

Velehanden.nl: what does it take to make a crowd?

von Nelleke van Zeeland und Ellen Fleurbaay

„VeleHanden“, was mit „Viele Hände“ übersetzt werden kann, ist die Online-Crowdsourcing-Plattform, die das Stadtarchiv Amsterdam aufgebaut hat. Sie ist bereits seit zwei Jahren in Betrieb, sodass es an der Zeit ist, zurückzuschauen, zu bewerten und die Ergebnisse zu würdigen. Eine kurze Geschichte des Stadtarchivs Amsterdam und seiner Online-Aktivitäten bietet die notwendigen Hintergrundinformationen zu diesem Projekt. Anschließend geht es darum, wie das VeleHanden-Projekt in einer Public Private Partnership initiiert und organisiert wurde. Das führt zur wichtigsten Frage bei Crowdsourcing: Wie schafft man eine ‚Crowd‘, und wie hält man sie zusammen? Am Ende stehen die Ergebnisse von VeleHanden: Was wurde erreicht? Wie ist die Qualität der Arbeit?

A short history of the Amsterdam City Archives

First of all some basic information on the Amsterdam City Archives. Since 2007 we are located in a beautiful building in the heart of the Amsterdam City Centre. We receive an average number of about 100.000 visitors a year. That is a figure we are very happy with, as before 2007, when we were housed in a picturesque little building on the outskirts of the city, we never attracted much more than a quarter of this number. And even this quarter was declining, in an alarmingly rapid pace. At the same time we were amazed to see how many visitors were attracted to our first website in 1998 and how fast the number of web visitors was

growing. Without much effort, in a few years the website attracted ten times more visitors than our reading rooms.

The website started in 1998 and, as most websites in those days, it was a sort of online brochure, offering illustrated stories about highlights in our collections, information on opening hours and other practical things. Then we saw an Image database, probably Getty's. We fell in love with it and wanted one for ourselves. In 2003 we had our own Beeldbank.

Then some visitors asked: 'Why an Image database? Your core business is archives, so where are the archives? We had to admit that this was an apt remark and we set

Mijn profiel: Nelleke (Stadsarchief Amsterdam | MAI) | Uitloggen

VeleHanden Home Projecten Gebruikers Nieuws Help

Iedereen kan archieven online toegankelijk maken.

Leuk en verslavend!

- Kies een project
- Voer gegevens in of beschrijf foto's
- Op je eigen pc of laptop
- Waar en wanneer het jou uitkomt
- Draag bij op een leuke en eenvoudige wijze!

Aan de slag

Meedoen is eenvoudig, probeer het nu uit.

Oefenen

Een greep uit de projecten

Officiersboekjes Legermuseum 1811-1940

Bevolkingsregisters Amsterdam 1874-1893

Glashelder!

Nieuws

Update aanpassingen website versie 7
10 januari 2014

VeleHanden een terugblik op 2013
9 januari 2014

Homepage

Bevolkingsregisters Amsterdam 1874-1893

Stadsarchief Amsterdam biedt een deel van zijn bevolkingsregisters aan om te indexeren op naam, geboortedatum en geboorteplaats.

[Invoeren](#) [Controleren](#) [Beheren](#)

Stand van zaken

47%

236.640 scans 82.319 dubbel ingevoerd

1.049 deelnemers 82.311 gecontroleerd

Berichten **Project info** Statistieken

Stadsarchief Amsterdam biedt de derde serie van zijn bevolkingsregisters aan om te indexeren op naam, geboortedatum en geboorteplaats.

De Bevolkingsregisters zijn geordend op adres. Per adres is geregistreerd wie er gewoond heeft en waarheen deze persoon verhuisd is. De index op adresgegevens is al gemaakt. Via VeleHanden willen we de persoonsgegevens toevoegen. Wekelijks zullen de ingevoerde gegevens worden gekoppeld en geëxporteerd naar de Archiefbank van Stadsarchief Amsterdam, zodat u snel resultaat ziet van uw werk. Wekelijks zullen de ingevoerde gegevens worden gekoppeld en geëxporteerd naar de Archiefbank van Stadsarchief Amsterdam, zodat u snel resultaat ziet van uw werk.

Projectleiders

Aan de slag

[Download de invoerinstructions](#)

[Download de controle-instructies](#)

Statistieken van vandaag

33 ingevoerd

25 gecontroleerd

[Alle statistieken](#)

Invoerders van vandaag

1	Leonore Rijnberg	10 Scans
2	R.S. Jansen	6 Scans
3	pieter Kloppenburg	3 Scans

[Bekijk alle](#)

Wat zijn Bevolkingsregisters?

