

USING AND IMPROVING MAPATHON DATA THROUGH HACKATHONS

Serena Coetzee¹, Victoria Rautenbach¹, Cameron Green¹, Kiev Gama², Nicolene Fourie³, Breno Alencar Goncalves²,
Nishanth Sastry⁴

¹ Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria,
South Africa - {victoria.rautenbach, serena.coetzee}@up.ac.za, u15055452@tuks.co.za

² Federal University of Pernambuco (UFPE), Recife, Brazil - {kiev,bag}@cin.ufpe.br

³ Council for Scientific and Industrial Research, Pretoria, South Africa - nfourie@csir.co.za

⁴ King's College London, London, Great Britain - nishanth.sastry@kcl.ac.uk

Commission IV, WG IV/4

KEY WORDS: crowdsourcing, mapathon, hackathon, OpenStreetMap

ABSTRACT:

Geospatial data is essential for the development of the blue economy: for sustainable coastal management of coastal areas and to unlock economic potential from marine and ocean resources. In developing countries, such as South Africa, there are often gaps in the data with significant implications for the blue economy. We conducted a project aimed at addressing these data gaps by experimenting with a circular process where geospatial data for selected areas on the South African coastline were collected through mapathons and used in applications that were developed during hackathons. We validated this circular approach with two iterations of mapathons and hackathons, and found that 1) the size and location of the map area need to be carefully chosen; 2) those creating the apps needed a huge amount of help in dealing with the geospatial data; and 3) any geospatial data is useful for the blue economy, not only data with a very specific purpose in the blue economy context, such as coastal access points. Overall, the geospatial data usability improved from one iteration to another and would certainly improve if more iterations were added. Similar to the deployment of mapathons for disaster relief, future research could focus on hosting hackathons for the rapid development of apps to assist with disaster relief operations. Generally, the hosting of mapathons and hackathons in lockstep is a novel way of exposing students to interdisciplinary collaboration in international teams with a common goal.

1. INTRODUCTION

The blue economy refers to the sustainable use of marine and ocean resources for economic growth and improved livelihoods in coastal areas. Geospatial data, such as boundaries of protected areas, access points to the coast line, landmarks and points of interests, is essential for the development of the blue economy. Such data is required for the management of coastal areas to ensure that resources are used in an effective and sustainable manner. At the same time, it is needed to unlock economic potential and to ensure that optimal value can be gained from marine and ocean resources for the development of the blue economy. In developing countries, such as South Africa, there are often gaps in the data with significant implications for the blue economy.

We conducted a project aimed at addressing these data gaps by experimenting with a circular process where geospatial data for selected areas on the South African coastline were collected through mapathons (portmanteau of 'map marathon') and used in applications that were developed in hackathons.

A mapathon is a collaborative effort aimed at collecting specific map data through remote mapping in places where OpenStreetMap (OSM) data is scarce or non-existent (Coetzee et al. 2018). For example, the Humanitarian OpenStreetMap Team (HOT) (<https://www.hotosm.org/>) and Missing Maps (<https://www.missingmaps.org/>), regularly engage volunteers in mapathons to assist with geospatial data collection for relief operations following a

disaster; and YouthMappers, is a world-wide network of university chapters engages student volunteers in mapathons to collect data for unmapped areas, to improve data locally, or to assist with disaster response through HOT (<https://www.youthmappers.org/>).

The scalable collection of map data through mapathons has been demonstrated, but the quality of data collected through mapathons has been debated (Haklay et al. 2010; Mooney & Minghini, 2017). Authors have suggested ways of improving the quality elements of precision and accuracy of such data (Shahid & Elbanna, 2015). In this project we focus on the quality element related to the usability of the data for geospatial applications in support of the blue economy. According to ISO 19157:2013, *Geographic information – Data quality*, usability is based on user requirements and can be described with reference to other quality elements (completeness, logical consistency, positional accuracy, thematic accuracy and temporal quality) or by providing specific quality information about a dataset's suitability for a particular application.

