Martijn Storms*

Maps in the crowd: results of a map georeferencing crowdsourcing pilot project

Keywords: georeferencing; crowdsourcing; accuracy; manuscript maps; navigational charts; Van Keulen

Summary: In 2015 Leiden University Libraries organised a crowdsourcing project on map georeferencing, called *Maps in the Crowd*. This pilot can be seen as part of a broader investigation how to catalogue a cartographic collection in a geographical way. Next to a geographical access to digitised old maps by georeferencing the scanned images, a map library has to deal with the facilitation of born-digital geodatasets, the presentation of its collections on various platforms and the sharing of digital maps and geodata with other institutions.
In this paper the process, quality and results of the georeferencing project will be evaluated and analysed. Among others, the following questions will be answered. Who is *the crowd*? How accurate are the georeferences done by that crowd? How efficient and useful is the reviewing process? How accurate are the maps? Is it useful to georeference all maps? Are all maps 'georeferencable'? Finally, the future steps of the geographical cataloguing of Leiden University's map collections will be discussed briefly.

The Van Keulen manuscript chart collection of Leiden University Libraries

At Leiden University Libraries, a collection of 334 manuscript charts produced by the Amsterdam publishing house Van Keulen is kept. The navigational charts have a striking uniformity: with only a few exceptions, they are drawn on sheets of c. 59 x 99 centimeters on which the compass lines are pre-printed. Comparable large sets of these charts are kept in the Biblioteca Anglicana in Rome (85 charts) and Staatsbibliothek zu Berlin (120 charts). It is determined with certainty that the Leiden collection formerly belonged to the archive of the Van Keulen firm. Especially the charts with the annotation 'origineele' (original) can be regarded as documents from the firm's archive (De Vries, 2005: 92). How the manuscripts charts functioned in the commercial activities of the publishing house still remains unclear. They are regarded as a 'unique by-product', 'an entirely new product' and 'a new product for commercial cartography' (De Vries, 2005: 91; 99; 116). Provenance research and the analysis of the sources used for the manuscript charts suggest that these charts were produced as a commercial product, assumedly more for collectors than for practical navigation. There is evidence that even out-of-date original master copies were sold to collectors, partly to embellish printed atlases (De Vries, 2005: 93-104). The reason why the charts weren't printed should be found in a limited demand: 'In those instances where Van Keulen suspected there was a certain degree of interest, but did not consider this sufficient to warrant the investment demanded by a printed chart, he opted for manuscript works that could be supplied on commission' (De Vries, 2005: 104). In the comprehensive research project, resulting in *The Van Keulen Cartography* catalogue in 2005 the sources of the manuscript charts were partly analysed. Three kinds of sources were distinguished:

* Leiden University Libraries [m.storms@library.leidenuniv.nl]

1) Information derived from skippers and captains.
2) Published printed charts in maritime atlases by other contemporary publishers;
3) Works kept in Van Keulen's own firm archive. For example, a set of 24 charts of the West Indies (West coasts of Africa, Caribbean, and the coasts of Brazil) are copied from a now broken up Vingboons Atlas, once in the possession of the Van Keulen's.

