



2011 Data Archiving and Networked Services (DANS)

© Sommige rechten zijn voorbehouden / Some rights reserved

Voor deze uitgave zijn gebruiksrechten van toepassing zoals vastgelegd in de Creative Commons licentie. [Naamsvermelding 3.0 Nederland]. Voor de volledige tekst van deze licentie zie <http://www.creativecommons.org/licenses/by/3.0/nl/>

Data Archiving and Networked Services (DANS)

Postbus 93067

2509 AB Den Haag

T 070 3446 484

F 070 3446 482

info@dans.knaw.nl

www.dans.knaw.nl

ISBN 978-94-90531-05-8

Redactie: M. Grootveld, J. van Egmond en B. Sørensen

Vormgeving: Ellen Bouma, Alkmaar

Druk: Bejo druk & print, Alkmaar



1. Aanleiding

Het reviewen van data is een oude wens van de directeur van DANS. Deze latente wens werd acuut na gesprekken met SURFfoundation over datakwaliteit. Zoals de kwaliteitsbewaking van wetenschappelijke publicaties plaats vindt door *peer review*, zo zouden ook datasets gereviewd kunnen worden. Reviews kunnen ook aanleiding geven tot verdere discussie over datasets, en zo bijdragen aan de vorming van communities van onderzoekers rond datacollecties. Natuurlijk zal het reviewen van data anders gaan dan het reviewen van boeken of artikelen. Ook het feit dat het online moet gebeuren zal de wijze van het reviewen van datasets beïnvloeden. Dataset-reviews zullen meer lijken op gebruikers-reviews van producten zoals digitale camera's of van hotels. De hotel-reviews van Booking.com zijn een goed voorbeeld van hoe reviews ingezet kunnen worden.

Eind 2010 is daarom een pilot uitgevoerd onder een aantal afnemers van datasets uit DANS EASY: zij kregen het verzoek om een door hen gedownloade dataset te beoordelen. Dit rapport beschrijft de opzet van de pilot, de uitkomsten en aanbevelingen.

Onlangs is overigens de rapportage verschenen van een – veel breder opgezet – onderzoek naar datakwaliteit in opdracht van SURFfoundation¹. Dit bevat de aanbeveling (p. 6) om te stimuleren dat afnemers ('hergebruikers') commentaar over kwaliteit toevoegen aan datasets.

¹ Graaf, M. van der en L. Waaijers: *Over kwaliteit van onderzoeksdata*. Stichting SURF, december 2010. Te vinden op <http://www.surffoundation.nl/nl/publicaties/Pages/Verken-nendonderzoek.aspx>. In een enquête onder bijna 400 universitaire hoogleraren en universitaire hoofddocenten uit verschillende disciplines zijn o.a. negen mogelijke kwaliteits-toetsen voorgelegd, zoals *peer review* van datasets, trainingen inzake datamanagement en het becommentariëren van de datasets door hergebruikers. Deze laatste optie bleek in alle disciplines een van de drie populairste opties.

2. Opzet van de pilot

Hoewel het bij gebruikers-reviews gebruikelijk is om kort na afname van het product de klant om een beoordeling te vragen, hebben we er bij de pilot voor gekozen om dit met terugwerkende kracht over een langere periode te doen. Dit levert in korte tijd meer beoordelingen op; dat is niet alleen leerzaam, maar ook wenselijk voor het tonen ervan op de website, bij de datasets in kwestie.

2.1 Survey

We hebben gekozen voor een online-survey, omdat dit het best aansluit bij de beoogde manier om de beoordelingen te vergaren. Zoals in andere DANS-onderzoeken is hiervoor SurveyMonkey gebruikt. Afnemers ontvingen een e-mail met, in de algemene tekst, de naam van de betreffende dataset en een unieke link naar de survey. Achter de schermen is de link gelegd tussen de individuele afnemer, de specifieke dataset en de unieke surveylink. Zo is duidelijk op welke dataset bepaalde beoordelingen betrekking hebben.

De vragen waren in SurveyMonkey gerubriceerd naar website, dataset en onderzoek. Deel 3 van dit rapport benoemt alle vragen en de verkregen antwoorden of scores.

2.2 Presentatie van resultaten

Na het sluiten van de survey is een deel van de uitkomsten gepubliceerd in DANS EASY, en wel bij de betreffende dataset. Op dit moment zijn per e-mail ook degenen geïnformeerd die deze datasets oorspronkelijk in DANS EASY hebben gedeponereerd, zodat zij kunnen zien welke beoordelingen anderen aan 'hun' dataset geven.

2.3 Respondenten en responses

Aanvankelijk hebben we een selectie gemaakt van alle personen die in de zes maanden voorafgaand aan de pilot een dataset uit de top-50 hebben gedownload. Die top-50 omvat overigens meer dan 50 datasets, omdat de 50^e en 51^e (t/m 61^e) even vaak zijn gedownload. Deze selectie omvatte

DANSEASY ELECTRONIC ARCHIVING SYSTEM
Thursday, Jan 27 2011, 11:8

Log in/Register

Home
SEARCH
Recently submitted

Humanities
Social sciences
Behavioral sciences
Socio-cultural sciences
Life sciences and medicine
Computational sciences

Terms and conditions
Help/Support

DANS website

www.dans.knaw.nl

DANS
www.dans.knaw.nl

KNAW
know.nl

ASDP
asdp.nl

Dataset Details - De steentijd van Nederland

Compact description | Full description | Data files

Title De steentijd van Nederland

Creator Dreefien, Jos (eds.)
Dreefien, Erik (eds.)
Dorrenbos, Marie-Francoise van (eds.)
Verbeek, Leo (eds.)