Hackathons are events with intensive short-term efforts in which programmers and participants of varied background intensively collaborate to produce proof-of-concept (i.e., not production-ready) applications (Kommssi et al., 2015). Such intensive efforts can be enormously useful for bringing out creativity and co-ordinating the human capital of a large number of participants from interdisciplinary backgrounds. Mapathons result in new geospatial data, available to anyone, and such data needs to be of an appropriate quality to be reusable. For our project, we

explored the use of hackathons for assessing the usability of geospatial data.

The goal of the project was to create and refine a process of scalable collection of high-quality map data. The key idea behind this international collaboration project, borrowing from other citizen science and crowdsourcing approaches, was to mobilize volunteers to scale up collection of high-quality geospatial data suitable and usable for application development in support of the blue economy.

In the remaining sections of the paper, we describe the circular process (section 2), present results from two iterations of this process (section 3) and discuss and conclude with lessons learnt (section 4).

2. CIRCULAR PROCESS

We conducted a circular process that started from an intensive mapping effort to create new data for unmapped coastal areas. We then crowdsourced the development of apps that use this data in a hackathon and provided feedback about the difficulties and successes to a second mapathon that focussed on improving the usability of the data for app development. We developed metrics to analyse the quality of the data collected and used this to evaluate how the maps improved from their initial state after the first mapathon, and from the first mapathon to the second mapathon.

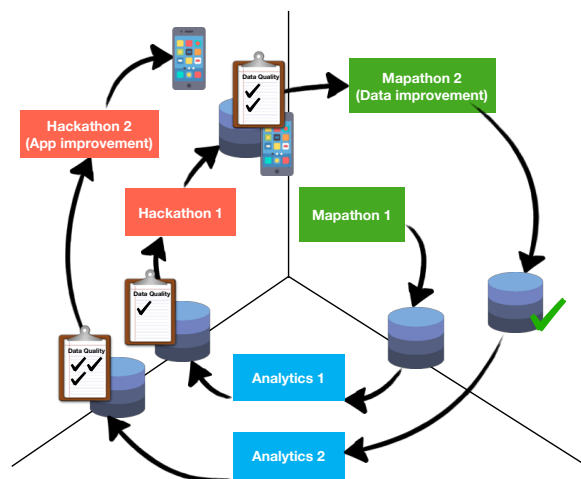


Figure 1. Circular process for collecting geospatial data in mapathons and assessing its usability through hackathons

Therefore, the main goal concerned the proposition and evaluation of a systematic circular process for collecting geospatial data and assessing its quality (Figure 1). The proposed process consists of using mapathons to scalably collect data, then have its usability validated by constructing apps of various purposes (related to the blue economy) in a hackathon, that will, in turn, help feedback enhancements on the geospatial data collected. The target domain to be explored as a pilot for this approach was in the development of the blue economy (Fourie et al., 2018).

3. RESULTS FROM TWO ITERATIONS OF THE CIRCULAR PROCESS

In this section, results for two iterations of the circular process are presented. Each iteration consisted of a

mapathon followed by a hackathon. The results from each event and iteration informed how we planned and conducted subsequent events.

3.1 First iteration

The first mapathon took place in March 2018 in South Africa over two days and was attended by 19 students who mapped over 2300 features in OSM (See Figure 2). The students were all enrolled for a final year geoinformatics module at the University of Pretoria, South Africa. The majority of the students had participated in a previous mapathon where they learned the basics of the mapping features using OSM.



