



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## Strategische dataproductie: representativiteit van data via crowdsourcing

van der Sijs, N.; Kirstein, Anna; Broeder, D.

2015

### **document version**

Peer reviewed version

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

van der Sijs, N., Kirstein, A., & Broeder, D. (2015). *Strategische dataproductie: representativiteit van data via crowdsourcing*. Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

# Bijlage 2:

## Enquête projectleiders

**1. Wat is uw naam en e-mail adres?**

7 antwoorden ontvangen.

**2. Wat voor crowdsourcingproject heeft u geleid? (titel, plaats, eventueel andere bijzonderheden)**

**3. Hoeveel vrijwilligers hebben er maximaal aan meegedaan?**

4 tot 497

**4. Hoeveel procent is er ongeveer afgevallen?**

De antwoorden variëren, maar duidelijk is dat er altijd een redelijk percentage afvallers is, behalve bij heel kleine projecten waarin de vrijwilligers onderling heel veel contact met elkaar hebben. Uit de grootste projecten blijkt bovendien dat veelal lichte procent van het werk gedaan wordt door 10% van de vrijwilligers.

**5. Wat waren de taken van de vrijwilligers?**

Transcriptie, metadata toevoegen, vertalen.

**6. Hoeveel tijd nam de projectleiding/coördinator van het project in beslag?**

Gemiddeld per project een dag per week.

**7. Heeft u zelf een technische voorziening gemaakt/laten maken? Zo ja: hoeveel tijd heeft dat gekost? Zo nee: heeft u gebruik maakt van een open source-applicatie, welke? Hoeveel tijd kostte het om die aan uw wensen aan te passen.**

Nee, hoogstens gebruik gemaakt van mogelijkheden van de instelling waarbij men werkzaam was. Nadeel hiervan was dat men daardoor tegen allerlei technische beperkingen opliep. Het grootste project heeft gebruik gemaakt van het platform VeleHanden van Picturae, en daar ook voor betaald.

**8. Hoe zorgde u voor kwaliteitsbewaking? Had u een aparte groep vrijwilligers die integraal of steekproefsgewijs het werk van anderen corrigeerde? Of liet u de correctie over aan een betaalde kracht?**

Meestal uitgebreide heldere instructies en een aparte groep vrijwilligers die controleerde; soms controleert de coördinator zelf een enkele keer is er een onderzoeksassistent.

**9. Organiseerde u een bepaalde beloning (een uitje, bijeenkomsten) voor de vrijwilligers?**

Iedereen heeft wel iets georganiseerd met verschillende beloningssystemen: een jaarlijkse bijeenkomst of een puntensysteem waarmee de vrijwilligers konden sparen voor een boek of museumbezoek.

**10. Maakte u gebruik van gaming?**

Nee.

**11. Hoe heeft u de vrijwilligers geworven?**

Via verschillende kanalen: de media of persoonlijke contacten.

**12. Had u een budget voor de organisatie van het crowdsourcingproject en hoe groot was dat?**

Gemiddeld een budget van een paar duizend euro; het grootste project had ongeveer € 50.000 aan manuren, dat niet alleen voor crowdsourcing werd toegepast maar ook voor scannen, databeheer en dergelijke.

**13. Had u een budget voor de techniek en hoe groot was dat?**

Geen van de projecten had een apart budget voor techniek behalve het grootste project dat 16.000 euro betaalde voor gebruik van VeleHanden voor een periode van negen maanden.

**14. Wat is uw eindoordeel: bent u tevreden over het crowdsourcingproject, heeft het opgeleverd wat u wilde, of heeft het meer tijd en energie gekost dan het opleverde?**

Alle projectleiders zijn erg tevreden, maar drie merken op dat het ook wel veel tijd kost; “erg leuk werk om te doen, en het levert heel veel resultaat op!”; “Ja, het was bijzonder interessant, maar het neemt wel veel tijd”.