Projectpage

ourselves to get an overview and catalogue of our archival collections online. We named it: Archiefbank, or Archives Database. Our colleagues said: well done! But our customers were not satisfied and they asked: 'Where is the button to see those documents? I cannot find it'.

Again an apt remark we thought. So we decided to start scanning. This time our colleagues laughed at us and asked us if, by the way, we had forgotten to calculate how long it would take to scan all those kilometres of documents. Of course we had not forgotten; it is an easy sum: with an average production of 10.000 scans a week it would take about 420 years to scan our collections. But the question is not about how much documents we have in our collections, the question is how much documents our customers ask from us. And that does not outnumber about 15.000 scans a week. That is why we started a 'scan on demand service'. This works out fine and nowadays we produce 15.000 scans a week. When a request for scans is received it takes three to five weeks to produce the scans and bring them online. A lot better than the calculated 420 years ...

However, then came the questions from customers that led to VeleHanden. This time the questions were not as simple as 'where is the button'. They were more like vague complaints: 'I cannot find it', 'It is too difficult' or 'Why is there so much'. It took us some time before we comprehended what was needed. Eventually, we came to the conclusion that our customers do want to find things, but they do not want to have to search for them. They want to know the history of their house, or the history of their family, but they do not want to make a study of how the population was administered in the past. They just want to type in a name or an address. In short, they want an index to all those documents.

We had some indexes made in India, but even in low-wage-countries, this is rather expensive. The average costs of indexing a scan vary from three to ten times the price of producing a scan. It would be absolutely impossible for us to find the money to pay for indexing on a large scale. What we needed for indexing were volunteers, and not just a few, but lots and lots of volunteers. And that is how the idea for this crowdsourcing project was born.

Can we build it?

To attract a real 'crowd' to our project it seemed better to us, not to operate on a local scale. A project where all archives in the Netherlands could benefit from, seemed more appropriate. So in 2011 we invited our colleagues to join us in a nationwide pilot project to index all Dutch Militia registers. As enrolment for the army was obligatory, the name of every Dutch 19-year-old boy is written at least once in these registers. So nowadays every man or woman with Dutch ancestors can find some of them in these registers. An interesting project for a real crowd.

However, a nationwide project costs a lot of money. So the first problem to solve was the funding problem. The solution for this problem was found in a public private partnership. You could call it co-creation, as there were various partners involved in making VeleHanden possible.

The first category of partners are the archival institutions. They were asked to pay for the scanning of their own documents. The Amsterdam City Archives offered to manage the scanning process and we were able to negotiate a very low price per page, simply because of the size of the project. So we as archival institutions gain a lot as we receive high quality and cheap scans plus an index and all we learn from being part of an innovative project.

Secondly, VeleHanden was supported by public funds that want to stimulate people to participate in cultural and socially relevant projects. Two public funds decided to join us: the Mondriaan Foundation and VSBfonds.

The third party that profits from the project are the users of the Militia registers website who like to have a look at, or to download our scans. These users pay € 0,50 per scan. Of course there is an important exception: our crowd-sourcers can earn scans for free.

And then there was a very important fourth partner in this project: the company that would own and exploit the crowdsourcing platform. We send out a Request for Proposal in which we offered a fixed price for the company that came up with the best plans. We did not ask for lowest price, but we asked for best quality and best warranties for continuity in the future.

We received seven enthusiastic proposals and we decided to contract two partners: one for the crowdsourcing platform VeleHanden.nl and one for the search system and webshop Militieregisters.nl. Geneabase built and owns militieregisters.nl and Picturae exploits the VeleHanden website. Currently, the crowdsourcing platform facilitates already ten institutions in eleven different projects. When an archive service wants to use VeleHanden, it has to pay a service fee that is related to the size, complexity and duration of the project. If new functionality is required for a particular project, it is up to Picturae to develop this functionality in dialogue with the archival institution. However, the archive service retains control over both the digital images and any metadata created by the volunteers during the project. We believe that the partnership therefore combines a commercial imperative for Picturae to support, develop and sustain

VeleHanden with the archival institutions' mission to promote online access and public engagement with archives.

In the next step of development, the actual creation of the platform for crowdsourcing, we invited a lot of people to help us 'co-create' VeleHanden. We realised that we needed to attract a huge and diverse community for crowdsourcing. This meant we had to attract a lot of people that we had never met before. Who are they, why would they participate and what is important to them? We needed to get to know these people a bit better. And we felt they had a right to participate in the development of the platform, as they were going to do the actual work on the website.