Figure 2. Mapathon participants in action



Figure 3. Locations of Agulhas and Mossel Bay in South Africa. Source for map of South Africa: UN Office for the Coordination of Humanitarian Affairs (OCHA)

Because the development of the blue economy was the target domain to be explored, we consulted a draft Coastal Management Plan for the City of Cape Town (2012) in order to identify relevant features to be mapped for the mapathon. The plan identified cul-de-sac, car parks, boat launch sites, roads and walkways as important features for determining access points to the coastline. It was decided to map these features for the coastal area between Agulhas and Mossel Bay (Figures 3 and 4). This area includes at least two national parks and is not as developed as other parts of the coastline, such as the Garden Route. Unlocking the potential of the blue economy, while at the same time ensuring sustainable coastal management is therefore relevant in this area.

Participants were tasked with mapping cul-de-sac, car parks, boat launch sites, roads and walkways from aerial or satellite imagery provided in the OSM id Editor (e.g. Bing imagery, Digital Globe and South African CD:NGI imagery). The mapathon area was divided into a grid of cells using TeachOSM (www.teachosm.org) so that each student could map a different part of the mapathon area.

These features to be mapped can be difficult to identify on imagery and we were not surprised by the relatively small

number of features mapped over two days. The quality of the features mapped was analysed using the Java OSM (JOSM) validation tool that makes use of the Java Topology Suite (JTS). We found four errors in the data contributed during the first mapathon and 239 warnings (e.g. unconnected roads). Features flagged with a warning had to be reviewed manually to ensure that they were mapped correctly. This check would be done by an experienced mapper during the validation process.

Lessons learned from the first mapathon include the fact that the relatively large mapathon area (6,953.6km²) and the features to be mapped resulted in a low rate of data contributions. Additionally, very little (if any) descriptive information was added for the mapped features. This can be explained by the fact that when features are mapped during remote mapping, the mapper identifies a feature from aerial or satellite imagery and then traces its shape before adding it to OSM. A mapper usually has little to no knowledge about the area being mapped during a mapathon and will thus not be able to add descriptive information (i.e. attributes) to the feature. For example, from imagery a mapper can readily identify a building's location and shape but cannot detect the name or current use of the building.

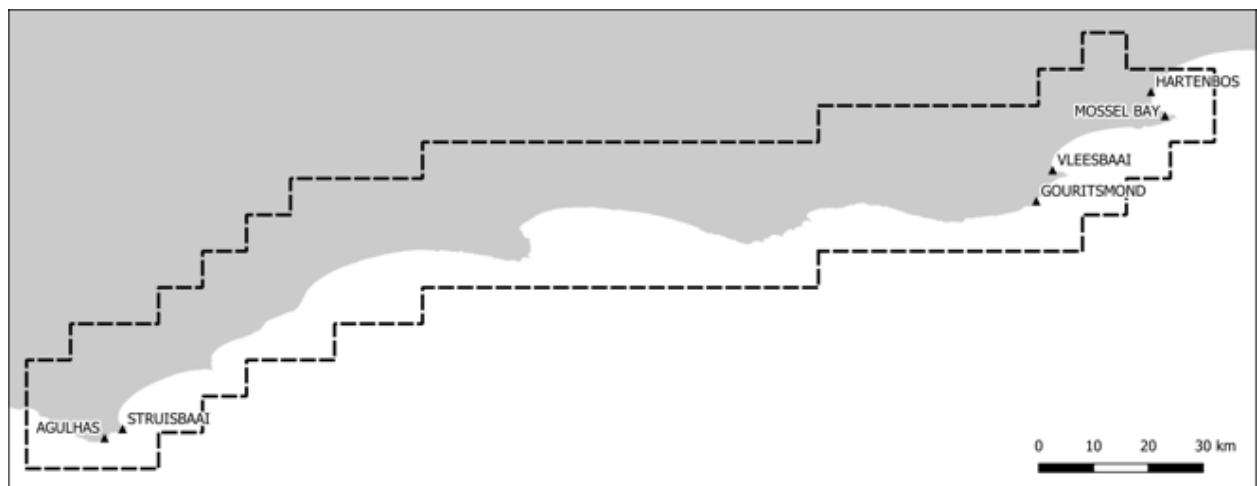


Figure 4. Coastal area for which data was collected in the first mapathon

The first hackathon took place in Recife, Brazil, in August 2018. There were 40 participants, most of them university students from varied backgrounds: architecture, cartography, computer science, computer engineering, design, geography and information systems. Although the event had more than 80 registrations, space restrictions limited the number of participants to a maximum of 40 people (Figure 5). The strategy to select the participants was to evenly distribute them among two groups: one with people experienced with maps and geospatial data; and another one experienced with software development.

