Kabinetsnota 'Informatie op Orde'

Van dementeren naar informeren

MARTIJN DE GROOT

Alle ministeries gaan in de komende jaren hun digitale informatiehuishouding op orde brengen. Daarnaast moeten de papieren archieven versneld worden bijgewerkt. Dat staat in de nota 'Verbetering Informatie- en Archiefbeheer Rijksoverheid', die het kabinet op 29 juni naar de Tweede Kamer heeft gestuurd.

Het stuk, ook wel aangeduid als de nota *Informatie op Orde,* is nog afgeleverd door de bewindslieden Pechtold en Van der Laan uit het tweede kabinet-Balkenende, met als aanleiding onder meer de nota 'Een dementerende overheid?' van de toenmalige Rijksarchiefinspectie. In het najaar zal het in de Tweede Kamer worden besproken.

De digitale informatiehuishouding van de overheid dreigt 'onbeheersbare proporties' aan te nemen en veel overheidsorganisaties kampen met een enorme achterstand in hun papieren archieven. Bij de eerste constatering speelt de 'digitale revolutie' een belangrijke rol; voor de tweede ontbreekt een rechtvaardiging. Woordvoerder Maarten Gresnigt van het ministerie van OCW wijst desgevraagd op de 'sneltreinvaart' waarmee de ambtenaren digitale uitingsvormen tot hun beschikking hebben gekregen, 'terwijl het informatie- en archiefbeheer bij de departementen van oudsher nog een papieren proces is. Digitale en papieren processen hebben lange tijd naast elkaar bestaan. Daarbij heeft de papieren informatiehuishouding geleidelijk minder aandacht gekregen'.

De problemen moeten volgens de nota nu wel worden aangepakt want 'overheidsinformatie moet vindbaar en toegankelijk zijn maar ook betrouwbaar, authentiek en volledig'.

Het inhalen van de achterstand in de papieren archieven lijkt nog het eenvoudigste voornemen. Een Rijksbreed project Wegwerken Archiefachterstanden t/m 1975 was al in 2005 van start gegaan en moet op 1 januari 2009 klaar zijn. Vanaf 2008 worden de 'werkvoorraden' van na 1975 onder handen genomen op basis van werkplannen die de ministeries dan klaar hebben.

Het aanpakken van de digitale archieven en vooral het aanpassen van de manier van werken opdat nieuwe stukken automatisch goed worden gearchiveerd is veel moeilijker, blijkt uit de visienota. Voor dat laatste doel heeft het kabinet volgens de nota op 1 januari 2008 een set van minimale kwaliteitseisen gereed – een z.g. *baseline* – die moet worden gecontroleerd door de departementale auditdiensten. Ook moet er een nieuwe aanpak worden ontwikkeld van de selectie van documenten en met name metadata die ervoor kunnen zorgen dat cultuurhistorische selectie al bij de vorming van documenten mogelijk wordt. Die metadata worden onderdeel van de *baseline*.

Voor dit alles zijn nieuwe regels nodig. Een deel van die regels kan in de vorm van een convenant tussen de departementen worden afgesproken. Een ander deel zal bestaan uit het op elkaar afstemmen van bestaande wetten en regels zoals de Wet Openbaarheid Bestuur, de privacywetgeving en de Archiefwet 1995. Hiernaar wil het kabinet vanaf het voorjaar van 2007 een onderzoek doen.

Voor informatie van wetenschappelijk belang heeft de nieuwe visie van het kabinet geen aparte gevolgen, verzekert het ministerie van OCW. Gresnigt: 'De nota richt zich op de kwaliteit van het informatie- en archiefbeheer van de rijksoverheid in het algemeen, ongeacht de vorm van de gegevens, en daarmee ook op een betere vindbaarheid en toegankelijkheid, ongeacht de gebruikersgroep'.