De projectleider van het grootste project schreef: “Ondanks het vele werk vind ik het nog steeds een zinvol en dankbaar project”; “Het project was opgezet als experiment in het kader van een groter digitaliseringsprogramma. We wilden we onderzoeken in hoeverre crowdsourcing financieel en kwalitatief een meerwaarde heeft in het digitaliseren van collectiedata. Die meerwaarde is zeker gebleken. Op financieel vlak bleek digitalisering middels crowdsourcing ongeveer even kostbaar als interne digitalisering van gelijksoortig materiaal. Aangezien het experiment echter een leerproces was zou een vervolgproject vrijwel zeker efficiënter kunnen worden uitgevoerd. De waardering voor het project en voor de te behandelen collectie was groot, evenals de outreach die het project heeft gegenereerd. De kwaliteit van opgeleverde data lag hoog.”

**15. Heeft u suggesties waarop mensen moeten letten als ze crowdsourcing gaan organiseren?**

- Goede instructie leveren en feedback geven.
- Persoonlijk contact, bv. via mail, wordt op prijs gesteld.
- Goed op fouten wijzen, maar tegelijk ook letten op excellentie: er zijn heel goede mensen bij die ook kunnen meedenken, en die kunnen dan ook ingewikkelder taken als correctie en samensmeden krijgen.
- Onderschat en overschat de vrijwilligers niet. Dus nauwkeurig naar ze luisteren, dan kun je het beste uit ze halen. Ze blijven dan langer betrokken bij het project, omdat ze datgene kunnen doen waar ze zich het prettigst bij voelen/het beste in zijn.
- Van te voren de deelnemers die zich aanmelden zo goed mogelijk vertellen wat de bedoeling is.

- Maak van te voren een goed projectplan, stem verwachtingen, randvoorwaarden en op te leveren producten van te voren goed af met alle betrokkenen, werk met projectfasen en waar nodig deelprojecten.

### **Uit de eindrapportage van het grootste project:**

- “- Denk niet al te veel na in naam van de crowd. Het is verleidelijk om voor aanvang van en tijdens een project te bedenken wat voor collectie en werkzaamheden de crowd wel en niet zal interesseren. Ons is gebleken dat dat meestal niet klopt. De crowd is divers en heeft uiteenlopende vaardigheden, kennis en interesses. Luister dus vooral goed naar je crowd, en stuur tijdens het project bij in plaats van vooraf op alle mogelijke issues te willen anticiperen.
- Vertrouw op je crowd. Het is niet nodig om voortdurend bovenop het project te zitten. In de eerste weken is goede support belangrijk, maar gaandeweg bleek steeds vaker dat sommige deelnemers net zoveel, zo niet meer inhoudelijke kennis hebben. Stimuleer dus vooral de crowd om elkaar te helpen en laat zien dat dit wordt gewaardeerd en gerespecteerd. Stuur alleen bij als dit echt nodig is. Aan de andere kant: kies de controleurs, die uiteindelijk bepalen hoe de data wordt aangeleverd, zorgvuldig. Vraag bij voorkeur mensen die op het forum actief zijn (die staan open voor vraag-en-antwoord) en laten zien dat ze echt begrepen hebben waar het om gaat.
- Geef de crowd gelegenheid om écht bij te dragen. Een vaak terugkerende discussie was de regel dat alle gegevens letterlijk moeten worden overgenomen, ook wanneer duidelijk is dat die gegevens niet correct zijn. Dit frustriert deelnemers. Zoals al vermeld weten sommige deelnemers meer van de materie en geeft het opschonen van de gegevens hen het idee dat ze een échte bijdrage leveren aan het wetenschappelijke werk. Platforms als VeleHanden hebben voldoende controlemogelijkheden om de kans op interpretatiefouten te beperken (pick lists, meervoudige invoer, 1-op-1 controle) en wanneer de gegenereerde gegevens bij inlezing worden voorzien van een noot waarin vermeld wordt dat de gegevens middels crowdsourcing zijn gegenereerd, dekt dit ons inziens het risico op onbruikbaarheid van de data voldoende af.
- Let bij het selecteren van het te transcriberen materiaal op de volgende zaken: 1) Wat zijn de terugkerende patronen van vast te leggen informatie? 2) Wat zijn er voor afwijkende opmerkingen, aantekeningen: waar staan die voor, moeten ze worden vastgelegd etc. Over alle afwijkingen worden namelijk vragen gesteld. Door alvast de antwoorden paraat te hebben wordt lang wachten voorkomen. Leg dit al voor de livegang voor aan de collectiebeheerders. 3) Denk goed na over de velden en de range van informatie die erin mag komen.”