Data driven ways to characterise human behaviour

Our personal data is often hidden away in siloes and not utilised to our benefit. If it could be liberated from these 'big data' silos and innovative privacy-preserving algorithms used to unlock its potential, it can provide both social good and a winning service for everyone.

We talk to Assistant Professor James Goulding and Dr Mark Iliffe about data-driven ways to characterise human behaviour.

What is Neodemographics, and how has the project developed?

The initial Neodemographics project was funded by Horizon four years ago. We built collaborations with large UK companies (such as Boots, Marks & Spencer and Tesco) to develop a series of novel data-driven techniques to shed new light on consumer behaviour. One example approach, based on "dynamic topic modelling", scours data for the underlying purchasing trends cutting across the market. Using these trends as building blocks, and re-assembling them in different proportions, it can characterize the distinct makeups of shoppers in a way that incorporates change over time. The team along with our private sector partners continue to explore where these exciting new insights take us, from new product development and consumer experiences to greater understanding of customer lifetime value.

Following the success of these initial approaches however, we were then able to take these 'Neodemographic' techniques outside of UK markets, and explore how they might help emerging economies through an ESPRC funded project called "Opening Developing World Markets by Using Personal Data and Collaboration".

What were the aims of this project, and what have you achieved so far?

A key factor in meeting the UN 2030 Agenda for Sustainable Development, is actually measuring progress. However, intelligence of this nature is traditionally extremely hard to come by in emerging markets – censuses, social infrastructures and open data often just don't exist in these regions. This hinders not only local companies, but seriously impacts on human and economic growth, deterring foreign investment from entering the market. Lack of geo-demographic intelligence in the developing world can also seriously hamper the progress being made by government organisations and NGOs, such as Red Cross and the World Bank.

And yet, while emerging countries are often infrastructurally poor, they are data rich. Most of the population in Tanzania, for example, own a mobile phone and carry out payment transactions with the device. We recognised that by engaging the private sector, we could leverage digital footprints to create new social data sets of a comparable quality, if not better, than traditional census and survey data.

This has grown from strength to strength and we now work with a wide network of data providers in these emerging economies to investigate just what market knowledge can be derived from that digital footprint data. We've fused mass transactional event logs with open geospatial data and developed new mathematical techniques to generate novel intelligence 'layers' - mobility models, transport networks, financial flow maps and new forms of data-driven geo-demographic segmentations - all aimed at supporting decision makers and business growth, and changing the way in which we think about collecting market intelligence. At the same time we were aiming to meet the goals of the sustainable development agenda.

What other projects have developed from the original research?

In the UK we have worked with psychologists, Boots PLC and a nutrient tracking company Nutracheck, to analyse patterns in behaviour when people make healthy or unhealthy eating choices. We've had a knowledge transfer partnership with a marketing company, Krow Communications, where we explored the space between business data, personality and marketing. These have actually now combined to a new project called 'machine-learned personality' – a novel method that can use big data to measure personality and predict economic and social behaviour.

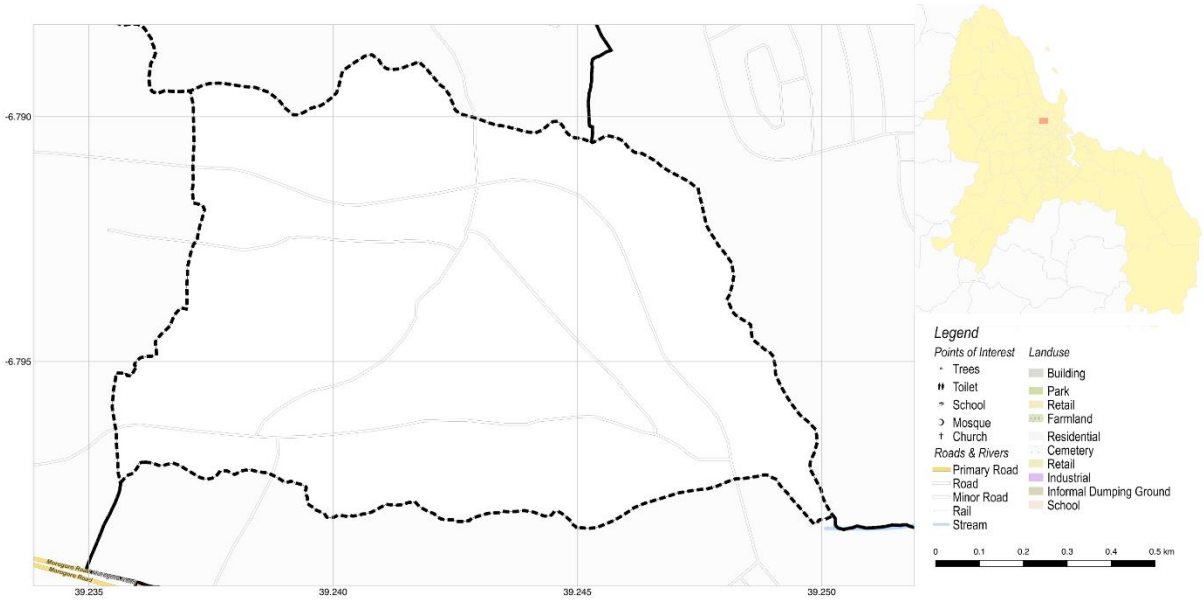
We have also received funding from the British Council/Newton Fund to model the transmission of dengue fever in Malaysia. Currently we have built predictive models (from climate and social media data) with around 80% accuracy in predicting dengue levels. The next stage of the project will bring in mobile phone data and environmental data (such as information on low-lying water where the mosquitos lay their eggs), to help predict where the next outbreaks may occur.