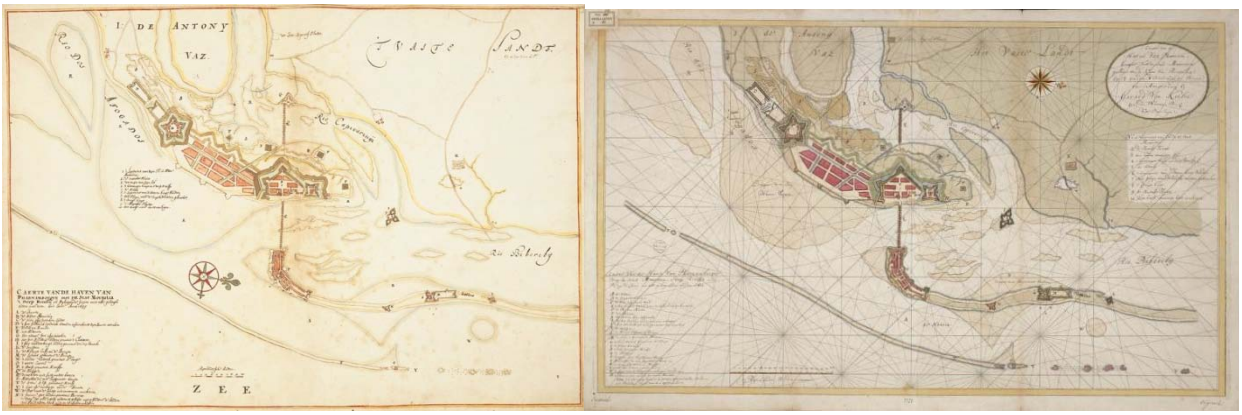

Figure 1: Van Keulen's chart of Recife, 1721 (right; UBL, COLLBN 003-09-015), copied after the fortification plan by Johannes Vingboons, 1639/1660 (left; Istituto Archeologico Historico e Geographico Pernambucano, Recife, nr. 41).

As far as it concerns copies of older maps, like those of Vingboons, one can question whether these out-of-date maps were salable. Again, they were certainly not bought for practical navigation, and at most by only a few collectors. For the maps related to Dutch Brazil, it is suggested that Van Keulen drew the charts to revive the memory of the period of Dutch rule in Brazil (Buve, 2008: 24-25). For only a sixth of the known total of Van Keulen manuscript charts, the sources are determined by De Vries (2005: 113). Further research to the sources of the charts may shed more light on the functions of this map collection. An unusual example is the 'chart' of Cyprus. This is an enlarged copy of the map of the island by Schenk and Valk (after 1676), including the decorated cartouche visualising the mythological birth of Aphrodite. Schenk and Valk copied the atlas map of Willem Jansz. Blaeu, 1634. We know that Van Keulen used the edition of Schenk and Valk because he added the borders of the four regions (Storms, 2011). Hopefully the origin of more Van Keulen charts will be traced in the future.

Figure 2. The map of Cyprus by Schenk and Valk (left; after 1676; UBL, COLLBN Port 174 N 256) and the derivative by Van Keulen (right; 1724; UBL, COLLBN 003-04-024).

Preparation of the pilot project

The reason Leiden University Libraries decided to start georeferencing its map collection is twofold. In the first place we want to make our collection better accessible. In anticipation to a new repository infrastructure that is in preparation, we formulated our desire to make our map collection searchable in a geographical way. An extra stimulus for this improvement of map cataloguing is the enrichment of our collection with two important collections of Dutch colonial maps.
In the coming years the image database of our map collection, and the recently obtained map viewer of the Royal Tropical Institute (KIT) map collection and the image database of the maps of the Royal Netherlands Institute for Southeast Asian and Caribbean Studies (KITLV) will be merged together within our new repository infrastructure (Storms, 2014). A second reason for georeferencing is that georeferenced maps can be used and analysed further in geographical information systems.
For the pilot georefencing project the Van Keulen collection (described above) was selected for two reasons: 1) This collection was catalogued and digitised almost completely. 2) The collection is relatively small and manageable. Nevertheless, the preparation of the project took a lot of time. In the first place, images of some of the charts were missing in our image database. The missing images were added and for some maps even new scans had to be produced. In some cases the catalogue record was updated.
The project is set up with the Georeferncer application of Klokan Technologies.[1] With this application comprehensive crowdsourcing projects are already started, among others with the map collections of the British Library and the David Rumsey collection at Stanford University. In the Netherlands we learned from the two projects that are executed with maps of Utrecht University Library, although that was not set up as a crowdsourcing project but as an education project with geodesy/geoinformatics students of Utrecht University of Applied Sciences (Heere & Van Egmond 2015).
Corresponding to the standards of Klokan, the metadata were delivered, as well as the digital images of the charts. A problem occurred, since a number of sheets were subdivided in two, four or even twelve maps. These sheets were digitally cut up in separate images, so these separate maps could be georeferenced individually. Separate metadata records were created for these maps.
For practical reasons this process was not executed for inferior inset maps, which are generally enlargement of parts of the map at the same sheet. As part of the brief metadata that is displayed in the Georeferencer application, links to both the catalogue record and the scan in the image database of Leiden University Libraries are added. Maps on subdivided sheets are linking to the metadata record and image database scan of the complete sheet.