Contributor Reijnen, J.H.M. (Dir.) (hoofdredacteur Archeologie)

Date created 2009-12-03

Copyright holder Stichting Archeologie

Publisher Stichting Archeologie

Description (abstract) In dit boek wordt een actueel overzicht gegeven van de stand van onderzoek naar de Nederlandse steentijd. Het met doordruid werd in ons land aan dit onderwerp gewijd artikel is gepubliceerd. Dit is nu het eerste, is te danken aan het grote succes van de jaarlijkse steentijdsdag, die in 2009 zijn vijftiende editie doorgaf.
De steentijd van Nederland provides an overview of the Palaeolithic through research in the Netherlands.

Temporal steentijd
Stone Age

Spatial Nederland
The Netherlands

Identifier dans.dans.knaw.nl/228455380529106347-1229215020410
x05099 (x059A-project)
0824-9117 (ISSN)
10-8371-45-2-5 (ISSN)

Persistent identifier urn:nbn:nl:0:3:3tgb-muf

Relation Voor deze dataset zijn enkele reviews beschikbaar

Afbeelding 1 In EASY wordt verwezen naar de website met data reviews

ongeveer 600 personen. 160 van hen hebben we uit de selectie verwijderd, omdat in EASY bekend is dat ze niet benaderd willen worden; afnemers van wie we op grond van hun buitenlandse e-mailadres hebben aangenomen dat ze mogelijk geen Nederlandstalige survey kunnen of willen invullen, hebben we eveneens buiten de selectie gehouden.

Personen die in de genoemde periode meer dan één dataset uit de top-50 hebben gedownload, hebben meer beoordelingsverzoeken gekregen (maximaal drie). In totaal zijn in eerste instantie 438 e-mails verzonden met het verzoek om een dataset te beoordelen. 65 personen hebben dit verzoek ingestemd, wat we te beperkt vonden. Daarop is een tweede groep personen geselecteerd, op basis van een top-75² van de afgelopen 12 maanden. Na deze tweede mailing zijn in totaal 1138 e-mails verzonden. Uiteindelijk, na een herinnering per e-mail, hebben 144 personen datasets beoordeeld; 128 van hen hebben de survey afgemaakt.

² Dit zijn voornamelijk Open Access files.

3. Uitkomsten

3.1 Overkoepelende percentages

144 personen hebben de survey geheel of gedeeltelijk ingevuld: 64,8% is onderzoeker, 7,8% student, 7% beleidsmedewerker en 20,3% vervult een andere functie (onder de respondenten bevinden zich geen journalisten). Dit is een respons van 12,7%. Bij de geaggregeerde scores hierna staat op hoeveel antwoorden ze gebaseerd zijn³.

100 % van de gesloten (score-)vragen is beantwoord, doordat de opzet van de survey een antwoord afdwong. Voor de scorevragen hebben we een vijfpuntsschaal gebruikt, plus de mogelijkheid om 'niet van toepassing' te kiezen.

Inhoudelijk bevatte de survey verschillende soorten open vragen:

- vragen waar in principe iedereen iets zou kunnen invullen, bijvoorbeeld de vraag naar trefwoorden. De vraag naar de reden om de dataset te downloaden is duidelijk vaker beantwoord (98 keer = 68%) dan de overige open vragen (tussen 17% en 40%).
- vragen van het type 'waarom niet?', dat wil zeggen, vervolgvragen die slechts werden voorgelegd aan een deel van de respondenten. Binnen deze subgroepen was de respons erg hoog (84% tot 100%).

Hierna geven we per vraag voorbeelden van antwoorden; de bijlage bevat de totale overzichten. Wanneer we de relevantie van de antwoorden op open vragen als maatstaf voor kwaliteit nemen, mogen we stellen dat de kwaliteit van de antwoorden hoog is. Ook de bereidheid om mondeling nadere toelichting te geven, is met 19% (na correctie) hoog.

3.2 Datasets

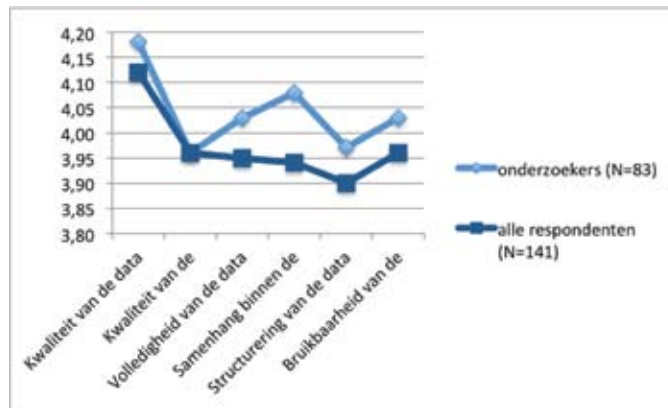
Gestelde vraag: *'Met 'dataset' bedoelen we de databestanden van één studie die bij elkaar horen, ook als u bijvoorbeeld slechts één bestand hebt gedownload. Hoe vindt u de data die u hebt gedownload op de volgende punten...?'* Afbeelding 2 toont hoe vaak de verschillende scores zijn gegeven.

³ Indien in SurveyMonkey onduidelijk is hoeveel respondenten een vraag voorgelegd hebben gekregen – tussen 128 en 144 – berekenen we percentages op basis van 144. Verder tellen we bij de open vragen antwoorden zoals 'n.v.t.' niet mee.

Onderdeel	Overall scores datasets						Gemiddelde score
	zeer goed (5)	goed (4)	niet goed en niet slecht (3)	onvoldoende (2)	slecht (1)	n.v.a. (0)	
Kwaliteit van de data	31	83	14	1	0	8	4.12 (129 beoordelingen)
Kwaliteit van de documentatie	26	80	20	4	1	6	3.96 (131 beoordelingen)
Volledigheid van de data	24	77	20	5	0	11	3.95 (126 beoordelingen)
Samenhang binnen de dataset (indien van toepassing)	20	60	22	2	0	33	3.94 (104 beoordelingen)
Structurering van de data (indien van toepassing)	19	64	22	4	0	28	3.9 (109 beoordelingen)
Bruikbaarheid van de bestandsformaten	27	74	17	6	1	12	3.96 (125 beoordelingen)

Afbeelding 2 Geaggregeerde scores voor dataset-aspecten

Wanneer we bij deze vraag de gemiddelde scores van de onderzoekers vergelijken met die van de gezamenlijke respondenten, blijken de onderzoekers vaak iets positiever:



Afbeelding 3 Gemiddelde datasetscores van onderzoekers en allen

Gestelde vraag: 'Welke trefwoorden zou u toekennen aan de dataset die u hebt gedownload, opdat andere onderzoekers deze set beter kunnen vinden?'

