Basissysteem HISGIS voor historische geo-informatie wil landelijk worden

THIJS HERMSEN

Sinds 1997 werkt de Fryske Akademy met diverse partners aan het project HISGIS (HIStorisch Geografisch InformatieSysteem), dat streeft naar 'een basissysteem voor het opslaan en ontsluiten van lokaliseerbare historische en landschapshistorische informatie voor heel Nederland'. Na Friesland, Groningen en Overijssel werd onlangs Utrecht toegevoegd. Meer provincies moeten volgen.

HISGIS wil zoveel mogelijk Nederlandse historische gegevens met een ruimtelijke component beschikbaar stellen voor iedere geïnteresseerde. Dat gebeurt *bottom up* met de kleinst mogelijke geografische bouwstenen: de percelen. Daardoor is het een tijdrovende en kostbare onderneming, die echter met elke uitbreiding een grotere waarde zal krijgen. HISGIS is voor iedereen te raadplegen: www.hisgis.nl. Wetenschappers kunnen voor hun eigen onderzoek in principe de beschikking krijgen over de kaarten en basisgegevens.

Minuutplans

Het fundament van HISGIS is het oudste Kadaster van Nederland, dat in 1832 gereed kwam met als doel grondbelasting op een eerlijke manier te heffen. De *minuutplans* van deze vroegste kadastrale kaarten zijn gevectoriseerd – omgezet in elektronische waarden – en vervolgens in het huidige coördinatenstelsel, het Rijksdriehoekstelsel gehangen zodat ze precies op alle mogelijke andere historische en actuele kaarten passen. De bijbehorende gegevens per perceel (over eigenaar, waarde en bodemgebruik) zijn gedigitaliseerd en aan elk object gekoppeld.

Referentiebestand

Volgens Hans Mol, historicus bij de Fryske Akademy en hoogleraar Friese geschiedenis in Leiden, zijn de mogelijkheden van HISGIS zowel technisch als historisch interessant omdat de basiskaart informatie biedt op perceelsniveau. Zo kunnen nieuwe kaarten worden samengesteld die niet alleen de situatie van omstreeks 1832 laten zien maar ook die van voordien. Het cultuurlandschap en de steden in Nederland zijn immers pas na de opmeting door het Kadaster drastisch veranderd. Het digitale bestand van het HISGIS kan daarom ook dienen als fundering en referentiebestand voor het ontsluiten van de vele prekadastrale en negentiende-eeuwse manuscriptkaarten waarover Nederlandse archieven beschikken.

Een ander voordeel is dat ruimtelijke data van voor en na 1832 op de kaart vastgezet kunnen worden. Door vanaf het perceelsniveau te aggregeren naar een hoger niveau zijn daarmee dan, vaak relatief snel, weer nieuwe kaarten te genereren.

Beroepsgroepen in Leeuwarden per pand in 1839. Blauw: arbeiders, roze: dienstboden, groen: rechters.

Demografische indicatoren

Een van de andere mogelijkheden die HISGIS de historisch onderzoeker biedt is het koppelen van de huisnummers van de percelen aan nieuwe gegevens. Dit is bijvoorbeeld gedaan voor Leeuwarden met de volkstelling van 1839 waarin van elke inwoner van de stad naam, geslacht, leeftijd, burgerlijke staat, geboorteplaats, religie, beroep en adres zijn opgetekend. Die data zijn dus ook per huis te groeperen en vervolgens te koppelen aan de huisnummers van de kadastrale percelen van Leeuwarden. Op deze wijze kan inzicht verkregen worden in de ruimtelijke verspreiding van diverse demografische indicatoren. Hans Mol en Peter Ekamper hebben laten zien waar precies in de binnenstad de dichtbevolkte huizen stonden waarin de lagere sociale klassen woonden: verspreid over de stad maar vaker in de buurt van de stadswallen. Wanneer opeenvolgende volkstellingen beschikbaar zijn kan de ontwikkeling van de bevolking nauwkeurig worden gevolgd. Volgens de onderzoekers is deze methode voor veel meer Friese en andere steden in Nederland mogelijk.

Grootschalige basiskaart

Vorig jaar is het systeem dat eerst alleen Friesland omvatte, uitgebreid met Groningen en Overijssel. Op 16 februari van dit jaar kon er Utrecht aan worden toegevoegd. Het streven is echter om elke provincie van Nederland in dit basissysteem op te nemen, zodat op den duur een grootschalige basiskaart 1832 met tal van verrijkingsmogelijkheden voor het hele land ontstaat. De Fryske Akademy wil in samenwerking met provinciale en ook landelijke partijen zoals het Kadaster het systeem verder uitbouwen. Gezien de brede gebruiksmogelijkheden – voor demografen, landschapsdeskundigen, archeologen, erfgoedbeheerders, genealogen, landbouwhistorici enz. – lijkt volgens de Fryske Akademy zo'n nationale aanpak meer dan de moeite van het investeren waard.

Dataset uitgangspunt voor meten wetenschappelijke kwaliteit

JEROEN SONDERVAN

Onlangs zijn er bij Amsterdam University Press (AUP) twee projecten afgerond waarbij een boek en een viertal artikelen, allemaal online gepubliceerd, zijn verrijkt met primaire onderzoekdata. Beide projecten werden als onderdeel van SURFshare 2008-2009 onder meer in samenwerking met DANS (Data Archiving and Networked Services) uitgevoerd. Wat verandert er bij deze nieuwe manier van publiceren?