DANS: 'Belangrijke wijziging in beleid'

Data-instituut DANS ziet de nota *Informatie op Orde* als een belangrijk signaal dat de overheid zich de kwetsbaarheid van digitale data op lange termijn realiseert. Voor het wetenschappelijk onderzoek is dat besef van groot belang omdat speciaal historisch, maar ook veel ander onderzoek vaak op informatie van de overheid berust. Heiko Tjalsma, senior expert research data van DANS ziet belangrijke wijzigingen ten opzichte van het bestaande beleid. 'Erkend wordt dat digitale documenten niet meer achteraf geselecteerd moeten worden, maar dat dit vooraf moet gebeuren, direct bij de vorming. Dat is wezenlijk want anders ben je te laat; na enige tijd kunnen de bestanden al onbruikbaar of onbegrijpelijk zijn geworden. Bovendien komt er bij de selectie nu meer aandacht voor de uitvoeringsdossiers en niet alleen voor het overheidsbeleid zelf. En juist in die uitvoeringsdossiers zit vaak een schat aan informatie voor historisch of sociaalwetenschappelijk onderzoek'.

Een ander punt is vooral voor DANS als nationale datacoach belangrijk: 'Er wordt nu ingezien dat het in het digitale tijdperk niet uitmaakt welke overheidsorganisatie de data beheert. Onderzoekbestanden van de overheid kunnen dus heel goed in een *trusted digital repository* worden bewaard, bij DANS of aan DANS gelieerd'.

Alleen Rijksoverheden

De nota Informatie op Orde geldt voor de Rijksoverheid: de departementen en de Hoge Colleges van Staat (zoals de Raad van State) en voor diensten zoals de Belastingdienst of Rijkswaterstaat. OCW-woordvoerder Gresnigt: 'Uitvoerende diensten vallen onder het beleid in zoverre ze onderdeel zijn van een ministerie. Over de verantwoordelijkheden en de uitvoering maken de ministeries afspraken met die diensten'.

De Zelfstandige Bestuurs Organen (zoals de Autoriteit Financiële Markten) vallen niet onder de nota, maar de ministeries waaronder ze vallen zullen die organen volgens de nota wel 'uitnodigen om de problemen op gelijke wijze aan te pakken en op te lossen'.

Letters, woorden en teksten: pijlers onder de data-infrastructuur

RENÉ VAN HORIK

Taal en literatuur worden op verschillende manieren wetenschappelijk bestudeerd. Zowel het materiaal als de instrumenten worden steeds vaker in elektronische vorm aangeboden. Het internet brengt efficiënte communicatie tussen onderzoekers en afstemming van onderzoeksinstrumenten binnen bereik. Enkele pijlers onder de benodigde data-infrastructuur komen hier voor het voetlicht.

De Digitale Bibliotheek voor de Nederlandse Letteren (DBNL: www.dbnl.org), opgericht vanuit de Maatschappij der Nederlandse Letterkunde, biedt een zeer goede website over de Nederlandse taal en literatuur. De site bevat literaire teksten, secundaire literatuur en aanvullende informatie zoals biografieën en portretten. Ook de digitale collecties van universiteitsbibliotheken en de Koninklijke Bibliotheek maken deel uit van de data-infrastructuur voor literair en taalkundig wetenschappelijk onderzoek. Het DBNL-project werkt samen met een groot aantal wetenschappers afkomstig van verschillende disciplines en vakgroepen.