There were four mentors with background on app development and experience in hackathon organization, but no one with experience in geospatial data (except student participants). Before developing the app, hackathon participants were introduced to the concept of a blue economy in South Africa and went under the following steps in a guided process: (1) Choosing a topic of interest; (2) Domain research; (3) User needs identification; (4) Challenge definition; (5) Solution proposal & development;

(6) Pitches. All seven groups that were formed delivered their apps. The resulting apps targeted two different domains related to the blue economy: fishing and tourism.



Figure 5. Hackathon participants in action

Based on feedback from participants in the first hackathon, a major stumbling block for using OSM data was the limited usefulness of the data collected during the first

mapathon for developing apps in support of the blue economy. Some participants reported scarce information about public restrooms, commercial and historic buildings. Two groups chose "fishing" as a theme but all OSM data referred to land features. Information about fishing spots, piers/berths which could be useful for the blue economy context, was not available. Generally, the groups reported a very low feature density in the mapathon area and suggested delineating a smaller area with more information for future mapathons. All these limitations in the OSM data led them to look other data sources.

The lack of a geospatial expert on site at the hackathon had an impact on the data used in the apps. The strategy to balance the participants based on their background create teams with complementary backgrounds. This, in part, counterbalanced the absence of a mentor expert in geospatial data, since each team had someone with that background.

3.2 Second iteration

The second mapathon was a one-day event in August 2018 attended by 16 students. All the students were returning mappers and had participated in more than two mapathons before. Based on the analysis of the data from the first mapathon and the feedback from the first hackathon, we decided to focus on easy to map features (i.e. buildings and roads) in a smaller area in the Mossel Bay Local Municipality (Figure 5).

During the second mapathon, 7647 new features were added and 326 features were updated. The majority (91%) of the added features were buildings. The reason for this would be that buildings are easy to identify on imagery, especially in the chosen urban area.