[1] http://www.georeferencer.com

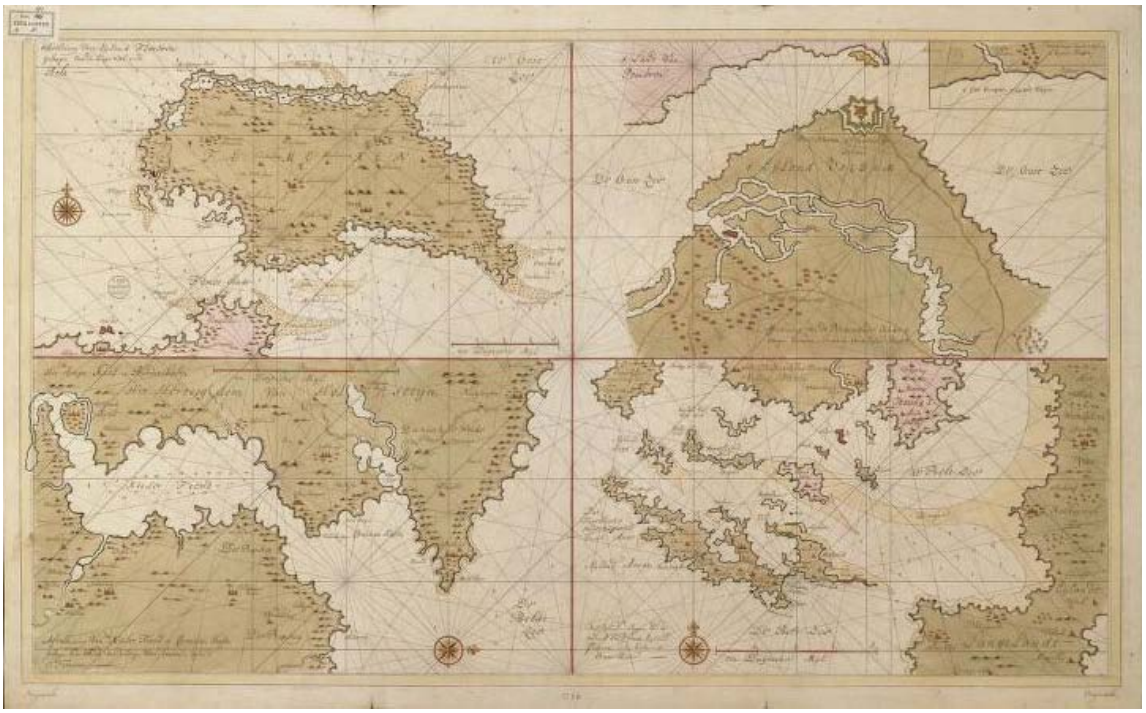

Figure 3: Four maps on one sheet: Fehmarn, Fort Peenemünde, Kieler Förde and the island Ærø and Langeland. Amsterdam: Gerard van Keulen, 1720 (UBL, COLLBN 003-01-011). The map of Peenemünde (upper right) has an inset map.