55 respondenten hebben gemiddeld 2,7 trefwoorden ingevuld; het merendeel is van inhoudelijke aard en daarmee in principe informatief voor andere onderzoekers. Waardeoordelen zijn nauwelijks gegeven.

Bij deze vraag werd gevraagd de trefwoorden te scheiden door komma's. Dat dit niet altijd duidelijk is gebeurd (sommige lange trefwoorden zijn voor meerdere uitleg vatbaar en 'wonen verhuizen woonwensen' lijkt toch echt een opsomming van drie trefwoorden), vertekent de resultaten enigszins. Wat opvalt is dat verreweg de meeste trefwoorden slechts

eenmaal zijn ingevoerd. Sommige daarvan lijken echter veel op elkaar, zoals 'steentijd' en 'steentijden'.

Gestelde vraag: *'Wat vindt u positief aan de gedownloadede dataset?'*

54 personen hebben deze vraag beantwoord. Een samenvatting is niet te geven, omdat de datasets varieerden. Enkele voorbeelden van antwoorden zijn:

- het longitudinale aspect;
- goed gescand, goede metadatabeschrijving erbij;
- pdf-formaat is altijd makkelijk omdat daar een goede zoekfunctie in het document zit;
- Doordat het databestand de oorspronkelijke respondentidentificatie bevat, kunnen altijd variabelen uit de afzonderlijke election-studies worden toegevoegd.

Gestelde vraag: *'Indien van toepassing, waarover bent u niet tevreden bij het gebruik van de dataset?'*

19 personen hebben deze vraag beantwoord. Een samenvatting is niet te geven, omdat de datasets varieerden. Enkele voorbeelden van antwoorden zijn:

- Omdat sommige vragen niet altijd gesteld worden; met name de *policy position* van alle partijen op alle *issues* zou altijd gevraagd moeten worden.
- Geen Engelstalige documentatie;
- Het zou prettiger zijn als het bestand niet was opgesplitst in 2 periodes, maar het is al geweldig dat dit geheel ontsloten is.

Gestelde vraag: *'Zou u deze dataset aan anderen aanbevelen?'*

92,2% van 129 personen heeft deze vraag bevestigend beantwoord, 7,8% ontkennend. Dit is een erg fraaie score.

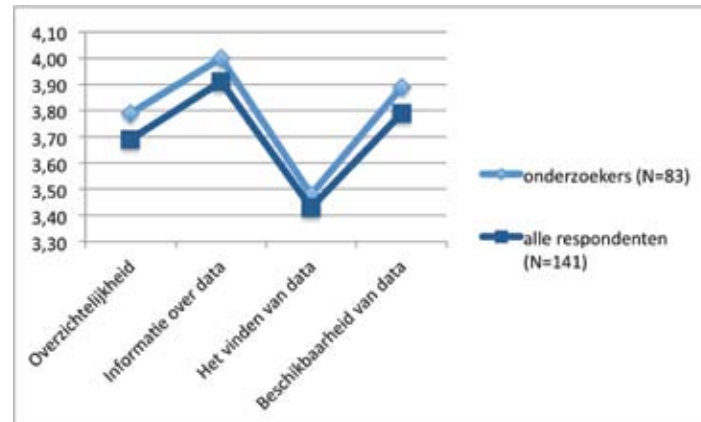
3.3 EASY-website

Gestelde vraag: *'Hoe vindt u onze website op de volgende punten...?'* Afbeelding 4 toont hoe vaak de verschillende scores zijn gegeven.

Onderdeel	Score						Gemiddelde score
	zeer goed (5)	goed (4)	niet goed en niet slecht (3)	onvoldoende (2)	slecht (1)	n.v.t. (0)	
overzichtelijkheid	13	51	37	10	0	1	3,69 (141 beoordelingen)
informatie over data	13	109	13	6	0	1	3,91 (141 beoordelingen)
het vinden van data	10	67	41	22	2	1	3,43 (142 beoordelingen)
beschikbaarheid van data	20	54	26	9	2	1	3,79 (141 beoordelingen)

Afbeelding 4 Geaggregeerde beoordelingen van de website

Opnieuw blijken de onderzoekers de gemiddelden omhoog te trekken:



Afbeelding 5 Gemiddelde websitescores van onderzoekers en allen

Bij alle vier de onderdelen is de groep 'goed'-zeggers het grootst. Het onderdeel 'Het vinden van data' krijgt verhoudingsgewijs de meeste lage scores (onvoldoende of slecht). De scores voor de website-aspecten liggen iets lager dan de scores voor de datasets (zie Afbeelding 2).

Gestelde vraag: *'Heeft u opmerkingen over de website of suggesties voor verbeteringen?'*

22 personen hebben deze vraag beantwoord. Van de 24 personen die bij de vorige vraag het onderdeel 'Het vinden van data' als onvoldoende of slecht beoordeelden, hebben er 10 bij deze vraag een suggestie gedaan of een toelichting gegeven. Met andere woorden: bijna de helft van de critici heeft deze gelegenheid benut.

