Nelleke van Zeeland[1]

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

The crowdsourcing website velehanden.nl (literally: many hands) is online for almost five years now. This gives us the opportunity to look back and evaluate: how did we initiate the project, what were the difficulties and successes and how will we continue?

Amsterdam City Archives

Before going into the details of VeleHanden, some basic information on the Amsterdam City Archives and our online activities will provide the context and background of this project. 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 and how fast the number of online visitors was growing. Without much effort, in a few years' time, 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 more practical things like that. Then we saw an image database, probably Getty's. We fell in love with it and wanted one for ourselves. So in 2003 we had our own what we call 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. We set ourselves to get an overview and catalogue of our archival collections online. We named it: Archiefbank, or Archives Database. Our colleagues praised us, 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, because 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. That does not outnumber about 15.000 scans a week. That is why we started a 'scanning on demand-service'. Nowadays we produce those 15.000 scans a week. When a request for digitization is received it takes three to five weeks to produce the scans and bring them online. A lot better than hundreds of years…

---

[1] Stadsarchief Amsterdam / Amsterdam City Archives

'I cannot find it'

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 administrated in the past. They just want to type in a name or an address, more or less like they are used to in Google. In short they want an *index* to all those scanned documents.

For most archival documents optical character recognition is not an option (yet), because the documents vary a lot and they are often handwritten. So making an index on the topics, names or addresses in all these documents requires a lot of manual labour. 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, not just a few, but lots and lots of volunteers. That is how the idea for this crowdsourcing project was born.

Hopefully this short history illustrates the mission of the Amsterdam City Archives and how VeleHanden fits neatly into this mission. We want to facilitate online historical research and therefor everybody should be able to use our archival collections at home 24/7. As I have pointed out our mission provoked a more demand driven approach, asks to think big in large quantities and is embedded in a project based working method. In practical sense this mission resulted basically in two online activities: a scanning on demand service and a crowdsourcing platform. Here, I will not focus on scanning on demand, but only on crowdsourcing.

Public-private partnership

So let's zoom in on crowdsourcing. Now we had the idea for VeleHanden, but to actually make it, we needed help. 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 chose Picturae to develop and manage the VeleHanden website. Nowadays, when an archive or other cultural institution 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' goal 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. 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 test the tools for data entry and for quality checks and to test login-procedures and other facilities. In two days 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. In six weeks, 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. But even more important than the meetings and the surveys was the online 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 website.

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. 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 site. They do all the quality checks and they help new arrivals. Five years later VeleHanden counts more than 8000 members and it still grows daily. So I think we managed to create a crowd.

VeleHanden

Time to look at the platform itself. The website consists of a homepage with general information. Each project has its own page, on which you can find more information about the task, the type of source, specific news items, statistics and so forth. Most projects are genealogical projects, in which participants are asked to enter the names, dates of birth and addresses of persons (for example in population registers). Other types of projects are also possible: video tagging, describing photographs, matching scans with existing metadata and even georeferencing. On another page you can get an overview of all participants. They can join anonymously; it is optional to enter a full name, location or motivation to participate. Each project has its own Forum, where participants can

exchange information, ask for help or share curiosities about the project. Statistics about the progress of a project are a very important motivator.

The tool for data entry is the core of VeleHanden. Data can be entered once or twice, or even more times, if you would like to. For genealogical projects we chose for double data-entry: the same scan is being presented to two individual participants and afterwards to a third one for control. While entering data the participants can use different buttons: with *too difficult* they can skip a scan and get another one, with *remarkable* they can tell a project manager that they found something special on the scan and the button *unusable* is for example for empty pages or book covers.

As a project manager you can appoint the controllers. In our project we always work with controllers from the crowd, but many institutions have their control done by one of their employees. He or she gets to see the two entries with the scan, with the differences highlighted. The controller can choose which entry is the right one or correct and fill out the right entry. If even the controller is not sure, he can use another button to contact the Expert, who is usually someone from the archival institution.

Because we want to show the output of a VeleHanden project as soon as possible, we have our work in progress indexes. Each week, all data that has been checked, is being transferred to our online indexes: archief.amsterdam/indexen. This way, the crowd gets to see what they helped create and mistakes or errors are noticed quickly. The production from VeleHanden is directly taken in use.

What's in it for the crowd?

We saw that we 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 think the work has to provide three things to the crowd: pleasure, praise and profit.

First of all: it must be a fun to use the website. It must be user friendly and really fast. The task itself must be fun, like a puzzle. On top of that it can 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: therefor we weekly add the newly indexed names in a work in progress index on our website. This is great for our crowd, because if I enter a name today, I will be able to search and find that name next week. For us 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 organised workshops, guided tours and even a lottery, where people can spend points. Although it is not the decisive factor to participate, it is important to reward your crowd.

Next to the points we sometimes organise events. Two examples: we threw a party to celebrate the ending of our largest project so far. Of course we had beer, wine, snacks and an interesting programme, but we also had a personalized artwork made by a Dutch artist, Jan Rothuizen, for one of our special volunteers. Another time we went with a small group on a canal boat, to get inspired for a project on localizing monuments in Amsterdam. We do not organise such events often, but it is important to say 'thank you' to your volunteers and to give them attention in an offline environment once in a while.

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, involved and hard-working as they are now.

Results

I already touched upon the results shortly. Today, the platform counts about 8.000 participants, mainly from the Netherlands, but also from abroad, for example Brazil, the United States, Sweden, Senegal and Australia. I even counted one from Luxembourg.

Their production is overwhelming. If I take all VeleHanden projects combined, the crowd has processed 3 million scans, which equals several millions of names, dates and addresses. If I only look at the six genealogical projects from the Amsterdam City Archives, there are over 3,5 million names transcribed. Calculated in working hours, this would cost a regular fulltime archivist more than forty years! And then he or she would not even take breaks. What you see in all crowdsourcing-projects is also true for VeleHanden: most of the work is done by a small hard core.

But quantity is not as important as quality. The question most asked is: is crowdsourcing prone to error? The answer is no. Most volunteers are really precise. They know how important their work is, because they often use the results in the indexes themselves as well. Moreover, we build in control mechanisms like the double entry and control system. So there are six eyes looking at the same scan, where in the old-fashioned way, there were only two. On top of that, there are a lot of ways to ask for help: on the Forum or via the Expert.

I will not argue that crowdsourcing is error free, but it is definitely not prone to error. If we think we encounter an 'error' in our systems, it usually appears to be an error in the original documents. Here, I give two examples from the population registers 1853-1863. You would think Jan Hendrik Leers could not be born in 1709, and that a volunteer made a mistake, because this would mean he would be at least 144 years old. But the document supports the volunteer and reveals the mistake of the civil servant of that time. The same

goes for Hendrika Vogelzang who seems to have been born on a date that never existed: 30-02-1885.

All in all we are very satisfied with the development of VeleHanden and the results so far. Over the past five years we have also launched some innovative new types of projects, like tagging and georeferencing, next to indexing. We will continue to do so, because we experienced that there are many possibilities of working successfully with a crowd.