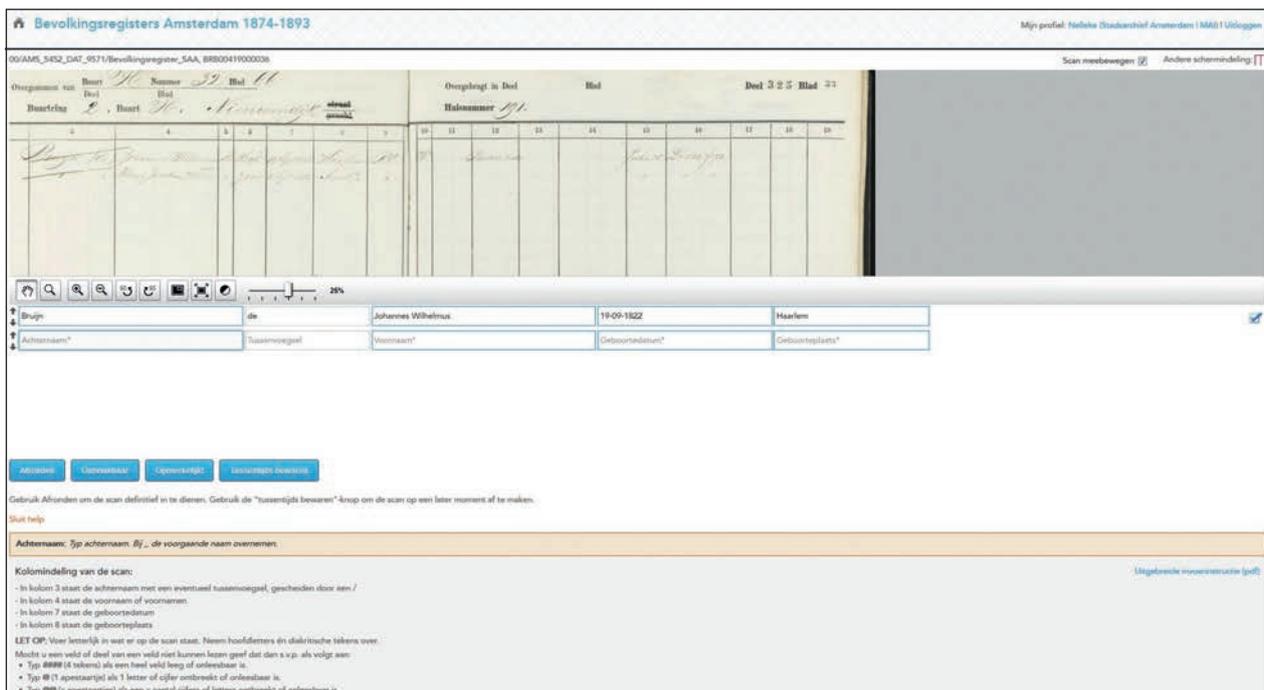
So we simply asked in one of the City Archives' newsletters for volunteers to join the project to help us testing the tools for data entry and for quality checks and to test login procedures and forum facilities. In two days' time 150 people volunteered and we had to stop enrolment for testing as our software-partner was getting a bit nervous.

We started the test period when the first tool, the tool for data entry was ready to be tested. We organised a meeting to get acquainted with our panel and to explain our ideas. They started the data entry, and at the same time we started developing the tool for control. Six weeks later, we held a second meeting, explained the working of the control tool and in the second test period our testers checked the data they had entered in the first six weeks and by doing this they tested the control tool.

All in all the testing took four months. We held another meeting and we organised several surveys, using Google analytics. Attendance to the meetings varied from twenty to fifty people. Even more important than the meetings and the surveys was the Forum. The Forum was a busy meeting-place for direct contact on a daily basis between testers and the project team. And it continues to be; direct communication is a very important feature of the site and we keep thinking of improvement.

All in all, this co-creation of our platform worked out very well and our software engineer told us she wished all projects could be done in this way. She enjoyed explaining how the site worked not only to us, but also to the actual users of the site. And she said it was very rewarding to be able to fix bugs as soon as people see them and to get a nice 'thank you' for it, instead of receiving complaints after you thought you had finished your job. On the other side, the volunteers were very happy to receive answers from the software engineer herself and they experienced that we really paid attention to what they said.

During this testing period the core of our crowd was formed. At November 3rd 2011, when we officially baptised VeleHanden and send it out in the world to grow, the testers transformed to ambassadors of the website. They do all the quality checks and they help new arrivals. At this moment, VeleHanden counts more than 2900 members and it still grows daily. So we can assume we managed to create a crowd.



