RegionaalSamenwerkingsorgaa n OVERHEID.Waterschap OVERHEID.ZelfstandigBestuursorgaan OVERHEIDop.DienstAgentschapInstelling OfProject

Functionele naam	Parameter in CQL query	Opmerkingen
Onderwerp / Categorie	Category	Onderwerp volgens de activiteitenindex of de onderwerpscodering van Buitenlandse Zaken (specifiek voor Tractatenbladen) DCTERMS.subject met 1 van de volgende schema's: OVERHEID.TaxonomieBeleidsagenda OVERHEID.TaxonomieBuZa
Datum totstandkoming van een verdrag (specifiek voor Tractatenbladen)	datumTotstandkoming	Datumveld (DD-MM-YYYY of YYYY-MM-DD)
Publicatiedatum	available	Datumveld (DD-MM-YYYY of YYYY-MM-DD)

Bijlage 2 – Voorbeeld metabestand changelog.xml

Voorbeeld van het volgende changelog.xml bestand /2008/07/02/changelog.xml:

```
<datetime>2008-07-02T14:54</datetime>
<changes>
  <change>
    <paths>
      <path>/2007/04/13/stb/stb-2007-4/stb-2007-4.pdf</path>
      <path>/2007/04/13/stb/stb-2007-4/stb-2007-4.xml</path>
      <path>/2007/04/13/stb/stb-2007-4/stb-2007-4.html.zip</path>
      <path>/2007/04/13/stb/stb-2007-4/stb-2007-4.odt</path>
      <path>/2007/04/13/stb/stb-2007-4/metadata_owms.xml</path>
      <path>/2007/04/13/stb/stb-2007-4/metadata.xml</path>
    </paths>
    <action>toegevoegd</action>
  </change>
  <change>
    <paths>
      <path>/2008/07/02/stb/stb-2008-5/stb-2008-5.pdf</path>
      <path>/2008/07/02/stb/stb-2008-5/stb-2008-5.xml</path>
      <path>/2008/07/02/stb/stb-2008-5/stb-2008-5.html.zip</path>
      <path>/2008/07/02/stb/stb-2008-5/stb-2008-5.odt</path>
      <path>/2008/07/02/stb/stb-2008-5/metadata_owms.xml</path>
      <path>/2008/07/02/stb/stb-2008-5/metadata.xml</path>
    </paths>
    <action>toegevoegd</action>
  </change>
  <change>
    <paths>
      <path>/2007/12/10/stc/stc-2007-12/stc-2007-12.pdf</path>
      <path>/2007/12/10/stc/stc-2007-12/stc-2007-12.xml</path>
      <path>/2007/12/10/stc/stc-2007-12/stc-2007-12.html.zip</path>
      <path>/2007/12/10/stc/stc-2007-12/stc-2007-12.odt</path>
      <path>/2007/12/10/stc/stc-2007-12/metadata_owms.xml</path>
      <path>/2007/12/10/stc/stc-2007-12/metadata.xml</path>
    </paths>
    <action>veranderd</action>
  </change>
  <change>
    <paths>
      <path>/2008/07/02/stc/stc-2008-6/stc-2008-6.pdf</path>
      <path>/2008/07/02/stc/stc-2008-6/stc-2008-6.xml</path>
      <path>/2008/07/02/stc/stc-2008-6/stc-2008-6.html.zip</path>
      <path>/2008/07/02/stc/stc-2008-6/stc-2008-6.odt</path>
      <path>/2008/07/02/stc/stc-2008-6/metadata_owms.xml</path>
      <path>/2008/07/02/stc/stc-2008-6/metadata.xml</path>
    </paths>
    <action>toegevoegd</action>
  </change>
</changes>
```



```
</change>
<change>
  <paths>
    <path>/2008/07/01/nds/nds-tk-2008D40334/nds-tk-
2008D40334.pdf</path>
    <path>/2008/07/01/nds/nds-tk-2008D40334/nds-tk-
2008D40334.xml</path>
    <path>/2008/07/01/nds/nds-tk-2008D40334/nds-tk-
2008D40334.html.zip</path>
    <path>/2008/07/01/nds/nds-tk-2008D40334/nds-tk-
2008D40334.odt</path>
    <path>/2008/07/01/nds/nds-tk-
2008D40334/metadata_owms.xml</path>
    <path>/2008/07/01/nds/nds-tk-2008D40334/metadata.xml</path>
  </paths>
  <action>verwijderd</action>
</change>
<change>
  <paths>
    <path>/2008/01/14/ah/ah-128031/ah-128031.pdf</path>
  </paths>
  <action>toegevoegd</action>
</change>
<change>
  <paths>
    <path>/2008/06/14/kst/kst-23412-14/kst-23412-14.pdf</path>
  </paths>
  <action>vervanging</action>
</change>
<change>
  <paths>
    <path>/2008/06/14/kst/kst-315784/kst-315784.pdf</path>
  </paths>
  <action>verwijderd</action>
</change>
<change>
  <paths>
    <path>/2008/06/14/kst/kst-23412-14/kst-23412-14.xml</path>
    <path>/2008/06/14/kst/kst-23412-14/kst-23412-
14.html.zip</path>
    <path>/2008/06/14/kst/kst-23412-14/kst-23412-14.odt</path>
    <path>/2008/06/14/kst/kst-23412-14/metadata_owms.xml</path>
    <path>/2008/06/14/kst/kst-23412-14/metadata.xml</path>
  </paths>
  <action>toegevoegd</action>
</change>
</changes>
```