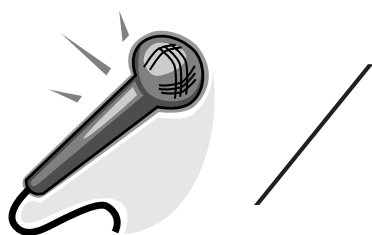


# Tolvrije digitale toegang tot de Sonttolregisters



George Welling en Jan Willem Veluwenkamp



Wanneer je tussen 1497 en 1857 met je laadruim vol goederen door de Sont voer en je betaalde geen tol aan de Deense koning, dan werd je vanaf slot Kronborg bestookt met kanonskogels. Als je dan toch nog levend aan wal kwam, mocht je ook nog eens de kosten van de kanonskogel betalen. De meeste schippers droegen wel keurig hun tol af en streken hun grootzeil als eerbetoon aan de Deense koning. Dat heeft een schat aan gegevens voor onderzoek opgeleverd: de Sonttolregisters.

**D**e Sonttolregisters bevatten de administratie van de tol die werd geheven door de ambtenaren van de koning van Denemarken”, zo begint Jan Willem Veluwenkamp, universitair docent geschiedenis van de vroegmoderne tijd (1500-1800) zijn verhaal. Samen met universitair docent Informatiekunde George Welling is hij een van de drijvende krachten achter het digitaliseren van de Sonttolregisters.

“Het vroegst bekende jaar is 1497 en de tol is geheven tot 1857. De tol werd geheven op alle schepen die door de Sont voeren, een smalle zee-engte tussen Denemarken en Zweden die de Noordzee en de Oostzee met elkaar verbindt. In de administratie staat de naam van de schipper, zijn woonplaats, de haven van vertrek, de haven van bestemming, welke goederen hij aan boord had en hoeveel tol hij daarover moest betalen. De naam van het schip vonden ze niet interes-

sant om op te nemen in de registers. Kijk, zij waren geïnteresseerd in de tolopbrengst, die moest berekend worden aan de hand van de spullen aan boord. Het schip was voldoende geïdentificeerd met de schipper en de datum van de doorvaart. Als het om een groot schip ging, en men zei weinig aan boord te hebben, dan was dat aanleiding om eens even te gaan kijken.

### **Aanleiding**

Tresoar, het Fries Historisch en Letterkundig Centrum, heeft een jaar of tien geleden de microfilms van de Sonttolregisters gekocht van de Statens Arkiver in Denemarken. Dit omdat ze dachten dat het een mooie bron voor Friese schippers was; die waren prominent in die tijd. In gesprekken met allerlei mensen is toen het idee ontstaan om de registers digitaal beschikbaar te gaan stellen.

Maar wat ga je dan doen? Ga je alleen voor de 18e eeuw alleen de Friezen uit die films in een database stoppen? Dat leek me niet zo handig want die hebben toch een context, een Nederlandse. Nederland heeft ook een context, dus toen kwamen we met NWO in gesprek. Wat zijn nu de mogelijkheden? Uiteindelijk hebben we gezegd: we digitaliseren alles. We hebben toen 1,2 miljoen euro gekregen van NWO uit het programma dat tot dan toe eigenlijk alleen maar werd gebruikt voor wetenschappelijke apparatuur, sterrenkijkers, medische scanners en dergelijke. We zaten net bij de eerste of tweede tranche van aanvragen die er gebruik van maakten dat ook het bouwen van dataverzamelingen als instrument werd beschouwd, zodat ook de humaniora van de subsidie konden profiteren. In 2008 hebben we de subsidie verworven en in 2009 zijn we aan het werk gegaan.

Het project is aangevraagd door de RUG en Tresoar. We zijn gelijkwaardige projectpartners, maar ik ben hoofdaanvrager bij NWO, met handtekening van het College van Bestuur natuurlijk. Maar ik heb het project dus niet verzonden, dat is in gesprekken met Tresoar, allerlei mensen uit het land en met George ontstaan. In de NWO-aanvraag moest een architectuur



van de database worden vermeld. Die kon ik niet maken, dus toen is George er al bij betrokken geraakt.”

### **Database**

“Van oorsprong ben ik economisch-sociaal historicus”, neemt George Welling het verhaal over. “Ik heb mijn dissertatie over het paalregister van Amsterdam geschreven. Het project was voor mij dus in meerdere opzichten interessant.