Data entry – archival project



Data entry – photo tagging

VeleHanden.nl

The platform VeleHanden.nl consists of several parts that we will briefly describe here. Of course this will make even more sense if one would logon to the website and browse around.

On the home page you will find general information about VeleHanden. What is it? Why should I join in in this project? What kind of projects can I work on? Each project has its own page with more specific information about the source the project is about, the task, the reward, relevant news items et cetera. Per project you can also find statistics on the overall progress of the project and which member entered or controlled most scans. We think this is also an important aspect for motivation of the crowd. Most of the projects are still genealogical projects, initiated by archival institutions, but there are also video tagging and photo tagging projects and even one in which you help determining louses and mites. This shows that VeleHanden is still developing.

On the member page ('Gebruikers') you find an overview of all volunteers who work on one or more projects. On the Forum people can ask questions or share interesting information with the crowd. The menu items News and Help (FAQ) speak for themselves.

Each member can choose a project to work on, and by clicking on the 'invoeren'-button, you start with the data entry. Data can be entered once or twice, or even more times, if you would like to. For most genealogical projects double data entry is chosen. A volunteer types the requested data and once he finished the scan and saves the data, a new scan immediately appears. We hope this motivates the volunteers and makes VeleHanden almost addictive. While entering data, the participants can use two buttons to contact the project managers. They can mark a scan as unusable ('onbruikbaar'), for example when it is a cover of a book, or as remarkable ('opmerkelijk'), if something special is noted. In the Militia registers project for instance we

De gecontroleerde versie				
Vegte	v der	Karel Ernst	04-07-1840	Gebourteplaats*
Rareband	Timmer	Elizabeth Maria Coletta	21-11-1845	Gebourteplaats*
Vegte	v der	Geertruida Elisabeth	21-08-1866	Gebourteplaats*
Vegte	v der	Lion	12-02-1876	Gebourteplaats*
Maat	ter	Frans Engbert	07-03-1850	Amstern
Vegte	v der	Wassenaer	29-05-1877	Amsterdam
Vegte	van der	Wassenaer	01-07-1879	Amsterdam

De twee ingevoerde versies	
Invoers1	Invoers2
1. Vegte v der Karel Ernst 04-07-1840 Amsterdam	1. Vegte v der Karel Ernst 04-07-1840 Amsterdam
2. Rareband Elizabeth Maria Coletta 21-11-1845 Amsterdam	2. Rareband Elizabeth Maria Coletta 21-11-1845 Amsterdam
3. Vegte v der Geertruida Elisabeth 21-08-1866 Amsterdam	3. Vegte v der Geertruida Elisabeth 21-08-1866 Amsterdam
4. Vegte v der Lion 12-02-1876 Amsterdam	4. Vegte v der Lion 12-02-1876 Amsterdam
5. Maat ter Frans Engbert 07-03-1850 Amstern	5. Maat ter Frans Engbert 07-03-1850 Amstern
6. Vegte v der Wassenaer 29-05-1877 Amsterdam	6. Vegte v der Wassenaer 29-05-1877 Amstern
7. Vegte van der Wassenaer 01-07-1879 Amsterdam	7. Vegte van der Wassenaer 01-07-1879 Amstern
8. Vegte van der Geertruida Elisabeth 21-08-1866 Amsterdam	8. Vegte van der Geertruida Elisabeth 21-08-1866 Amstern
9. Vegte van der Henri George Zoon 29-01-1881 Amsterdam	9. Vegte van der Henri George Zoon 29-01-1881 Amstern

Control tool

asked the crowd to use the 'remarkable-button' when they recognised famous persons. This led to a nice summing up of heroes on Militieregisters.nl.

The control tool is only accessible if you have been appointed controller by a project manager. The data entered by both participants is shown next to each other and next to the scan and differences are highlighted. The controller can see in an instant where he should pay extra attention to. He can choose one of the entries by clicking on it, or if he disagrees with both, he can enter the correct data himself. If he does not know what is right either, he can use the problem button to send an email to an expert, someone from the archival institution. This option is not used by the controllers often though. As a default setting both participants are rewarded with VeleHanden points, but if the controller suspects misuse, he has the power to take points. This is also a rare exception.

Pleasure, praise, profit

We saw that we have already passed the problem of making a crowd. Now the important challenge is to maintain a crowd. Of course we have thought a lot about it: it started during the development of VeleHanden, but thinking of ways to keep the crowd still continues. We believe the work has to provide three things to the crowd: pleasure, praise and profit.

First of all: it must be a real pleasure to use the website. It must be user friendly and really fast. The task itself must be fun and it must be nice to come into contact with like-minded people that you get to know after a while.