Results of georeferencing by the crowd

The pilot georeferencing project was a great success. In only two weeks time, all 393 maps were georeferenced by the crowd. It is hard to judge the quality of georeferencing by the crowd, because the accuracy of the original charts appeared to be quite low. In some cases it was hard to find the number of at least five corresponding locations on the modern map, necessary to georeference the chart and to apply the MapAnalyst accuracy analysis tool, included in the Georeferencer application. Although a tutorial and instruction video were available, not all participants georeferenced the maps as was the intention. Common mistakes that were made include:

1) Adding less than the minimum of five control points.
2) Linking of different locations (usually by accidentally clicking one time too much on either the old map or the modern reference map).
3) A few charts cover (parts of) the Pacific Ocean. In the Georeferencer application it is not possible to visualise a proper overlay of a map that crosses the International Date Line. This problem was pragmatically solved by only selecting control points on one side of the date line.
4) Clipping the map wrongly, by clicking on the control points in stead of the corners of the map image.
5) Inaccurately clipping, probably without zooming in to cut carefully along the borders of the map image.

Figure 4: On this chart of the Southern Pacific Ocean, only control points are placed in Latin America. Therefore the western part of the map could not be correctly placed. Tasmania on the old map is placed at the location of New Zealand in the overlay (Amsterdam: Gerard van Keulen, 1712, UBL, COLLBN 003-08-017).

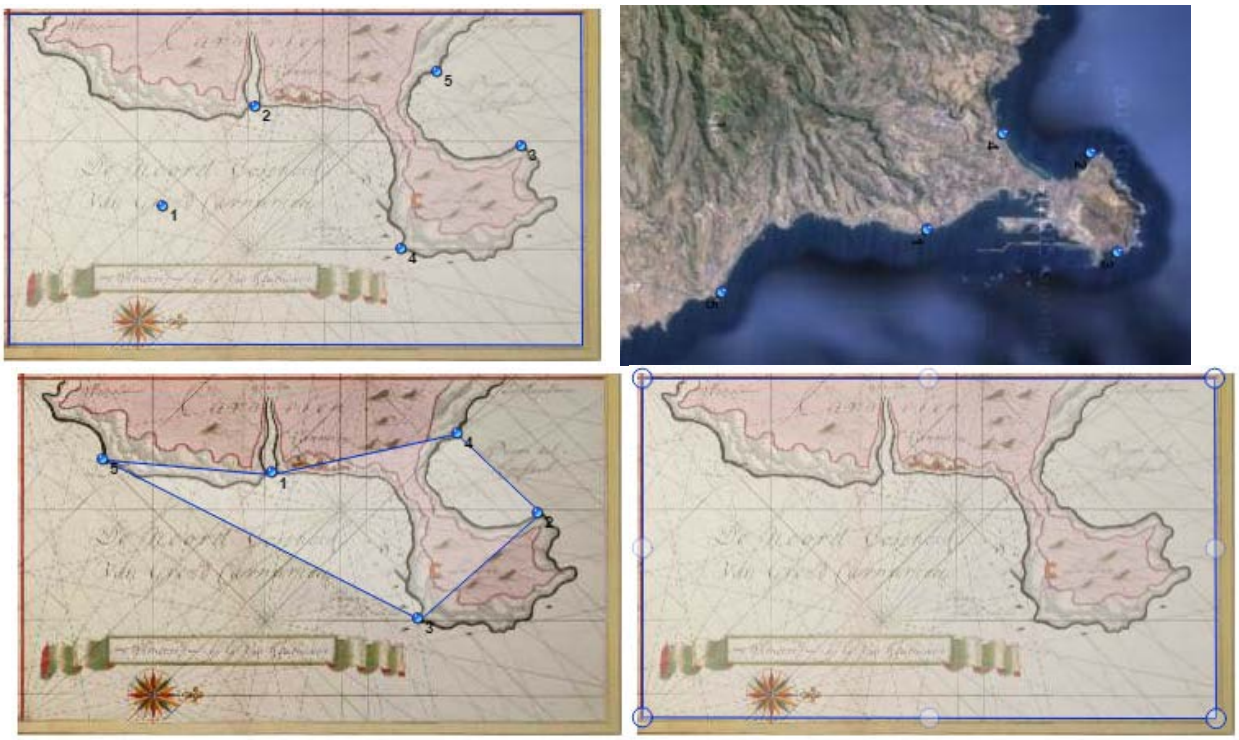

Figure 5. Example of linking different locations by mis-clicking (upper left), compared with Google Satellite (upper right, rotated for comparison) and wrong (lower left) and right (lower right) clipping of the map image. In this example the chart of Northeast Gran Canaria (Amsterdam: Gerard van Keulen, 1710, UBL, COLLBN 003-07-013: 4) is used.