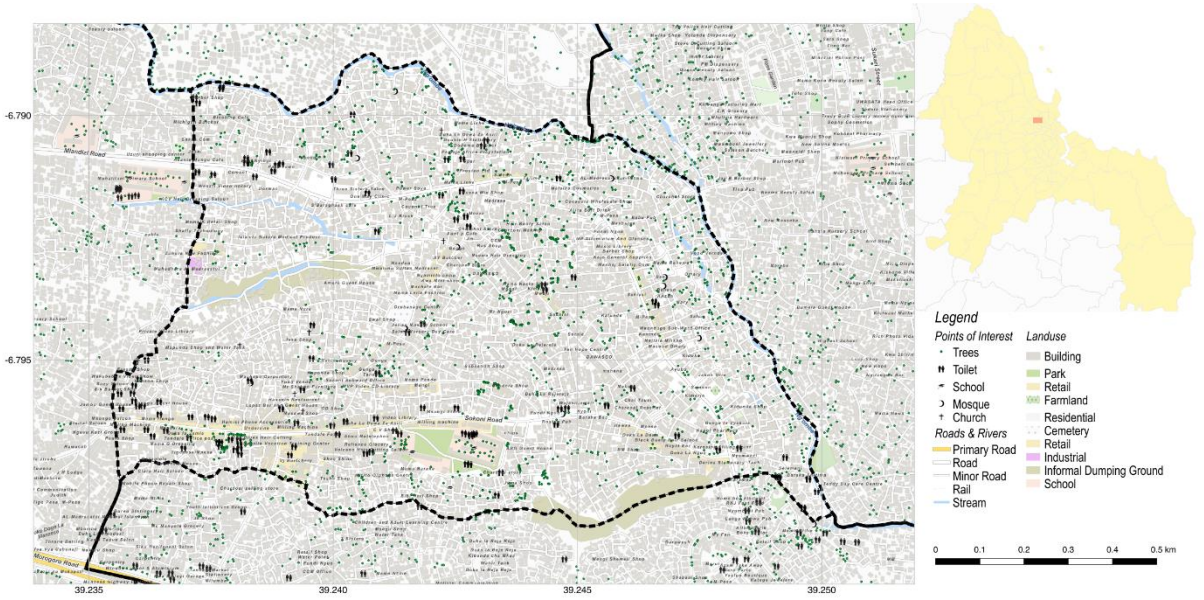
We were also awarded an ESRC-NERC-DFID grant for 'Big Data for Flood Resilience in East Africa' where the team have investigated how digital footprint data can predict, identify and inform decisions during flood and other natural hazard events. The research informed DFID's Big Data for Climate Change and Disaster Resilience Report (publication to follow).

We also have an ongoing programme of using big data for development, and this has been recognised by the Gates Foundation in awarding us funding for Financial Data Mapping in East Africa. A new tool for understanding financial accessibility is being developed as part of this work. This will combine new data streams from transactional records and aerial imagery, developing new mathematical and machine learning techniques. Ultimately, this will serve to illuminate areas of financial deprivation, supporting policy makers and financial service providers to resolve the challenge of access to financial services.