Een behoorlijk deel van de antwoorden betreft de zoekfunctionaliteit en metadata in EASY; deels hiermee samenvallend heeft ook een

behoorlijk deel betrekking op archeologische datasets. Een aantal voorbeelden:

- Google zoekmachine aanbieden i.p.v. zelfgemaakte zoekmachine. Wat er nu zit werkt redelijk tot goed, maar de Google *engine* is sneller en werkt toch wat beter.
- Het kost me net wat te lang om de NKO-data te vinden. Zoekfunctie werkt matig.
- Ik heb vanuit mijn taak als regio-archeoloog vooral de behoefte aan definitieve rapporten. Deze zijn veelal moeilijk te vinden omdat veelal geen 'hits' op specifieke rapportnummers komen. Ook kun je in de algemene pagina deze rapportnummers niet zien. De invoerdatum is voor mij van minder belang, maar deze staat daar wel.

Aan het eind van de survey hebben we gevraagd welke informatie we in EASY over de respondent mogen vermelden bij de beoordeling van de betreffende dataset. Enerzijds is dat de functie, zoals onderzoeker (64,8%) of student (7%). Anderzijds zijn dat naam en organisatie van de betrokkene. Bijna 60% van 128 personen wil anoniem blijven; de anderen gaan ermee akkoord als de gegevens waarmee ze zich bij EASY hebben geregistreerd hier zichtbaar zijn.

3.4 Onderzoek

Gestelde vraag: *'Wat was uw belangrijkste reden om deze dataset te downloaden?'*

98 personen hebben deze vraag beantwoord. Vervolgens hebben we zelf de antwoorden gerubriceerd als 'voor onderzoek' (56 maal), 'uit interesse' (12 maal), 'voor studie over onderwijs' (7) en overig (23). Enkele voorbeelden:

- GIS onderzoek naar verkiezingsresultaten populistische partijen;
- Document geeft opgravingsverslag van een project in mijn woonplaats;
- scriptieonderzoek

Gestelde vraag: *'Helpt deze dataset u om uw onderzoeksvragen te beantwoorden?'*

68,8% van 128 personen heeft deze vraag bevestigend beantwoord,

10,2% ontkennend. ('niet van toepassing' werd aangeboden voor respondenten met een ander doel dan onderzoek.) De ontkennende groep kreeg vervolgens de multiplechoicevraag gepresenteerd '*Zo nee, waarom niet?*'. Antwoorden hierop waren '*niet relevant genoeg*' (53,8%), '*de inhoud van dataset wijkt af van wat ik verwachtte*' (38,5%) en '*niet actueel genoeg*' (0%); de vrije antwoorden waren:

- ontoegankelijk;
- niet gedetailleerd genoeg voor wat ik zocht;
- pure nieuwsgierigheid.

Het is duidelijk dat gebrek aan relevantie de voornaamste oorzaak is wanneer een dataset niet bijdraagt aan het beantwoorden van een onderzoeksvraag; gebrek aan actualiteit daarentegen speelt geen rol.

Gestelde vraag: '*Heeft u gepubliceerd met gebruikmaking van deze dataset? En zo ja, wat is de referentie naar de publicatie?*'

15,6% van 128 personen heeft deze vraag bevestigend beantwoord, 84,4% ontkennend. In totaal hebben we referenties ontvangen van 10 verschenen en 5 te verschijnen publicaties.

Gestelde vraag: '*Bent u van plan om te publiceren met gebruikmaking van deze dataset?*'

43% van 128 personen heeft deze vraag bevestigend beantwoord, 57% ontkennend.

Overig: neveneffecten

Het is voorstelbaar dat het uitsturen van een survey waarin datasets met name worden genoemd, de aandacht voor DANS EASY of voor bepaalde datasets vergroot. Daarom hebben we vóór en na de verzending van de surveymails gekeken naar webstatistieken, in het bijzonder het verkeer dat direct vanaf de DANS-site naar datasets in EASY gaat. In dit verkeer is een licht positieve trend te zien, maar de verschillen zijn dermate klein en grillig dat er geen duidelijk verband met dit project te benoemen is. Veruit het grootste deel van de gebruikers die EASY binnenkomen via de DANS-site gebruiken de 'Nieuw in EASY'-pagina. In de periode van de survey zijn er niet meer klachten of vragen binnengekomen dan anders.

4. Conclusies en aanbevelingen

4.1 Samenvatting en conclusies

De **respons** op ons verzoek tot data reviewing is 12,7%. Als we dit beperken tot de mensen die de survey helemaal hebben ingevuld, ligt het responspercentage bij 11,3%.

De gemiddelde scores voor de bevroegde aspecten van de **datasets** liggen rondom de 4 op een schaal van 1 tot 5, met 'kwaliteit van de data' op de eerste plaats (4,12). Dat is iets om blij mee te zijn. Bij de onderzoekers (65% van de respondenten) liggen de meeste gemiddelden nog een fractie hoger. Maar liefst 92% van de respondenten zou de dataset aan anderen aanbevelen, wat een krachtige indruk geeft van de kwaliteit van de datasets.

Datasets worden in 57% van de gevallen gedownload voor **onderzoek**. Dat is een vrij laag percentage, maar we moeten hier in het oog houden dat het een open vraag was waarvan we de antwoorden zelf hebben gerubriceerd. Mogelijk zou een multiplechoicevraag tot een andere verhouding leiden (aanbeveling: maak er een gesloten vraag van). Bovendien kan het te maken hebben met de datasets in de pilot: misschien zijn er toevallig relatief veel bij die eerder voor beleidsmedewerkers dan onderzoekers interessant zijn.

In bijna 70% van de gevallen helpt de dataset om de onderzoeksvragen te beantwoorden; gebrek aan relevantie is de voornaamste oorzaak wanneer een dataset niet bijdraagt aan het beantwoorden van een onderzoeksvraag. 16% van de respondenten heeft de dataset al gebruikt voor **publicaties**, terwijl bijna de helft (44%) van plan is om dit – alsnog of nogmaals – te doen.