The used instruction video led to some confusion, because in that example parts of the water were clipped away. The sea, sometimes with indications of depths, sandbanks, cliffs, and islands, is an inherent part of the map image. For the Van Keulen charts title cartouches and compass roses are

generally drawn in the map images, surrounded by sea. Therefore we didn't want to clip these parts away too. Only the framework and insets are not part of the georeferenced map image and had to be left out.

Most maps, however, were georeferenced more or less correctly. Nevertheless, it was decided to carefully review and adjust all georeferenced maps in this pilot project. However, in the Georeferencer application it is always possible adjust or improve the georeferencing of a map for every participant, even after the georeferencing is approved by the reviewer. In retrospective, in the reviewing process too much time is spent to adjust the georeferences.

In total, 41 participants contributed to the pilot crowdsourcing project. Strikingly, about 75% of the maps are georeferenced by the three most active contributors, where the top contributor took almost half of all georefencing work on his behalf. From experiences of other crowdsourcing projects it is known that '10% of the volunteers do 90% of the work' (Van der Sijs, 2015: 21). Moreover, a natural selection process takes place. Volunteers who have affinity with the subject will georeference more maps and do better than other participants who find it hard to georeference a map.

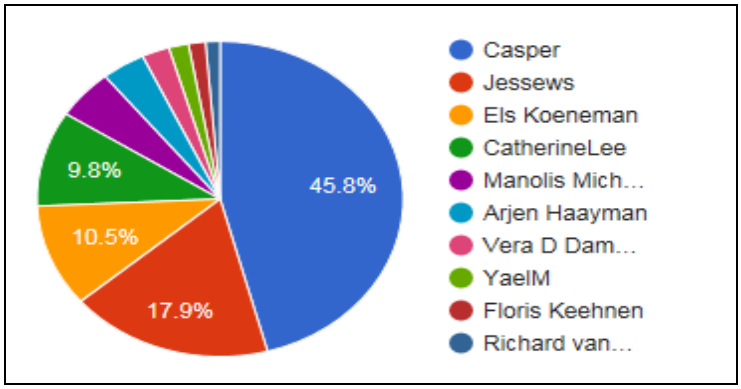

Figure 6: Diagram showing the top 10 contributors of the pilot project, according to the number of control points added.

After all 393 Van Keulen charts were georeferenced and approved by the reviewer a complete set of geographical data was available. A nice widget that is provided by Klokan is a dot map, where every dot indicates the (center of a) location of a georeferenced map. Although this is less sufficent for small scale maps of larger areas, it gives a good indication of the geographical distribution of the regions mapped on the, in this case, Leiden corpus of Van Keulen manuscript charts; especially, because the collection consists mostly of the large scale plans of single ports, bays and relatively small strips of coast. The highest concentration of charts is in the Mediterranean. A second cluster are charts of western and northern Europe, including the southern coasts of Norway, the Danish islands, the northern coasts of Germany, the Gulf of Finland and the coasts of northern France and the British Isles. A smaller amount of charts is concerned with the coasts of West-Africa, the Caribbean and South-America. Remarkably, charts of Asian coasts are completely lacking in the Leiden corpus. Probably these charts were already removed from the Van Keulen archive, before the manuscript charts were offered for sale to Leiden University.

Figure 7: Dot map indicating the geographical distribution of areas covered by the Leiden Van Keulen charts.