In wezen is de database niet complex. Hij bestaat uit een aantal hoofdtabellen en hulptabellen. Het is vooral veel. Als je bijvoorbeeld tegen elkaar af wilt zetten in een kruistabel hoeveel vaarten er hebben plaatsgevonden van welke plaats naar welke plaats over 100 jaar, dan kun je je misschien voorstellen dat er nogal veel variabelen komen. Dat loopt dan het interne werkgeheugen uit. Als hij dan tot het punt komt dat-ie moet swappen naar disk en tijdelijk dingen op schijf moet zetten, omdat hij het niet meer in zijn interne geheugen kan houden, dan wordt het lastig. We proberen ten koste van al-

les te voorkomen dat dat gebeurt. En dat houdt in dat je mensen sommige dingen niet mag toestaan. In principe kan het wel, maar je kunt het beter in stukken doen.

Naast de hoofdtabellen in de database is er een hulptabel waar van iedere doorvaart is vastgelegd op welke scan van de film hij staat. Soms staat een doorvaart op een enkele scan, soms staat hij op meerdere scans. Dan zijn er nog hulptabellen voor de standaardisatie. We hebben een tabel waarin alle voorkomende vormen van plaatsnamen zitten en iedere voorkomende vorm van de plaats wordt herleid naar een code. Die code wordt herleid naar de standaardvorm in een van de talen die we gebruiken. Je kunt zoeken op plaatsnamen zoals ze in de bron voorkomen en op de gestandaardiseerde versie van de plaatsnaam, bijvoorbeeld København.

De standaardisering van de geografische namen is mijn levenswerk zo ongeveer geworden. Voor nog niet alle plaatsen heb ik de juiste standaardvorm. Ook heb ik besloten dat we alleen die plaatsen accepteren waarvan we de



# 'er zitten nu

## 450.000

# namen in de database'

coördinaten daadwerkelijk kunnen vaststellen. Die kun je dan doorverbinden met Google Maps.

We hebben de database pas een tijdje geleden openbaar gemaakt. Tot die tijd was het echt in ontwikkeling en liepen de processen ook wel eens vast. Als je zo'n database volgens de regels van de kunst helemaal genormaliseerd in elkaar wilt zetten, heb je hulptabellen nodig om het allemaal voor elkaar te krijgen. Dat ging allemaal veel te langzaam. We hebben daarom nu ook een dedicated server waar alleen deze database op staat. De response is aanzienlijk sneller geworden. Als je kruistabellen gaat maken met een half miljoen records wordt het exponentieel erger als je dingen tegen elkaar af gaat zetten. We hebben eindeloos gezocht om het sneller te krijgen, het moest toch kunnen. Uiteindelijk hebben we de database platgeslagen en redundante informatie toegevoegd. De responsetijd is nu redelijk, in 4/5 sec. kun je een kruistabel eruit krijgen. Eerst was dat meer dan een halve minuut. Dan is iedereen vertrokken.

### Nabewerkingen

"De oorspronkelijke invoer is handgeschreven Deens, over meer dan 3 eeuwen. En die Vroegmoderne tijd kende geen vaste spelling. Dus alles wat erin zit, plaatsnamen, persoonsnamen, goederennamen, namen van maten, dat is op allerlei verschillende manieren gespeld. Je kunt voor de stad Groningen misschien wel 15 of 20 verschillende varianten tegenkomen. Dat is voor alle plaatsnamen, persoonsnamen en goederennamen zo. Als je iets uit een database wilt halen, moet je zoektermen invullen. Als je "Groningen" invult, krijg je inderdaad wel alle records met de naam Groningen, maar niet die met alle varianten. Dus dat betekent dat je resultaat eigenlijk zinloos is. In de nabewerking moet je dan in een extra veld de standaard toevoegen. Met dat standaardiseren voor de geografische namen ben ik al een eind gevorderd, er zitten nu 450.000 namen in de database. De productnamen wachten nog, de persoonsnamen is een ander verhaal. Een volgend probleem is dan nog: we hebben maten en gewichten die

erg variëren. Je zou op een gegeven moment lasten met schipponden en met ponden kunnen weergeven. Daar zit nog een extra probleem aan, want de maten zijn de maten van de haven van afvaart. Dus een last uit Hamburg is helaas niet hetzelfde als een last uit Amsterdam."

"Een roggelast uit Hamburg", preciseert Jan Willem, "want we hebben het hier ook nog over roggelasten, houtlasten en teerlasten. Dat is om het simpel te houden. Dus als je het over nabewerking hebt: we zouden natuurlijk graag willen zeggen zoveel kilo teer, zoveel ton teer in kilo's uitgedrukt, maar nu staat het er nog in tonnen of vaten. Als je werkelijk uiteindelijk de gebruiker helemaal ten dienste zou willen zijn, dan moet je alles helemaal standaardiseren.