De datasetscores liggen enigszins boven de scores voor de **website**. Ook inzake de website zijn de onderzoekers gemiddeld positiever dan de anderen. In deze rubriek staat het aspect 'informatie over data' (3,91) bovenaan, terwijl het 'vinden van data' relatief laag wordt gewaardeerd

met 3,44. Dit laatste cijfer komt overeen met het grote aandeel opmerkingen over de **zoekfunctionaliteit** van EASY; dit is duidelijk een terrein waar onze afnemers wensen hebben.

Positief te waarderen is zeker dat bijna 20% van de respondenten bereid is om hun antwoorden mondeling toe te lichten. Dit wijst op een hoge mate van **betrokkenheid**. Ook de voorbeelden die respondenten geven in de open vragen wijst in deze richting: men neemt de datasets, het reviewen van datasets en EASY serieus.

Het aantal respondenten dat **tags** (trefwoorden) heeft toegekend, ligt met 38% enigszins beneden onze verwachting; weliswaar op hetzelfde niveau als de vraag 'Wat vindt u positief aan de dataset', maar duidelijk lager dan de vraag 'Wat was uw belangrijkste reden om deze dataset te downloaden' (68%). De bescheiden respons kan komen doordat het een te ongebruikelijke vraag is; in dat geval zou het in de toekomst kunnen helpen als EASY tags presenteert en zoeken daarin mogelijk wordt.

4.2 Aanbevelingen en overwegingen

Onze voornaamste aanbeveling is om een continue vorm van datareviewing door afnemers van datasets in te voeren. Zowel de medewerking aan de pilot als het hiervoor genoemde onderzoek van SURFfoundation wijzen in die richting. Deze aanbevolen nieuwe functionaliteit van EASY kan deels profiteren van materiaal uit de pilot en moet deels verder worden ontworpen.

Gezien de opmerkingen op de website doen we verder de aanbeveling om de zoekfunctionaliteit binnen EASY te verbeteren; met de huidige verbeteringen in EASY wordt hierin al deels voorzien door meer browse-mogelijkheden.

We ronden dit hoofdstuk af met een aantal overwegingen voor de invoering van een continue data review, verdeeld over de aspecten proces, survey, koppeling aan EASY en trefwoorden.

Proces

Een goede mogelijkheid lijkt om, net als bij een hotelboeking, een week na het downloaden van een dataset de afnemer een reviewformulier te sturen. Bij automatisering hiervan is het essentieel dat er een link wordt gelegd tussen de dataset en de afnemer, om de scores bij de juiste dataset te kunnen presenteren.

Een automatische download van resultaten uit SurveyMonkey in bijvoorbeeld CSV-formaat is helaas niet mogelijk. Aan het automatisch toevoegen of actualiseren van scores, gemiddelden, tags en dergelijke in EASY is in de pilot geen aandacht besteed.

Het is overigens niet vanzelfsprekend dat beoordelingen automatisch in EASY verschijnen. Het is beter om een – terugkerende – redactionele taak in te voeren voor het controleren van beoordelingen. In de pilot hebben we geen ongepaste uitingen gezien, maar bijvoorbeeld wel trefwoorden die feitelijk suggesties zijn voor metadatavelden, zoals ‘plaatsnaam’.

Ook het automatisch attenderen van depositors op reviews van ‘hun’ datasets valt te overwegen.

Survey

De bestaande vragenlijst is herbruikbaar. Wel kan men opnieuw bezien welke vragen relevant zijn voor andere afnemers en welke voor DANS zelf: is het bijvoorbeeld nodig om permanent de EASY-website te laten beoordelen? Om frequente downloaders niet lastig te vallen, zou dit onderdeel optioneel kunnen zijn. De vraag of er publicaties zijn verschenen op basis van de dataset vervalst wanneer de review kort na het downloaden wordt verstuurd.

Er blijken heel wat buitenlandse adressen onder afnemers te zijn, dus het loont de moeite om de reviewvragen tweetalig te maken. Een vraag is dan of ook de presentatie van beoordelingen en dergelijke tweetalig zou moeten zijn.

Koppeling aan EASY en presentatie van scores

In de pilot is de beoordeling van een dataset toegankelijk gemaakt via het metadatumveld Relation (zie Afbeelding 1). De presentatie van de beoordelingen is zelf geen onderdeel van de data review geweest, dus we weten niet hoe toegankelijk en informatief gebruikers dit vinden.



Afbeelding 6 Een voorbeeld van de resultaten van de data reviews

Het verschil tussen open en restricted acces hebben we in de pilot genegeerd; we stellen niettemin voor dat reviews van datasets met restricted access algemeen zichtbaar zijn, zelfs voor bezoekers die niet zijn ingelogd. Dan zijn ze namelijk maximaal informatief voor geïnteresseerden.

In de pilot hebben we ervoor gekozen om reviews pas te tonen als de betreffende dataset minstens tweemaal is beoordeeld. Het is onduidelijk wat het meest wervend is: überhaupt een score tonen, of pas als die een 'solide' basis heeft, of zelfs pas als een 'solide' aantal datasets een 'solide' scoringsbasis heeft. Het aantal scores waarop het gemiddelde is gebaseerd is in elk geval belangrijke informatie voor een potentiële afnemer.

Zeker wanneer EASY wijzigingen in beoordelingen automatisch weerspiegelt, is het voor interne monitoring wenselijk om snel een overzicht te krijgen welke datasets beoordeeld zijn.

Trefwoorden

Diverse websites presenteren *tag clouds* als navigatie-instrument. Dat is ook voor EASY een overweging, mits er een voldoende representatief geachte tagverzameling is gegroeid (of we duidelijk stellen dat het een experiment betreft). Voor representativiteit zal het nodig zijn de afzonderlijk toegekende trefwoorden te aggregeren, bijvoorbeeld door spellingvariatie te beperken. Het consolidatieproces van een tagverzameling kun je bovendien bevorderen door automatisch tagsuggesties aan te bieden, zodat gebruikers tags niet zelf hoeven bedenken en intypen.