Accuracy analysis of the Van Keulen charts

The aspect of the metrical accuracy can now be analysed relatively easy in detail, because the MapAnalyst visualisation tool is incorporated in the Georeferencer application. This led to the conclusion that the Van Keulen charts are not very accurate in general. Especially the large scale charts and plans of single ports, bays and estuaries are remarkably inaccurate and not detailed. Although it was known that the charts for navigating ports were 'not generally conspicuous for their accuracy' (De Vries, 2005: 100), the accuracy of the charts on a smaller scale were not very accurate too. However, it can be stated, with some reservations, that the charts of larger areas are more accurate than the charts of small areas. A typical aspect of charts leads to higher inaccuracy values in the analysis: that is the practice that capes, bays and river mouths are exaggerated on navigational charts. In other words, these coastal charts do not have a uniform scale. If, for example, only capes are chosen as control points, the overall coastline will be located somewhat inland when viewing the overlay on the modern map.

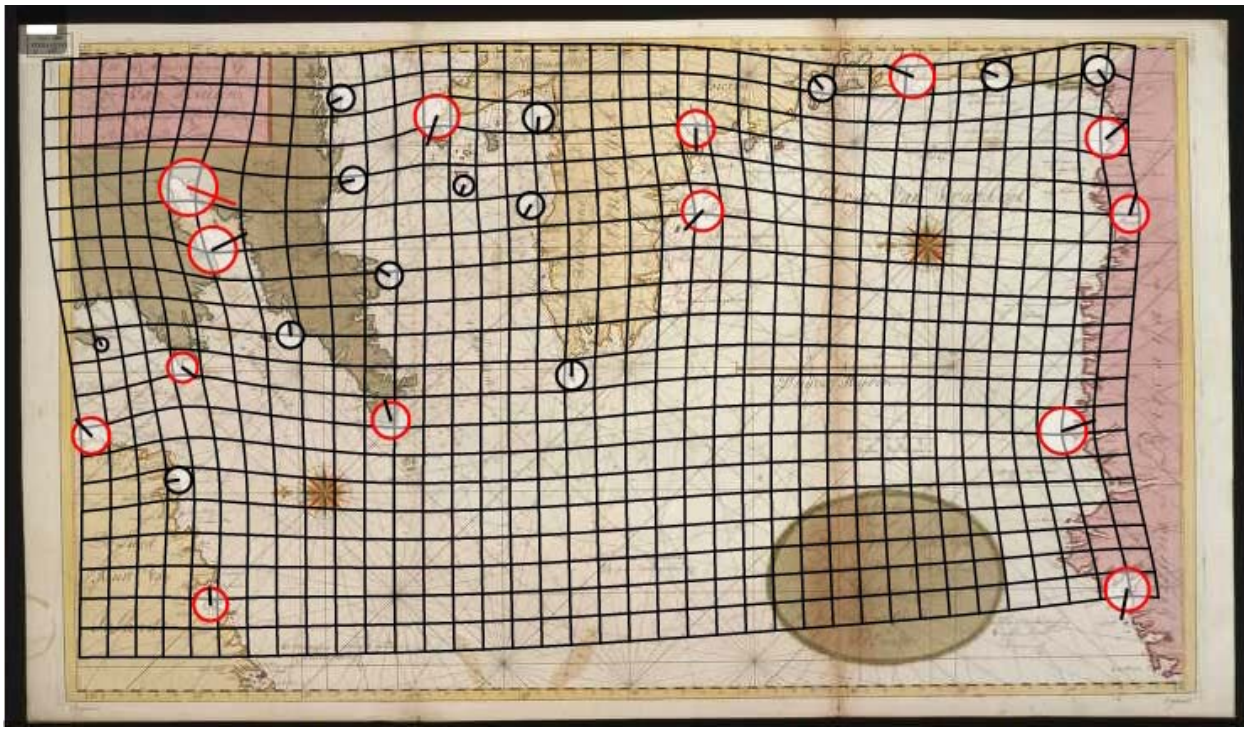

Figure 8: Distortion grid and displacement circles and vectors (MapAnalyst app.) projected on the chart of the Irish Sea, English Channel and Bay of Biscay. Amsterdam: Gerard van Keulen, c. 1720 (UBL, COLLBN 003-03-027).