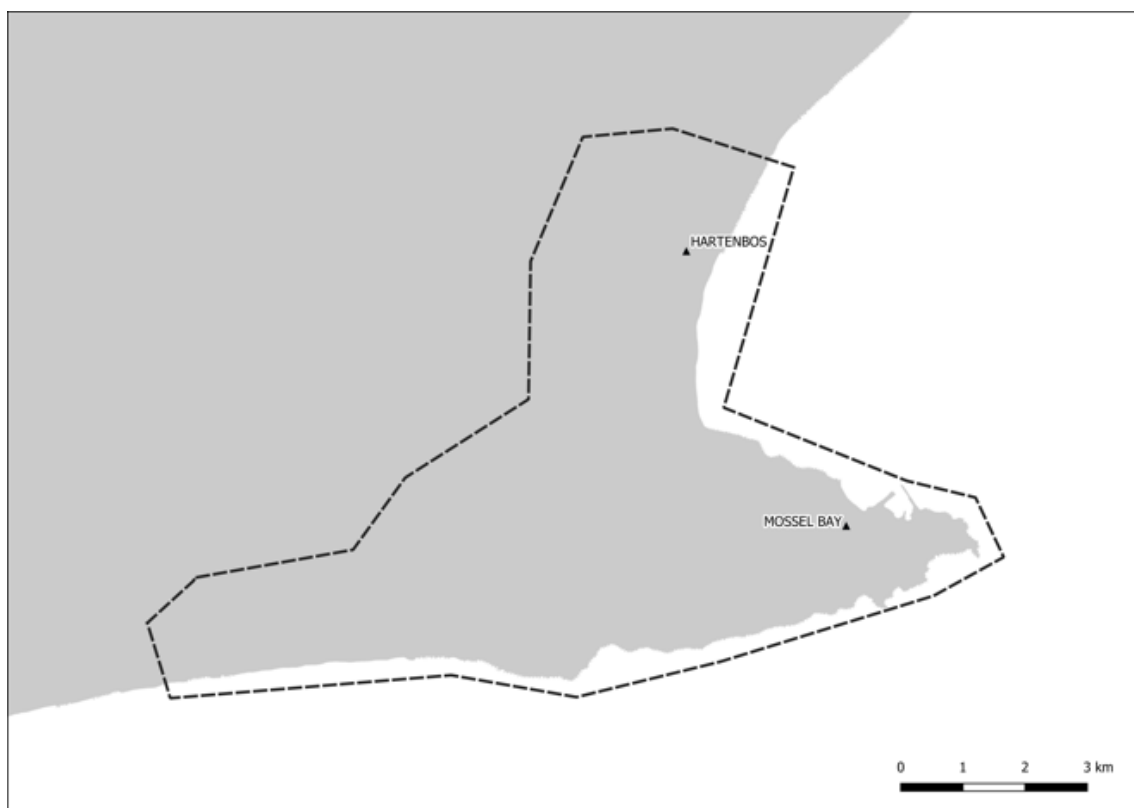


Figure 5. Coastal area for which data was collected in the second mapathon

For the quality evaluation, we did not use the JOSM topology validator for this mapathon, but rather focussed on assessing the quality of tags (i.e. descriptive information) added. We only found 5 features that were incorrectly tagged, in all cases on purpose, e.g. one student added their personal information as descriptive information for two points of interest in the area. These errors were corrected after the mapathon. Additionally, we also evaluated features what were updated, based on the logic that features update frequently would be more correct than features mapped for the first time (Haklay et al., 2010; Mooney and Corcoran 2012).

In terms of number of features added, mappers were more productive in the second mapathon, because the features

were easy to identify and readily present in the mapathon area. We also found the smaller area to be more manageable and the spatial distribution of the contributions was denser than in the first mapathon.

Taking lessons learnt from the first hackathon into account, a second hackathon took place in Brazil in January 2019. There were two major differences: (1) the presence of an geospatial expert who came from South Africa; (2) the usage of OSM data was mandatory. There were 20 participants and the event followed the same structure as in the first hackathon. Four apps were produced, all of them targeting tourism. Apps were mostly using data such as restaurants, hotels and general points of interest (Gama et al. 2019).

We collected feedback from three participants who had also attended the first hackathon. They considered the smaller area in the second mapathon to be a major advantage. They also thought the presence of a geospatial expert had two significant advantages. The first one, as expected, concerned the support with OSM and geospatial data generally, which was easier with examples given by the expert, as well as on site guidance on what type of data to use. The second aspect was not expected, concerning the fact that the geospatial expert became a stakeholder able to partially validate user needs in each app. The fact of having a South African provided access to someone with a potentially better understanding of the target audience. This fact made the participants more confident about their app than in the previous hackathon.

4. DISCUSSION AND CONCLUSION

For this project, we validated a circular approach to mapathons and hackathons by targeting sustainable growth of coastal areas, i.e. the blue economy. We believe this process has been successful, although with limitations. It was possible to generate apps with relevance to the studied context and to provide feedback about the data generated in OSM, allowing to adapt the objectives on a second mapathon, as the process expected.

We experimented with a circular process where geospatial data for selected areas on the South African coastline were collected through mapathons and used in applications that were developed in hackathons. We went into the process with the idea that there needs to be a link between creating new data and creating applications based on the data. However, we learned from the process that 1) the size and location of the map area need to be carefully chosen; 2) those creating the apps needed a huge amount of help in dealing with the geospatial data itself, as they had not used geospatial data in apps before; and 3) any geospatial data is useful for the blue economy, not only data with a very specific purpose in the blue economy context, such as coastal access points.