In a linked project, team members have also worked extensively with the World Bank on the Ramani Huria (http://ramanihuria.org/) - an effort to map over 1.3 million residents in Dar es Salaam, Tanzania, Africa's fastest growing city. As part of this, we have collected vast amounts of geodemographic data through community mapping of neighborhoods (crowdsourcing information from local community members). We've built on these open data outputs with the mining of mobile phone data (through collaboration with an African telecommunications company, TiGo). Thanks to this experience, we have been able to extend the work via funding from Horizon to analyse aerial imagery of the whole city collected from drones and, in partnership with companies such as IBM research Africa and DigitalGlobe, we are examining how these datasets can automate analysis of road conditions, land use, and the serious issues of

slum expansion due to rapid urbanization.



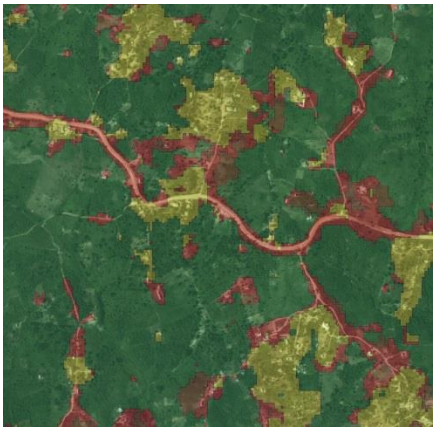


All map data © OpenStreetMap contributors

Give us an example of impact from your research…

As part of our research have helped scale out the mapping in Dar es Salaam out to cover 3.5 million people - around half the population, and this has provided a de facto baseline map of the fastest growing city of Africa. Dar currently has a population of 5 ½ million people, but eventually it is predicted to become a mega city of over 10 million in the next 15 years. This unplanned rapid urban expansion, along with global warming, has resulted in the city suffering biennial flooding with no adequate draining network.

The extensive mapping has enabled us to identify the most flood prone areas, predict impacts and work with individual communities to build community resilience groups. These will help maintain infrastructure, clean drains and provide dumping sites to reduce the risk and impact of flooding. It has also pinpointed priority areas for development organisations and the government to build infrastructure.



In addition, analysis of mobile phone data has allowed us to generate mobility models for the city of unparalleled granularity, which is now informing transport policy in the region. Much of the data we have contributed to has emerged from community mapping, and is entirely free and open on Open Street Map.

Since the launch of Ramani Huria in 2015 and as of Feb 2017:

Over 1.3 million citizens affected

29 neighbourhoods mapped

1254 km waterways mapped

3396 km roads mapped

450 mappers trained

10 disaster prevention teams established

What is the key learning from the international work?

Two or three years ago the country would have depended on traditional government agencies to solve the developmental problems, but there is now the growing realisation that this is not the way forward, and we can provide a shortcut through the process by mediating with private sector bodies and NGOs. By training the next generation in new digital approaches, they will return to their countries and be ready to support the development of new processes within government institutions that are going to be better, leaner, and more efficient than the current traditional slow-moving ones.

The key opportunity is working with communities in the places where we carry out the research, and this includes academic institutions. We are therefore building a network across East Africa that can come together and do this sort of work with us.

And the future?

Future research challenges involve investigating how such community-curated open data becomes sustainable over time, how we can generate business models that encourage the private sector to contribute data for development, and understanding how the skills gained by community members and government officials can be leveraged for wider societal and economic benefit.

The problem is that as fast as a city can be mapped and citizens' behaviour understood, rapid urbanisation ensures that data is out of date – even with Big Data approaches. We are now embarking on a project that brings together all the datasets (satellite, drone images, call data records) and by partnering with other organisations, will build models of not just how a city has been or how it is *now*, but how it *will be* in the next few years. We believe these near-future predictions will be invaluable for informing successful policy decisions such as transport planning and health.

We are also very encouraged by the potential - if we can achieve so much in Tanzania, what about other African countries and other emerging nations? We are now expanding the Neo-demographics approach to South Africa and Kenya, as well as building relationships in Botswana, Uganda, Malaysia, Vietnam and Cambodia as potential areas for future work collaborating with different partners.

Any final comments?

Although we have worked on vastly different application areas – health and wellbeing, marketing, community mapping - they are all underpinned by a common process and goal - to unlock the potential of doing social good via the personal data locked away in silos and unavailable to us.

As Neodemographics developed and started to produce real economic, societal and policy benefits, we recognised a need to provide our services on a larger scale, and are now training the next generation to take the skills into industry and other organisations. To this end the University has established the N-LAB as a permanent centre of excellence for International Analytics