Een traditioneel artikel waaraan onderzoekgegevens worden toegevoegd wordt ook wel een verrijkte publicatie genoemd. De afgelopen jaren verschijnen er naast online wetenschappelijke artikelen en monografieën, ook steeds meer 'ruwe' onderzoekgegevens in online dataarchieven. Deze archieven kunnen worden ingezet om dwarsverbanden te leggen met traditionele online publicaties. Dergelijke publicaties maken meer kans en kunnen het wetenschappelijke debat daadwerkelijk verder helpen als de artikelen of boeken, maar ook de datasets, zich in een open en tegelijkertijd duurzame publicatieomgeving bevinden. Onderzoekgegevens krijgen hierdoor een heel andere status dan enkele jaren geleden. Toen bleven dit soort datasets vaak achter gesloten deuren, of ze waren simpelweg niet machineleesbaar en duurzaam opgeslagen, waardoor lang niet iedereen toegang tot de data kon krijgen. Met alle afspraken over standaardisering voor dataopslag is dit in de afgelopen jaren sterk verbeterd.

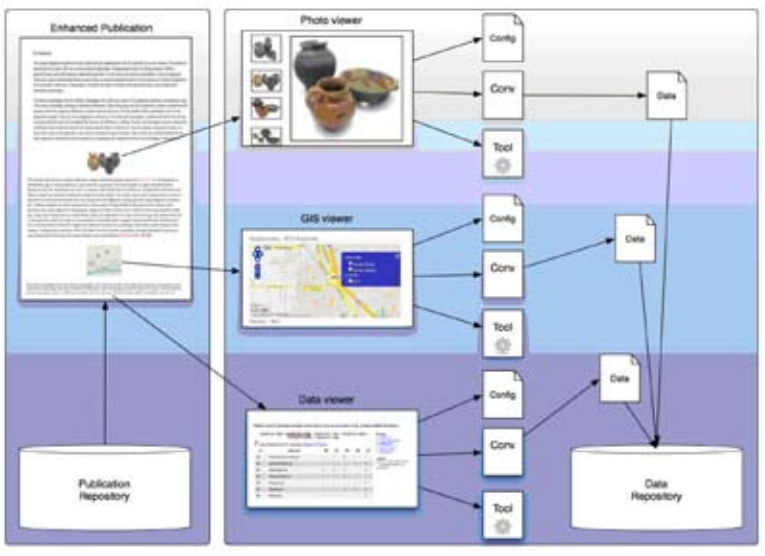
Voor wetenschappelijke uitgeverijen ligt het belang van het verrijken van publicaties in de meerwaarde die dat kan bieden aan de lezers, ergo onderzoekers. Dit roept direct een aantal vragen op, zoals: Aan welke eisen (zowel technisch als inhoudelijk) zou een verrijkte publicatie en de daaraan gekoppelde dataset moeten voldoen? Welke toepassingen wil de lezer kunnen gebruiken? En wat is de impact van dergelijke artikelen voor de verschillende processen (bijvoorbeeld het lezen en het publiceren)? Dit zijn belangrijke vragen waarmee een academische uitgeverij en institutionele repository rekening moeten houden bij het aanbieden van dit soort diensten voor lezers en auteurs. Een nauwe samenwerking met onderzoekers en onderzoekgroepen, institutionele archieven, uitgevers en universiteitsbibliotheken is daarbij een vereiste.

Onderzoekers denken vaak hoofdzakelijk vanuit het traditionele artikel en/of boek en zijn niet gewend om hun ruwe datasets op een (inter)actieve manier in te zetten om hun verhaal kracht bij te zetten. Dit vraagt om een culturele verandering bij de onderzoekers. De infrastructuur om met dergelijke publicaties aan de gang te gaan is aanwezig en zal verder worden uitgebreid. Nu is het aan de onderzoeker om er ook gebruik van te maken. Daarbij is het van groot belang dat verrijkte publicaties zichtbaarder worden en dat deze extra punten opleveren voor de publicatielijst; anders zullen auteurs niet snel extra tijd steken in het geschikt maken van hun dataset.

De vraag is natuurlijk hoe de kwaliteit van datasets is te meten. Artikelen worden door redacties en collega's gereviewd, maar voor datasets is dit nog niet goed geregeld. Er wordt al wel nagedacht over het peer reviewen van data; zie bijvoorbeeld het SURFfoundation-rapport *Verkennend onderzoek Over kwaliteit van onderzoeksdata* uit 2010. Uiteraard worden elders ook initiatieven ontplooid, maar het is van belang dit in de komende tijd goed te coördineren. Zo is het een voorwaarde om de kwaliteitsbeoordeling en duurzame opslag van onderzoekgegevens institutioneel te regelen. Dit kan niet onder de verantwoordelijkheid van uitgevers of andere commerciële partijen vallen. In tegenstelling tot artikelen of boeken (de interpretatie van ruwe datasets door onderzoekers), zijn datasets over het algemeen eigendom van een universiteit of onderzoekgroep, en te belangrijk om niet volgens algemene standaarden en afspraken duurzaam op te slaan.

In plaats van het traditionele publicatiemodel centraal te stellen bij het meten van wetenschappelijke output, zou dus de dataset en daarmee de verrijkte publicatie(s) het uitgangspunt moeten zijn. Alleen dan zullen auteurs bereid zijn hun kostbare tijd te investeren in het integreren van ruwe onderzoekgegevens in hun publicaties. Die toevoeging is een waardevolle 'verrijking' voor disseminatie van onderzoek.

www.surfgroepen.nl/sites/JALC-project/default.aspx
www.watveteranenvertellen.nl/
www.datasealofapproval.org



De verschillende onderdelen van een verrijkte publicatie