We found that mapping efforts concentrated in smaller area that correlated to the number of mapathon participants to be more useful for our purpose. Additionally, the availability of unmapped features and the ease with which they can be identified also plays a role in the mapathon productivity. Choosing an area in which geospatial data is more readily available was also more useful for app development during the hackathons. Finally, the geospatial expert who attended the second mapathon greatly helped the developers and a positive side effect of was the additional role he took on as of stakeholder who could validate apps targeting his country.

The circular process was a useful learning experience. In future, we will be better prepared for the fact that students with limited geospatial background need additional guidance for using geospatial data in app development. Similar to the deployment of mapathons for disaster relief, future research could be focused on hosting hackathons for the rapid development of apps to assist with disaster relief operations. The intensive efforts resulting from hackathons can be enormously useful for bringing out creativity and co-ordinating the human capital of a large number of participants from interdisciplinary backgrounds. At the same time, this would help raise awareness of the usability and suitability of the geospatial

data that is being collected. Generally, the hosting of mapathons and hackathons in lockstep is a novel way of exposing students to interdisciplinary collaboration in international teams with a common goal.

ACKNOWLEDGEMENTS

The research in this project was supported by the Royal Academy of Engineering Global Challenges Research Fund through the Frontiers of Development Programme.

REFERENCES

City of Cape Town, 2012. *A preliminary assessment towards a Coastal Access Management Plan for the City of Cape Town in terms of section 18(1) of the Integrated Coastal Management (ICM) Act (No. 24 of 2008)*, Draft. City of Cape Town Environmental Resource Management Department.

Coetzee S., Minghini M., Solis P., Rautenbach V. and Green C., 2018. Towards understanding the impact of mapathons—Reflecting on YouthMappers experiences. *International Archives of the Photogrammetry, Remote Sensing and Information sciences*, Vol. XLII-4/W8, 35–42.

Fourie N., Coetzee S., Rautenbach V., Green C., Gama K., Goncalves B.A., Sastry N., Agarwal P. and Xuehui, H., 2018. Generating Coastal Area Data with Mapathons and Hackathons. *SciDataCon-IDW 2018*. Gaborone, Botswana, 5-8 November 2018.

Gama K., Rautenbach V., Green C., Goncalves B.A., Coetzee S., Sastry N. and Fourie N., 2019. Mapathons and Hackathons to Crowdfund the Generation and Usage of Geographic Data. *Intl. Conf. Game Jams (ICGJ), Hackathons and Game Creation Events*, San Francisco, March 2019. <https://doi.org/10.1145/3316287.3316293>.

Haklay, M., Basiouka, S., Antoniou, V. and Ather, A., 2010. How Many Volunteers Does It Take To Map An Area Well? The validity of Linus' law to Volunteered Geographic Information', *The Cartographic Journal*, 47(4), pp. 315–322.

ISO 19157:2013, *Geographic information – Data quality*. International Organization for Standardization (ISO), Geneva, Switzerland.

Komssi M., Pichlis D., Raatikainen M., Kindström K. and Järvinen J., 2015. What are Hackathons for? *IEEE Software* 32, 5 (2015), 60–67.

Mooney, P. and Corcoran, P., 2012. Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, 4, pp. 285–305.

Mooney P. & Minghini M., 2017. A Review of OpenStreetMap Data. In: Foody, G. et al, *Mapping and the Citizen Sensor*. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bbf.c>

Shahid, A.R. and Elbanna, A., 2015. The Impact of Crowdsourcing on Organisational Practices: The Case of Crowdmapping. *ECIS 2015 Completed Research Papers*. Paper 166. ISBN 978-3-00-050284-2
http://aisel.aisnet.org/ecis2015_cr/166 (1 April 2019).