

Kwaliteitsbeleid metadata ontsloten in het NGR

In het overleg van het GI-beraad van juni 2014 is het verbeterplan metadata besproken.

De voorzitter concludeert dat het belang van metadata door iedereen onderschreven wordt en tegelijk nog verbetering behoeft. De monitoring, uitgevoerd door Geonovum, blijft daarom gehandhaafd en zal met de voortgang van maatregelen ter verbetering aangevuld worden. Eén van de maatregelen is om een kwaliteitsbeleid op te stellen. Voorliggend memo is het voorstel.

Inleiding

Geografische informatie infrastructures zorgen er niet alleen voor dat gegevens effectief en efficiënt worden ingewonnen maar stimuleren en faciliteren ook het gebruik van deze gegevens. Het gebruik van gegevens van verschillende organisaties kent interoperabiliteitsproblemen. Datasets en services worden steeds vaker alleen via portalen aangeboden. Metadata is zo steeds meer de enige toegang tot deze datasets en services om ze te verkrijgen en te gebruiken. Meestal is er geen communicatie meer met de beheerders van de datasets en services. Dit stelt hogere eisen aan de kwaliteit van de metadata. Daarnaast vereist het automatisch verwerken van metadata in portalen een nauwgezette beschrijving van metadata; een machine kan een foutje in een code niet interpreteren.

Dit memo definieert het kwaliteitsbeleid voor het gebruik van metadata en daaraan gekoppelde datasets en services in het NGR. Dit memo geeft inzicht in de procedure om tot structurele kwaliteitsverbetering te komen.

Vertrekpunt

Sinds oktober 2012 wordt er een structurele monitoring door Geonovum op de metadata van het NGR uitgevoerd. Uit de monitoring¹ van december 2013 en juni 2014 blijkt dat de helft van de organisaties gemiddeld onder de 65 (van de 100) scoort. Aanpassingen lijken niet structureel te worden doorgevoerd, waardoor de kwaliteit over een langere periode schommelt.

De monitoring wordt uitgevoerd op een beperkt aantal metadata elementen die essentieel zijn voor het **gebruik** van de beschreven data. Het gaat daarbij met name om de gebruiksvoorwaarden, links naar de datasets en services en mailadres. Deze elementen maken deel uit van het Nederlands metadata profiel.

Problemen met het gebruik van metadata

Bij het gebruik van informatie over datasets en services in het NGR en andere portalen speelt een aantal (technische) problemen. De problemen zijn onder andere zichtbaar in niet werkende links naar de datasets en services en het ontbreken van gebruiksvoorwaarden. Daarnaast wordt er in veel gevallen niet gereageerd op mailtjes naar het contactadres. De dataset of service is daardoor niet te gebruiken of te downloaden en er is geen mogelijkheid om in contact te komen met de dataprovider.

Uit reacties op de monitoring blijkt dat organisaties die data aanbieden, het lastig vinden om voor de volle 100% valide te zijn en te blijven. Het gaat om een redelijke hoeveelheid metadata die op orde gehouden moet worden en daarnaast neemt de complexiteit met verwijzingen en gebruik van identifiers toe.

Rol van gebruikers

¹ https://www.pdok.nl/sites/default/files/bibliotheek/ngr_metadata-monitor_juni2014.pdf



Gebruikers van het NGR hebben ook een rol bij het verbeteren van de kwaliteit van de metadata; ze kunnen problemen die ze signaleren direct doorgeven aan de dataproviders met de feedbackvoorziening. Daarnaast kunnen ze metadata waarderen door een vijf sterren systeem, waarop ook gesorteerd kan worden. De monitoring van metadata kwaliteit die op de metadata in het NGR wordt uitgevoerd, kan data providers verder ondersteunen bij het verbeteren en in stand houden van de metadata kwaliteit.

Het voorstel²

De verantwoordelijkheid voor de inhoud van de metadata blijft onveranderd liggen bij de dataprovider. De monitoring³ wordt vier keer per jaar uitgevoerd op een beperkt aantal metadata elementen die essentieel zijn voor het **gebruik** en bijdragen aan de **vindbaarheid** van de beschreven data. Bij een score onder de norm, zie bijlage A, kan ze niet worden gebruikt. Om de data via de metadata weer bereikbaar te maken worden onderstaande stappen uitgevoerd;