Some of the large scale charts of small areas bear very few locations that can be found on the modern reference map. In some cases, these charts were almost 'ungeoreferencable', despite the location of the mapped region was known.

Figure 9: Almost 'ungeoreferencable'. Chart of the estuary of the river Po. Amsterdam: Gerard van Keulen, 1720 (UBL, COLLBN 003-06-016: 4).

Lessons learned and future plans

The pilot georeferencing crowdsourcing project 'Maps in the Crowd' is successfully completed in 2015. In two weeks time 393 charts were provided with geographical coordinates. However, the preparation of the project and the reviewing process took more time than was expected. The fact that some charts are very inaccurate and sometimes almost 'ungeoreferencable' should not be seen as a problem. With the addition of the coordinates they are now searchable in a geographical way, the main goal of this project. Inaccurate maps are probably not very useful for GIS analysis but that is up to decide for the researcher who want to use our georeferenced maps.

In the mean time we continued to prepare maps from our collections for Maps in the Crowd.[2] In February 2016 a set of 6678 Dutch colonial maps of Indonesia from the KITLV collection were made available. Because of this significant scaling up an intern was attracted to assist with the preparation work, and more attention was paid to publicity aspects. We now made our own instruction video (with subtitles in English), a general video about the project, posters and leaflets, and a blog[3] with regularly new information. The instruction manual is improved, with extra attention to avoid the common mistakes that were made in the pilot project. In collaboration with our communication officer, we made sure the media (local and national newspaper, local radio, magazines, social media and blogs) paid attention to our project. It is important to reach a wider public, because the amount of maps is now much bigger than in the pilot project. After one month almost 1,000 maps (c. 15%) were georeferenced.

Figure 10: Header of the Maps in the Crowd blog with logo. Detail of the town plan of Batavia by Mattheus du Chesne of 1652. Amsterdam: Johannes Janssonius, 1657 (UBL, KITLV D E 23,1).

[2] The members of the project team are subject librarian and GIS-specialist Patrick Gouw, curator of maps and atlases Martijn Storms, both of Leiden University Libraries and intern Jonathan Bos, masterstudent history at Leiden University.

[3] http://blogs.library.leiden.edu/mapsinthecrowd/

Literature

Buve, R. (2008), Een selectie uit de Nederlandse cartografie van Latijns Amerika, 17de en 18de eeuw in het bezit van de Universiteitsbibliotheek Leiden = Uma seleção de cartografia neerlandesa sobre a América Latina dos séculos XVII e XVIII em posse da Biblioteca de Universidade de Leiden. In R. Buve & M. Storms, *De Nederlandse cartografie van Latijns Amerika. Kaarten uit de Collectie Van Keulen en de Collectie Bodel Nijenhuis = A cartografia neerlandesa da América Latina. Mapas de Coleção Van Keulen e da Coleção Bodel Nijenhuis*. Leiden: Universiteitsbibliotheek Leiden, 22-59.

Heere, E. & M. van Egmond (201X), Het georefereren van oude kaarten door studenten Geodesie/Geo-Informatica. *Geo-Info* 12 (4): 26-32.

Storms, M. (2011), 'Van de beste eylanden eene' The Dutch mapping of Cyprus. In digital form, https://socrates.leidenuniv.nl/R/-?func=dbin-jump-full&object_id=2702145

Storms, M. UB Leiden zet Azië op de kaart. Complete kartografische collectie KIT blijft behouden in Leiden. *Geo-Info* 11 (2): 34-37.

Sijs, N. van der (2015), 7 vragen en antwoorden over crowdsourcing. Lessen van het Meertens Instituut. *Informatie Professional* 19 (8): 20-23.

Vries, D. de (2005), The manuscript charts, In D. de Vries, G. Schilder et al, *The Van Keulen Cartography Amsterdam 1680-1885*. Alphen aan den Rijn: Canaletto, 91-116.