Schatbewaarder

Naast het onderzoek dat aan de letterenfaculteiten van universiteiten wordt verricht bestaan er enkele toonaangevende instellingen in Nederland op het gebied van taalkundig en literair wetenschappelijk onderzoek: het Huygens Instituut, het Meertens Instituut, het Max Planck instituut voor psycholinguïstiek en het Nederlands Instituut voor Lexicologie (INL). Het Huygens Instituut (www.huygens.knaw.nl/) is geinteresseerd in de toepassing van nieuwe methoden en technieken in ontsluiting en tekstwetenschappelijk onderzoek. Op het gebied van de Nederlandse taal bestudeert het Meertens Instituut (www.meertens.knaw.nl) zowel de geografische als sociale variatie. Het Max Planck Instituut voor psycholinguïstiek (www.mpi.nl) richt zich op onderzoek naar de mentale processen gerelateerd aan het begrip en gebruik van taal. Het Instituut voor Nederlandse Lexicologie (www.inl.nl) beschouwt zich als de schatbewaarder van de Nederlandse taal door bijvoorbeeld de Taalbank bij te houden, waarin alle Nederlandse woorden voorkomen. Deze en andere instellingen zijn betrokken bij de creatie en het gebruik van een groot aantal digitale databronnen en onderzoeksinstrumenten, zoals *parsers* en *lemmatisers*.

Tekstcorpora zijn van groot belang voor linguïstisch onderzoek en worden gebruikt om theorieën te toetsen. Er bestaat al een behoorlijk aantal corpora, maar deze zijn niet volledig qua type en tijdspanne. Een corpus dient verrijkt te worden om de onderzoeksvragen te kunnen beantwoorden. Het Corpus Gesproken Nederlands (CGN) is bijvoorbeeld een belangrijk onderzoekscorpus. Daarnaast zijn er diverse taaldatabases, bijvoorbeeld op het gebied van dialecten, zoals de Syntactische Atlas van de Nederlandse Dialecten (SAND), te vinden via de website van het Meertens Instituut.

De Digitale Bibliotheek voor de Nederlandse Letteren biedt een zeer goede website over de Nederlandse taal en literatuur.

Scala van onderzoek

Databronnen en onderzoeksinstrumenten zoals deze maken een breed scala aan wetenschappelijk onderzoek mogelijk. Historici richten zich met hun onderzoek bijvoorbeeld op de dynamische relatie tussen taal en maatschappij. Taalveranderingen treden immers vaak op waar contact is geweest tussen verschillende talen of verschillende variaties van een taal. Ook kan het perspectief van het cultureel erfgoed centraal staan in onderzoek. Meer fundamenteel linguïstisch onderzoek kan zich bezig houden met de veranderingen van de syntaxis van het Nederlands, bijvoorbeeld op het gebied van het gebruik van persoonlijke voornaamwoorden.

In de literatuurstudie speelt de data-infrastructuur een belangrijke rol bij het vernieuwen van de bestaande kennis. Daarbij moet rekening worden gehouden met de relatie tussen eigentijdse oordelen en de literaire geschiedenis. Hiervoor is een corpus van secundaire literaire bronnen nodig die gebruikt kunnen worden om literaire teksten te kunnen evalueren. Literaire tijdschriften zijn hierbij van het grootste belang, want zij vormen de 'barometer van de smaak'.

Goede fundering

Door de computerondersteunde auteurschapidentificatie wordt steeds vaker toegepast bij de analyse van gedigitaliseerde teksten. Het onderzoek richt zich dan op stijlkenmerken van auteurs door gebruik te maken van speciale algoritmen. Zo kan bijvoorbeeld vastgesteld worden wie de auteur is van een anoniem pamflet.

De data-infrastructuur op het gebied van taalkunde en literatuurwetenschap heeft een goede fundering en maakt een breed scala aan wetenschappelijk onderzoek mogelijk. Meerjarige onderzoeksprogramma's, zoals *Stevin* (www.nwo.nl/stevin/) op het gebied van taal en spraaktechnologie en *Catch* (www.nwo.catch) voor verbeterde toegang tot digitaal cultureel erfgoed, dragen bij aan verbeterde data-infrastructuur voor wetenschappelijk onderzoek. Uitwisseling van expertise, standaardisatie van datacreatie, data-opslag en data-uitwisseling zijn gebieden die meer aandacht vragen, waardoor de data-infrastructuur nog beter zou kunnen worden.