- Als een metadata beschrijving onder de norm scoort, wordt de NGR contactpersoon van die organisatie per mail geïnformeerd dat de kwaliteit niet voldoende is en gewezen op mogelijke consequenties daarvan; de data is niet vindbaar en bruikbaar.
- De organisatie wordt verzocht binnen drie maanden hun metadata te verbeteren. Verbeterpunten worden aangegeven in de monitoringsrapportage. Als ondersteuning om te kunnen voldoen aan de criteria gewenst is, wordt dat door Geonovum gegeven⁴.
- Als na drie maanden, bij de volgende monitoring blijkt dat de verbetering niet of niet voldoende is geweest wordt persoonlijk contact gezocht met die organisatie en ondersteuning geboden bij verbetering.
- Als vervolgens na drie maanden (dus een half jaar na eerste signalering) wederom de score en daarmee de vindbaarheid en bruikbaarheid niet substantieel verbeterd is, wordt de metadata niet meer getoond in het NGR of gemarkeerd als onjuist/onvolledig.
- Metadata die voldoende is aangepast wordt weer getoond in het NGR, de metadata beheerder is zelf in staat de metadata weer te publiceren.
- Aparte afspraken zijn nodig met de provincies en mogelijk andere organisaties, omdat zij gezamenlijk de metadata aanleveren middels harvesting.
- De tooling voor het berekenen van de scores kan onvoorziene resultaten geven. In deze gevallen wordt in overleg met de dataprovider naar een oplossing gezocht.
- De ondersteuning wordt alleen gegeven op het verduidelijken van de standaarden en de daaruit voortvloeiende benodigde aanpassingen van de metadata. De metadata beheerder is zelf verantwoordelijk voor het doorvoeren van de aanpassingen in de metadata.

Geonovum bewaakt *het kwaliteitsniveau* van de metadata in het NGR, zolang dit niet door gebruikers rechtstreeks of een community wordt opgepakt. Door ook te sturen op het vergroten van het gebruik van NGR en gebruikersfeedback aan dataproviders, is er voor dataproviders een grotere prikkel om de kwaliteit op peil te houden.

² Dit voorstel is in consultatie geweest. Reacties van het PBL, gemeente Delfzijl, Alterra/WUR, BIJ12 en het Kadaster zijn hierin verwerkt.

³ Metadata monitor NGR, Handleiding voor kwaliteitsverbetering metadata
<http://www.geonovum.nl/sites/default/files/Handleiding%20metadata%20monitoring%20v0.3.pdf>

⁴ Geonovum besteedt hieraan in 2015 maximaal 10 dagen. Deze zijn onderdeel van de opdracht strategisch tactisch advies PDOK.



Bijlage A; de norm

De norm wordt gebaseerd op een beperkt aantal metadata elementen die essentieel zijn voor het **gebruik**. Het gaat daarbij met name om de gebruiksvoorwaarden, links naar de datasets en services, het protocol en mailadres. Deze elementen geven een maximale score van 65 bij dataset metadata en 72 bij service metadata, zie bijlage B en C. Deze elementen in de metadata moeten volledig en correct zijn. Titel, samenvatting en trefwoorden dragen in sterke mate bij aan de **vindbaarheid** van data. Deze elementen geven een maximale score van 17 bij dataset metadata en 6 bij service metadata. Deze scores bij elkaar bepalen de norm; Bij een score onder de 82 (van de 100) voor dataset metadata en 78 bij service metadata is de data slecht te vinden via de metadata en kan ze niet worden gebruikt.

**Bijlage B; Metadata elementen van datasets en te behalen score**

Metadata elementen van datasets	Controle Omschrijving	Score goed	Score matig	Score slecht
Overige beperkingen	<p>Moet voldoen aan een opendata of Geo Gedeeld licentie. Zie http://www.geonovum.nl/geostandaarden/gebruiksvoorwaarden</p> <p>Het is van belang dat hier een geldige licentie wordt ingevuld. U heeft de keuze tussen de Publiek Domein, de Creative Commons Zero (CC0) of een Geo Gedeeld licentie (inclusief geldige URL naar uw Geo Gedeeld licentiebepalingen).</p>	20	0	0
Overige beperkingen	Moet een korte beschrijving bevatten over de url zoals geogedeeld of geen beperkingen	5	0	0
(Juridische) Toegangsrestricties	Het is verplicht om het veld juridische toegangsrestricties te vullen met de waarde 'otherrestrictions'. Een andere domeinwaarde is ongeldig.	2	0	0
(Juridische) gebruiksrestricties	Het wordt aanbevolen om het veld met de Juridische gebruiksrestrictie leeg te laten of te voorzien van 'otherrestrictions'. Als meer keer other constrains of andere waarden voorkomen is de beoordeling slecht	2	0	0
Protocol	<p>Is een geldig protocol opgegeven voor de URL waarmee de dataset wordt gedistribueerd? (bijv. download, website, OGC:WMS of OGC:WFS Zie ook bijlage 5.1)</p> <p>Als er een 'Online Resource' URL (check 2) is gekoppeld is de opname van een protocol verplicht (en vice versa).</p> <p>Als er geen url aanwezig is en er is geen protocol ingevuld, dan wordt protocol goed gekeurd en krijgt dus ook 10 punten</p>	10	0	0