Bijlage: antwoorden op de open vragen

Gestelde vraag: 'Wat vindt u positief aan de gedownloadede dataset?'

Correctie van de antwoorden: drie maal 'n.v.t.' verwijderd

- Dat er een overzichtswerk wordt opgenomen in de database.
- Hoge kwaliteit data
- Ik kende het boek, nu is hij als pdf beschikbaar.
- bijzonder compleet
- uitgebreide gegevens over wonen
- Het is gewoon een erg interessante, diepgravende en uitputtende dataset voor het onderzoeksgebied en de periodes.
- Goed gescand, goede metadatabeschrijving
- zeer uitgebreid, natuursteen door E. Kars
- Hoeveelheid historische data, bruikbaar formaat
- veel documentatie erbij
- Volledigheid, nagenoeg geen ontbrekende data.
- dat het inzicht geeft in wat ik zoek
- uitgebreidheid van de dataset en het feit dat het om een voor Nederland representatieve steekproef gaat.
- Het longitudinale aspect
- complexiteit, omvattend
- omvang van zowel het aantal respondenten als van de informatie per respondent
- De rapportages zijn in pdf alle eenvoudig te downloaden. Van de ROB- onderzoeken zijn daarnaast de veldwaarnemingen, dagrapportages, ideeën, determinaties, tabellen etc etc terug te vinden. M.a.w. er is veel beschikbaar.
- Het feit dat deze vrijelijk ter beschikking gesteld wordt
- Alle gegevens beschikbaar op internet
- zeer grote representatieve landelijke steekproef. Onmisbaar voor iedereen die zich op het onderzoeksterrein van woning begeeft.
- de volledigheid ervan, omvat gegevens die je zelf niet kan verzamelen
- de compleetheid
- Fijn dat deze per artikel zijn te downloaden
- het bevat het gehele proefschrift met de bijbehorende data
- Het is eigenlijk een boekwerk dat digitaal beschikbaar is. Hierdoor wordt het makkelijker beschikbaar.
- over de jaren heen compleet
- Volledige documentatie van het betreffende proefschrift
- forse omvang van de enquêterespons, veel gegevens over woonomgeving, naast woonsituatie zelf.
- toegankelijkheid
- Na het gevonden te hebben, ben ik blij dat ik het kon downloaden

- grote omvang
- grote steekproef
- longitudinaal karakter
- Doordat het databestand de oorspronkelijke respondentidentificatie bevat, kunnen altijd variabelen uit de afzonderlijke election-studies worden toegevoegd.
- hij is er, er zijn geen alternatieven
- Opsplitsing van de verschillende onderdelen
- lange tijdreeks
- op een redelijk eenvoudige wijze kan ik kennis nemen van interessant onderzoek in het hele land. De originele rapporten geven een diepgang die ik ook zoek.
- hoeveelheid en kwaliteit variabelen. opname van partnerkenmerken.
- Overzichtelijk, en duidelijk geschreven. PDF formaat is altijd makkelijk omdat daar een goede zoekfunctie in het document zit.
- Het is de enige zo uitgebreide database over woonvoorkeuren
- veel onderwerpen, loopbanen, beide partners
- Uitgebreid, veel bedrijven meerdere jaren achtereenvolgend opgenomen in steekproef
- bruikbaar voor onze vereniging
- De snelle beschikbaarheid is zeer waardevol.
- niet direct bruikbaar
- hetgeen aanwezig is, is makkelijk te downloaden
- weergave
- Het betreft een landelijk opgezette steekproef, hetgeen in 1966 zeer uitzonderlijk was. Voor bestudering van het politieke systeem is 1966 een interessant jaar (opkomst studentenprotesten, jongerenbewegingen)
- 1. Alleen de indeling van het hele methoden en technieken veld al is handig. We hebben het onlangs gebruikt om voor onze onderzoekers bruikbare literatuur te ordenen.
- 2. Uiteraard is het jammer dat het systeem na 2004 niet meer wordt bijgehouden, maar door de diensten van DANS kan men nog handig zoeken in de literatuur van 1958-2004. Ik gebruikte SRM vaak om onderzoekers te wijzen op voorbeelden van toepassingen van minder vaak gebruikte modellen en technieken op hun eigen terrein.
- volledig en documentatie
- compleet; volledig definitief rapport
- Makkelijk op te vragen
- uitgebreid, diverse thema's

Gestelde vraag: 'Wat was uw belangrijkste reden om deze dataset te downloaden?'

De coderingen hebben we zelf toegekend en zijn subjectief.

12 maal i uit interesse
56 maal o voor onderzoek
7 maal s voor studie of onderwijs
22 maal R restcategorie van moeilijk in te delen antwoorden

i Interesse

i Interesse in alles wat met prehistorische archeologie te maken heeft.

i nieuwsgierigheid

i wederom pure interesse

i Interesse in reisafstand, vervoermiddelgebruik HBO studenten.

i Document geeft opgravingsverslag van een project in mijn woonplaats

i Interesse naar het tijdvak

i De gegevens in zien

i interesse

i Mijn interesse voor cultuurhistorie en archeologie

i interesse inhoud

i Mijn algemene interesse in historie van mijn woonomgeving

zoek naar gegevens over archeologie

kijken of het nuttig is deze te koppelen aan eigen data sets

nodig voor onderzoek

onderzoek

literatuuronderzoek binnen een archeologische rapportage

onderzoeksdoeleinden

vergelijking vindplaatsen

Informatie inwinning

GIS onderzoek naar verkiezingsresultaten populistische partijen

achterhalen sociale context van sportbeoefening in NL

Ik wilde gegevens van mijn eigen respondenten kunnen vergelijken met een nationale steekproeg

onderzoek naar hypotheek van ouderen

nodig voor onderzoek

werkgerelateerde/onderzoekgerelateerde vraagstelling

Ivm eigen onderzoek op het aangrenzende Marine Vlieggkamp Valkenburg.