### Instrument

Dat neemt niet weg dat voor dat dat gebeurd is die database ook nuttig is voor heel veel mensen. Natuurlijk voor liefhebbers die voorouders zoeken of de geschiedenis van hun eigen plaats of regio willen bekijken. Dat is toch een grote groep mensen. Professionals die onderzoek doen in deze bronnen, kunnen naar Denemarken gaan en de bladzijden omslaan of ze kunnen

werken met dit instrument dat we hebben gemaakt, ook al is het dan nog niet volmaakt door dat gebrek aan standaardisering. Als je geïnteresseerd bent in een bepaalde goederenstroom in een bepaalde periode tussen bepaalde plaatsen, dan kun je dat binnenhalen. Voor jou set kun je dan bewerkingen gaan doen. Dat hoort bij het historisch métier."

"Ja", bevestigt George, "en die 450.000 die er nu inzitten zijn gestandaardiseerd voor wat de plaatsnamen betreft. Daar ben ik echt met de hand helemaal doorheen gegaan. We dachten dat het een dalende curve zou zijn met langzaam maar zeker steeds minder varianten. Maar de tijd leidt weer tot andere schrijfwijzen. De politieke verhoudingen in het Oostzeegebied veranderden ook nog al eens. In Finland hebben bijvoorbeeld een boel plaatsen Zweedse en Finse namen, de Zweden hebben een tijdje de Noord-Duitse kust in handen gehad. Dat geeft ook weer andere plaatsnamen. Een Russische kapitein die een Duitse haven beschrijft, enzovoort. Gelukkig werden de betaalmiddelen doorgaans uitgedrukt in Lübische daalders, die bestonden weer uit 24 schellingen.

### Tijdrovend

Het derde punt: de mensen die het voor ons zit-ten in te voeren, van de sociale werkplaats Breed in Nijmegen, zijn onderhand redelijk geschoold in het lezen ervan, maar we hebben ze gevraagd op te schrijven wat ze zien en daar komen soms ook weer de raarste verbasteringen van."

Jan Willem voegt toe: "Het is lastig werk. Het is niet simpel. Je kunt over veel dingen twisten. Ik vind dat Breed goed werk doet."

"Een dag in de week besteed ik aan nabewerking", vertelt George. "Soms ben ik er ook de hele week mee bezig, dan weer even niet. Ik wacht tot er weer zo'n invoerset binnenkomt. Dat gaat meestal per ingevoerde microfilm. Uiteindelijk wordt het in Leeuwarden op de server gezet en beschikbaar gemaakt. Dan vindt de plaatsnaamstandaardisatie nog plaats. Ruim 500.000 doorvaarten zijn er nu beschikbaar. Het is nog lang niet compleet."





“Ik verwacht dat wij over een jaar de periode 1630-1857 online zullen hebben”, stelt Jan Willem Veluwenkamp. “Voor de periode voor 1630 hebben we een ander probleem. Voor dat jaartal noteerden de tollenaars de doorvaart niet in één boeking, maar in meerdere. Gemiddeld in 3 boekingen, voor alle verschillende tollens die ze moesten betalen een andere. In een voor ons soms irrationele indeling. Het aantal records dat we in te voeren hebben over die periode is daardoor verdrievoudigd. Alleen de naam van de schipper verbindt de verschillende boekingen met elkaar en die wordt dan ook nog regelmatig op verschillende wijze geschreven. Voor dezelfde doorvaart kunnen ze de ene boeking op 2 mei zetten, maar als er dan op 4 mei weer een boeking is, is dat dan dezelfde doorvaart, maar op een ander moment geboekt, of is het de terugreis? Dat soort dingen kom je tegen. We moeten er met alle betrokkenen nog maar eens over praten hoe we dat gaan oplossen. Breed zal het invoerwerk dan niet meer doen. Die zal bij 1630 ophouden. We overwegen serieus om dat een vrijwilligersproject te maken als crowdsourcing. We zien daar wel een goede kans. Het moeten mensen zijn die de liefhebberij hebben om dat oude schrift te lezen.