Metadata elementen van datasets	Controle Omschrijving	Score goed	Score matig	Score slecht
URL	<p>Verwijzen alle bronnen naar geldige URL's (WMS/WFS/download/etc)?. Bij services dient de basis-URL van de service te worden ingevoerd (dus exclusief specifieke requests zoals ?request=GetCapabilities en protocolaanduidingen zoals &service=WMS).</p> <p>uitzonderingen</p> <p>1) Een score slecht als er wel een URL is opgegeven, maar geen metadata element protocol</p> <p>2) Als check 5_protocol de waarde dataset bevat dan mag de url een intern link zijn met meer dan 10 karakters en wordt deze check als goed beoordeeld</p> <p>3) Als er geen url aanwezig is en geen metadata element protocol is ingevuld wordt deze check als goed beoordeeld</p>	20	0	0
Verantwoordelijke organisatie metadata: e-mail	Zijn de contactgegevens per organisatie consistent ingevoerd? De contactgegevens moeten voorzien zijn van een geldig e-mail adres waarbij bij voorkeur een functioneel e-mail adres wordt gebruikt. (bijv. beheerder@organisatiennaam.nl)	10	5	0
Titel van de dataset	Een titel heeft tussen de 3 en 75 karakters.	2	0	0
Trefwoorden	De trefwoorden dienen per stuk in een afzonderlijk veld te zijn opgeslagen (dus niet komma gescheiden in hetzelfde veld).	10	1	0
Omgrenzende rechthoek (Extent dataset)	Valt de extent van de dataset (rechthoek) binnen de Nederlandse grenzen inclusief (NCP) en is de schrijfwijze van coördinaten correct.	2	0	0
Samenvatting	Is de samenvatting aanwezig en is de samenvatting niet te lang. Bevat tekst tussen de 25 en 2000 karakters.	5	0	0
Unieke identifier van de bron	Is de UUID voor de dataset aanwezig en uniek? Gebruik bij voorkeur geen accolades en hoofdletters.	6	1	0
Metadata unieke identifier	Is de UUID voor de metadata aanwezig en uniek? Gebruik bij voorkeur geen accolades en hoofdletters.	6	1	0

**Bijlage C; Metadata elementen van services en te behalen score**

Metadata elementen van services	Controle Omschrijving	Score goed	Score matig	Score slecht
OtherConstraints	Moet voldoen aan een opendata of Geo Gedeeld licentie. Zie http://www.geonovum.nl/geostandaarden/gebruiksvoorwaarden Het is van belang dat hier een geldige licentie wordt ingevuld. U heeft de keuze tussen de Publiek Domein Mark, de Creative Commons Zero (CC0) of een Geo Gedeeld licentie (inclusief geldige URL naar uw Geo Gedeeld licentiebepalingen).	20	0	0
OtherConstraints	Moet een korte beschrijving bevatten over de url zoals geogedeeld of geen beperkingen	5	0	0
AccessConstraints	Het is verplicht om het veld juridische toegangsrestricties te vullen met de waarde 'otherrestrictions'. Een andere domeinwaarde is ongeldig.	1	0	0
UseConstraints	Het wordt aanbevolen om het veld met de Juridische gebruiksrestrictie leeg te laten of te voorzien van 'otherrestrictions'. Als meer keer other constrains of andere waarden voorkomen is de beoordeling slecht	1	0	0
Service Type	Is een correct 'Service Type' gebruikt uit de domeinlijst uit het NL Profiel 1.2 voor services? Als voorbeeld moet hier de tekst 'view' staan als het een WMS service betreft.	5	0	0
Resource locator	Is er een geldige URL opgenomen in Resource locator? Is deze benaderbaar vanaf het internet?	15	0	0
Metadata point of contact	Zijn de contactgegevens per organisatie consistent ingevoerd? De contactgegevens moeten voorzien zijn van een geldig e-mail adres waarbij bij voorkeur een functioneel e-mail adres wordt gebruikt. (bijv. beheerder@organisatienaam.nl)	2	1	0
Resource title	Een titel heeft tussen de 3 en 75 karakters.	2	0	0
Keyword value	De trefwoorden dienen per stuk in een afzonderlijk veld te zijn opgeslagen (dus niet kommagescheiden in hetzelfde veld).	2	1	0
Resource abstract	Is de samenvatting aanwezig en is de samenvatting niet te lang. Bevat tekst tussen de 25 en 2000 karakters.	2	0	0



Metadata elementen van services	Controle Omschrijving	Score goed	Score matig	Score slecht
ConnectPoint Linkage	Er wordt in de monitor gecontroleerd of de URL van de 'Connect Point Linkage' vanaf het internet benaderbaar is. Daarnaast moet deze identiek zijn aan de inhoud van het veld 'Resource locator' (Zie: check 6)	25	10	0
Coupled resource	Het veld 'Coupled resource' is verplicht als er links naar de datasets waarop de service opereert beschikbaar zijn. Voor dataservices is het dus verplicht. Dit element geeft informatie (metadata) over de data die in de service beschikbaar wordt gesteld. Het veld coupled resource bestaat uit 2 attributen. Het attribuut xlink:href bevat een URL naar de metadata XML van de dataset die in de service is ontsloten. Deze link moet werken en geeft de metadata van de dataset terug waarop de service gebaseerd is. In het attribuut UUID is de identifier van de bron opgenomen (de dataset UUID). Deze komt dus overeen met de unieke identifier van de bron in de metadata van die dataset.	15	5	0
Metadata identifier	Is de UUID voor de metadata aanwezig en uniek? Gebruik bij voorkeur geen accolades en hoofdletters.	5	3	0