M.n. interesse in de veldwaarnemingen en interpretaties mbt de logistieke inrichting van het wijdere landschap, ligging geulen, wegen, militaire/civiele installaties

Doen van onderzoek

Promotieonderzoek, scriptieonderwerpen voor studenten

VOC gegevens over opvarenden, met name van kamer Delft.

- Om onderzoek te doen naar verhuisgedrag en buurtkeuze
- gegevens over woningmarkt Helmond/SRE
- om analyses voor Helmond en de regio te kunnen doen
- Specifieke interesse voor onderzoek in Rhenen
- archeologisch onderzoek in de Achterhoek
- gebruik in verder onderzoek
- Informatie winnen over kiezersgedrag
- Te gebruik als achtergrondinformatie voor het uitvoeren van archeologisch prospectief onderzoek in Oost-Nederland
- gebruik van analyseresultaten in een artikel over wijk economie
- onderzoek en onderwijs
- eigen specialisme middeleeuwen zuidnederland
- analyse woningmarkt
- nodig voor wetenschappelijk onderzoek
- verkennen van overlap met eigen onderzoeksinteresse
- de gegevens nodig voor mijn onderzoek
- De NKO-reeks overdekt een lange periode, zodat veranderingen in de tijd (trends) goed onderzocht kunnen worden.
- arch. onderzoek
- controle/vergelijking
- inhoudelijk. – Voor gebruik in verder onderzoek
- Nodig voor data-analyse
- aansluiting op mijn onderzoeksvraag
- Voor artikel over woningbehoefte van specifieke doelgroepen
- onderzoek
- Herijking onderzoek naar cultuurparticipatie in Noord-Brabant
- Type data waar ik naar op zoek was zit er in. Contact met OSA maakte duidelijk dat een goed georganiseerd onderzoeksinstituut de dataverzameling en verwerking georganiseerd heeft
- inzien eerder uitgevoerd archeologisch onderzoek
- Een van mijn onderzoekers had bij ICPSR 17360 gedownload en kwam met vragen bij mij.
- inventarisatie wat er aan data over mediagebruik beschikbaar is voor mogelijke onderzoeksartikelen. heb de data (nog) niet gebruikt
- Behoeft om specialistische ontwikkelingen op het hele terrein van methoden, technieken en statistiek te kunnen volgen.
- vraag van een collega
- onderzoek in de buurt
- onderzoek
- bekijken van vragen en vraagformuleringen ivm eigen onderzoek om te kijken of dezelfde vragen gesteld konden worden zodat vergelijking mogelijk was.
- bekijken van vraagformulering ivm vergelijkbaarheid met vragen over gezondheid in eigen onderzoek
- Onderzoek
- onderzoek naar achtergronden omgeving

- O nodig om data te vergaren voor vergelijkend onderzoek.
- O nodig voor vergelijkend onderzoek

R werk

R informatie opzoeken

R Enige in zijn soort

R wilde betreffende boek voor één hoofdstuk raadplegen

R algemene mobiliteitsgegevens Nederland

R Behoeftte aan informatie

R De dataset digitaal hebben

R Om de data ook digitaal te hebben

R Ik heb deze dataset volgens mij niet via deze weg gedownload (wellicht alleen documentatie). Ik gebruik deze data wel maar dan via directe online access by het CBS. Die bestanden zijn volgens mij uitgebreider dan die via DANS.

R Heb deze data volgens mij niet gedownload (wellicht alleen de documentatie)

R Ik heb aan deze publicatie meegewerkt

R hij bevat gegevens die ik nodig heb

R sociale mobiliteitsdata

R meer info verkrijgen

R weet niet meer

R Het opzoeken van bepaalde informatie over het grafveld.

R De documentatie, niet de data.

R Opzoeken van specifieke informatie

R opgraving binnen het gebied

R opgraving in de omgeving

R opgraving binnen de regio

R geen zin om naar de kast te lopen? Ik weet het niet meer precies.

S Studie

S scriptieonderzoek

S Het stond op de leeslijst van een module die ik volg bij de Rijksuniversiteit Groningen.

S onderwijsdoeleinden

S gebruik in het onderwijs

S gebruik in het onderwijs

S gebruik voor cursus steentijd door auteur van het boek.

Gestelde vraag: 'Heeft u opmerkingen over de website of suggesties voor verbeteringen?'

Correctie op de antwoorden: drie maal 'n.v.t.' verwijderd

- De archeologische datasets zijn lastig te vinden als je niet weet waar je moet zoeken. Datasets per archeologische organisatie is bv zeer handig, maar alleen te bereiken via de website van EDNA (e depot Nederlandse Archeologie)
- Het vinden van data kan overzichtelijker door archeologie als kopje op te nemen in de linkerkolom.
- ik zou graag specifiek willen kunnen zoeken op archeologische datasets, ipv alleen op humanities.
- Soms is het onhandig dat de datafiles en de inhoudsopgave op verschillende pagina's staan. Dan wil je een artikel en moet je de positie in het boek (mbv een bullet) onthouden voordat je 'm kan downloaden...
- Google zoekmachine aanbieden i.p.v. zelfgemaakte zoekmachine. Wat er nu zit werkt redelijk tot goed, maar de Google *engine* is sneller en werkt toch wat beter.
- Zoeken op –toegevoegde– trefwoorden zou handig zijn
- zoeken op trefwoorden levert vaak geen hit op bij data die er wel zijn!
- Er is nog te veel data alleen beschikbaar voor Archeologische groepen..... maar er wordt niet uitgelegd waarom dat is. Het betreft praktisch altijd onderzoek dat door gemeenschapsgeld bekostigd wordt. Dus ik snap er helemaal niets van.
- Het kost me net wat te lang om de NKO-data te vinden. Zoekfunctie werkt matig.
- Graag archeologie als aparte discipline vermelden; dat maakt het zoeken nog gemakkelijker
- Ik heb vanuit mijn taak als regio-archeoloog vooral de behoefte aan defintieve rapporten. Deze zijn veelal moeilijk te vinden omdat veelal geen 'hits' op specifieke rapportnummers komen. Ook kun je in de algemene pagina deze rapportnummers niet zien. De invoerdatum is voor mij van minder belang, maar deze staat daar wel.
- zoektermen leiden vaak niet tot vondst (bijv. plaatsnaam) terwijl zoeken in een reeks wel antwoord oplevert
- Ik heb gemerkt dat een zoekopdracht vaak geen resultaten geeft wanneer er leestekens in de zoekopdracht staan. Hierdoor zullen mensen kunnen denken dat een bepaalde publicatie niet in het archief is terwijl dit wel het geval kan zijn.
- Ik vind het vinden van rapporten lastig
- De discipline archeologie aan het zoekscherm toevoegen
- Ik heb me aangemeld omdat ik een geïnteresseerde ben voor informatie over cultuurhistorie en archeologie.
Een deel van die informatie is voor mij als niet lid van een erkende organisatie niet toegankelijk. Dat is jammer.