Then you must feel and be told that the work you are doing is appreciated very much, not only by the archives, but by all historical researchers now and in the future. Working on VeleHanden is a socially relevant activity. It may sound obvious, but this is an important feature of all volunteer work: volunteers have to feel appreciated and praise is a simple, but an important reward. Important at this point is also to show the results: as Amsterdam City Archives we

weekly add the newly indexed names in a work in progress index on our website. This is great for our crowd, because if a participant enters a name today, he will be able to search and find that name next week. For us as project team it is great as well, because we see our indexes grow with about 15.000 to 20.000 names a week, which is incredible.

And, last but not least, you have to gain something as well. At VeleHanden you receive what we call VeleHanden points, to be compared to Airmiles. These points are not worth a lot; an estimate would be that you earn € 0,50 an hour. But still, it is a sign of appreciation and we learned this is important to our crowd. Each archive can decide how many points they want to reward their volunteers and can decide what kind of 'products' they offer. Usually, it is access to scans, which are otherwise only downloadable for money. But we also have sent flowers or chocolates and organised workshops, guided tours and even a lottery. Although it is not the decisive factor to participate, it is important to reward your crowd. An example that illustrates this, is the birthday of the VeleHanden project of the population registers. We decided that on that particular day, people would be awarded five instead of the regular three points for each scan. Where the average of indexed scans was 250 a day, on the birthday it suddenly rose to 591.

These three p's, pleasure, praise and profit, are the key factors which we will continue to monitor and develop to maintain our crowd, in hope we can keep them as enthusiastic and hard-working as they are now.

Results

To conclude some more details about the results of crowdsourcing. In almost two years, VeleHanden counts 2900 participants, mainly from the Netherlands, but also from abroad, for example from Brasil, the United States, Sweden, Australia and even Senegal.

Their production is overwhelming. If I take only the genealogical projects into account, there are over 4,5 million names transcribed. Calculated in working hours, this would

take us, regular archivists, more than fifty years. And then we would not even take breaks. So that is huge. On top of that, they really seem to like it, because they are also really active on the Forum. In two years they started 1280 forum topics and responded to them more than 5000 times.

What you see in all crowdsourcing-projects: most of the work is done by a small hard core and this is also the case with VeleHanden. Fifteen participants have processed more than an amazingly amount of 10.000 scans each.

But quantity is not as important as quality. The question most asked is: is crowdsourcing prone to error? Last year, Ellen Fleurbaay and PhD researcher Alexandra Eveleigh from University College London, tackled this question in a paper delivered at the ICA conference. They gave a three-fold answer to this question, which we will summarise here.

Firstly, support and guidance is essential for crowdsourcing. Instructions are provided in several different places and formats: as a manual, while indexing and in FAQ. People can ask for help on the Forum or use the remarkable button to bring a scan under extra attention of a controller or project leader. These guidelines and communication tools make the crowd confident.

Secondly, VeleHanden uses a double entry system: as we explained earlier two different people independently index the same scan. And then on top of that, a third person checks the data that the first two have entered. This third volunteer can see both the scanned document and the two sets of data that have already been entered. The third person's job is to check and decide on the right choice. So three independent persons review a scan, something that would be far too costly if you were not to work with a crowd.

And lastly, a question in return is: what do we consider an error? Sometimes an entry in our index looks like a mistake, but eventually turns out to be right. We will show you one example from the Militia registers.

Searching on the impossible year or birth 1, gives 33 results. So you would think these are 33 mistakes. But take for example Hendrikus Ebeling, who cannot have been born on June 12th in the year 1. As he is registered in 1861, and all boys had to sign on at the age of 19, it is obvious that Hendrikus must have been born in 1842, as were all the other boys registered on the same page. But the strict instruction for the volunteers is to copy exactly what they read in the scan and as you can see, according to the scan his date of birth is June 12th in the year 1. So in the index you can find him born in the year 1 and this is not a mistake made during indexing, but an error in the nineteenth century administration.

So even if you find strange years or dates of birth, you cannot be sure they are mistakes. In the Militia register index of over one million records, there are 316 of these 'impossible' years of birth, and this is only 0.027% of the index, a reassuring percentage. Especially if you consider that some of those 316 might even be indexed right.

So all in all we are very satisfied with the development and results of VeleHanden so far. Of course, there are still points of improvement and we still learn every day. ■



Nelleke van Zeeland
Stadsarchief Amsterdam
nvanzeeland@stadsarchief.amsterdam.nl



Ellen Fleurbaay
Stadsarchief Amsterdam
efleurbaay@stadsarchief.amsterdam.nl