### Tolvrij en toekomstvast

De geschiedenis van de Friese schipperij is nu ook een klein programma binnen NWO”, vertelt Jan Willem. “Een postdoc en twee AIO's doen daar onderzoek naar. Er wordt samengewerkt met een historicus in Duitsland waarbij ze naar Schotse schippers gaan kijken. In 2009 en 2010 hebben we hier symposia gehad met een stuk of 16 sprekers. Komende oktober hebben we weer zo'n bijeenkomst. We zijn nu twee jaar verder en hebben een aanzienlijk beter instrument. Ikzelf ga geen andere onderzoeken doen. Ik houd me bij de Friese schippers en ben van plan me de komende jaren toch voornamelijk bezig te houden met die database. Het vrijwilligersproject, als dat van de grond komt, gaat ook veel aandacht eisen en allerlei nabewerkingen moeten gebeuren. Ik vind dat ik daar energie in moet blijven steken. Anderen kunnen dan leuke onderzoekjes doen. We hebben contact met een student in Rostock die een scriptie schrijft en met allerlei vragen komt. Mensen uit Noord-Engeland die laten horen dat ze bezig zijn met een proefschrift.”

“Ik heb onlangs thuis de films van de paalgelden bekeken”, vertelt George. “Die zijn ooit, toen ik mijn dissertatie schreef, met filmlezers ingevoerd - hoeveel beschadiging dat op film geeft! Het feit dat je het nu digitaal hebt is als conservering zoveel beter. Vooral de elektronisch aangedreven lezers beschadigen de film zo erg. Het origineel mag je al lang niet meer inzien.”

Jan Willem Veluwenkamp voegt toe: “De kracht van ons digitale register is dat je altijd naar het origineel terug kunt. Bij iedere record zit een scan die je met een speciaal programmaatje kunt vergroten. Er zijn al verschillende oplettende gebruikers geweest die meldden: dat heeft u verkeerd.

Er wordt niet betaald voor het gebruik van de database. Dat wilden wij ook graag. Je zou geld kunnen vragen en van de opbrengst nieuwe projecten kunnen financieren, maar zo hebben we dat niet afgesproken en zo doen we dat ook niet. We heffen geen tol. “ ❏

**Wat krijgen we nou?** Dat kan toch niet waar zijn? Het College van Bestuur vindt de Universiteitskrant te duur. De universiteit is niet meer van plan de personeelsadvertenties in de UK te zetten en dus: stoppen ermee! Alsof de UK niets meer is dan de achterkant van de personeelsadvertenties van de RUG! Bij de hamer van Thor: de belediging! Wraak. Actie! En wel nu! Waar zijn de spandoeken? Die moet ik toch nog wel ergens hebben. Verdorie nog aan toe, waar zijn die dingen nou? Nu ja, dan maar wat lakens. Rode verfpot. Wat denkt dat CvB wel. Die tent van ze valt als een incoherent stel



Langs de zijlijn

winkeltjes uit elkaar als er niet iets als een medium bestaat. Een krant die althans de illusie van een eenheid in stand kan houden. Hoe zeg je dat kort en bondig, zoiets? Geen UK, geen RUG!! Zo, die zit. Ha, ha!! Op naar het redactiekantoor.

‘Nee hoor. De UK wordt helemaal niet opgeheven.’ De dame achter de balie kijkt me met grote ogen aan. Een beetje verbouwereerd. Ik laat mijn spandoeken zakken. ‘Maar in de krant stond toch . . . en op jullie eigen website . . . en op de website van RTV-Noord . . . dat het College jullie te duur vindt?’

‘Het College vindt de papieren krant te duur. En dat zijn wij wel met het College eens. Vandaar dat we nu een digitale UK gaan maken. We zien dat als een nieuwe uitdaging. We staan er volledig achter, hoor.’

‘Dus ik hoef niet . . .’

‘Nee hoor, u hoeft niet te demonstreren. Alles is in goed overleg geregeld.’

Ik sla me voor mijn kop. Duizelig verlaat ik het pand. Dat uitgerekend mij dit nu moest overkomen. Wat een blamage. Terwijl ik zelf in januari nog wel in Pictogram geschreven heb: “Tegenwoordig stelt de ontwikkeling van de ICT je voor de opgave over je werk zelf na te gaan denken. En ook nog eens over de rest van je leven. Alle journalisten - om maar bij dat voorbeeld te blijven - kunnen hier over meepraten. Wat is een krant, tegenwoordig? Niemand die het nog precies weet.’

Dan krijg ik eens een keertje gelijk en dan heb ik dat weer eens niet door. Zo wordt het natuurlijk nooit wat met mezelf.

Ik heb (soms) gelijk!!



• [www.soundtoll.nl](http://www.soundtoll.nl)