Het menu om info te vinden is redelijk hanteerbaar. Het vooraf selecteren van gebieden werkt niet altijd, merk ik.

- Het vinden van specifieke data is nog niet altijd makkelijk.
- Veel data is helaas nog niet beschikbaar betreffende archeologisch onderzoek; ligt ook aan het feit dat lang niet iedereen zijn of haar data op tijd toe stuurt
- Het zoeken zou geavanceerder kunnen door ook zoeken in de datasets zelf toe te staan of te faciliteren.
- misschien oudere data-sets en rapporten
- Meer data zou welkom zijn, b.v. meer rapportages gezien het feit dat het daar nog wel eens fout gaat.
- om de archeologische startpagina terug te vinden en te kunnen zoeken op organisatie is erg omslachtig.
Steden die op hun eigen website wel rapporten als pdf hebben maar niet op DANS. Vreemd.

Gestelde vraag: 'Welke trefwoorden zou u toekennen aan de dataset die u hebt gedownload, opdat andere onderzoekers deze set beter kunnen vinden?'

Een overzicht van twee datasets die verhoudingsgewijs vaak beoordeeld zijn. Niet-inhoudelijke trefwoorden zijn verwijderd.

Dataset: De steentijd van Nederland

Trefwoord	Aantal
'steentijd' 'verhart'	1
archeologie	1
culturen	1
geologie	1
mesolithicum	1
nederland	1
neolithicum	1
overzicht	1
paleolithicum	1
prehistorie	2
regionaal	1
steentijd	1
steentijden	1

Dataset: WoON2009: release 1.2 – Woononderzoek Nederland
(voor overheid en universiteiten)

Trefwoord	Aantal
energieverbruik	1
hypotheken	1
ouderen	1
wonen	2
wonen verhuizen woonwensen	1
woningbehoefte onderzoek	1
woningen	1
woon	2
woonlasten	1
woonwensen	1

Gestelde vraag: 'Heeft u nog opmerkingen naar aanleiding van dit onderzoek?'

Correctie op de antwoorden: drie maal 'nee' verwijderd en doublures verwijderd van twee respondenten die meer dan een survey hebben ingevuld.

- je mag me altijd bellen ;)
- Prima initiatief dat DANS. Ga zo door!
- Ik blijf ontevreden over de restricted publicaties van archeologische rapporten. In mijn ogen dienen die openbaar te zijn!!!
- Ben wel blij dat er een instantie is die dit beschikbaar stelt.
- Ik gebruik DANS alleen voor NKO, dus ben nogal een beperkte gebruiker.
- niet alle vragen zijn goed te beantwoorden als je nog bezig bent met de analyses
- Heb deze data (in een uitgebreidere versie) direct via CBS online access gebruikt en niet de DANS EASY versie. Gepubliceerd obv CBS-versie.
- kwaliteit ligt volgens mij vooral aan de snelle beschikbaarheid van datasets; m.a.w. open access ten aanzien van de archeologie; uw systeem is daar een belangrijke stap in voorwaarts

Eind 2010 heeft DANS een pilotonderzoek uitgevoerd om het reviewen van onderzoeksdata te stimuleren. 140 mensen hebben de kwaliteit beoordeeld van de dataset(s) die zij uit het EASY-archief hadden gedownload. Maar liefst 92% van de respondenten zou de dataset aan anderen aanbevelen. De gemiddelde scores voor verschillende aspecten – zoals kwaliteit van data en documentatie en structurering binnen de dataset – liggen rond de 4 op een schaal van 1 tot 5; ‘kwaliteit van de data’ staat bovenaan met 4,12. De dataset helpt in 70% van de gevallen om onderzoeksvragen te beantwoorden. Overigens downloadt niet iedereen vanwege een onderzoeksvraag; ook ‘pure interesse’ wordt als reden genoemd. Dit zijn waardevolle uitkomsten, ook voor de onderzoekers die deze datasets oorspronkelijk in EASY hebben gedeponereerd. Deze publicatie doet verslag van de uitkomsten uit de pilot.

Data Archiving and Networked Services (DANS) biedt duurzame toegang tot digitale onderzoeksgegevens. Hiertoe bevordert DANS dat wetenschappelijke onderzoekers gegevens duurzaam archiveren en hergebruiken via het online archiveringssysteem EASY. Tevens verschaft DANS met Narcis.nl toegang tot duizenden wetenschappelijke datasets, e-publicaties en andere onderzoeksinformatie in Nederland. Daarnaast biedt DANS training en advies, en doet het instituut onderzoek naar duurzame toegang tot digitale informatie.

Gedreven door data zorgt DANS er met zijn dienstverlening en deelname in (inter)nationale projecten en netwerken voor dat de toegang tot digitale onderzoeksgegevens verder verbetert. Kijk voor meer informatie en contactgegevens op www.dans.knaw.nl.

DANS is een instituut van KNAW en NWO.

