

Institution: University of Nottingham		
Unit of Assessment: 17 - Business and Management Studies		
Title of case study: Neodemographics – Informing UN Sustainable Development Goals via “Big” Consumer Data and AI Techniques		
Period when the underpinning research was undertaken: 2014 onwards		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI
Dr James Goulding	Assistant Professor	2007-present
Dr Gavin Smith	Assistant Professor	2011-present
Dr Bertrand Perrat	Assistant Professor	2018-present
Prof Andrew Smith	Professor	2003-present
Period when the claimed impact occurred: 2016 onwards		
Is this case study continued from a case study submitted in 2014? N		
<p>1. Summary of the impact</p> <p>A multi-disciplinary research programme using mass consumer and citizen activity data (complex ‘big’ data sources, such as mobile banking, cell-tower activity, and purchasing patterns). In combination with novel methodological innovations in statistical modelling and machine learning, this work has produced otherwise inaccessible socio-demographic indicators for UN Sustainable Development Goals (SDGs). These have underpinned large transportation programmes (World Bank), producing measurable benefits to both citizen mobility and CO₂ reduction in Dar es Salaam, Tanzania (SDG11). Furthermore, the research has led to establishment of new systems and best practices in local government in East Africa (Zanzibar Department of Roads); contributed to health and wellbeing via Zanzibar Ministry of Health’s perinatal mortality assessment program (SDG3); and in 2020, began use by Anti-FGM organisations to help rescue girls in the Mara region of Tanzania (SDG5).</p>		
<p>2. Underpinning research</p> <p>A dearth of demographic data continues to seriously impede SDG efforts in Africa, leaving decision makers in the dark. The World Bank’s chief-economist has called this a “statistical tragedy” (UNFPA), with traditional data collection processes (e.g. surveys) often being prohibitively expensive for local government to undertake, logistically complex, and rapidly out-of-date (due to swift population change, high levels of transient mobility and rapid urbanization). This lack of data at local government level, renders evidence-based decision-making practically impossible, severely impeding progress towards UN SDGs. Dr Goulding and his team’s solution has been to develop a range of novel methods and systems in consumer and citizen analytics, which link previously unattainable datasets, held by private-sector companies, to pressing transport, social and health policy contexts. This research has resulted in the development of a range of techniques that have generated unique insights into mobility, poverty, socio-demographics and vulnerability. Working with a range of NGOs and international development organizations, these have seen direct application in large-scale SDG initiatives in East Africa over the past 5 years.</p> <p>In 2014, Goulding and Smith, part of N/LAB, a multidisciplinary centre for consumer analytics at University of Nottingham, demonstrated that the physical and social mobility of local populations could be estimated via cellular phone customer data (Call Detail Records or “CDR”). Supported by EPSRC and TiGo (a leading Tanzanian Telco), they achieved this through development of new statistical methods in the prediction and understanding of human mobility [1]; and novel techniques able to model consumer “event series” at a micro level using stochastic processes [2]. Much “big” data collected by companies arrives in this form - a series of temporal logs reflecting, for example, purchase events, phone calls, mobile money exchanges, etc., often accompanied by a spatial dimension. The team’s research enhanced methods that deal with ‘noisy’, partial and non-stationary (volatile) data of this nature - characteristics which are highly prevalent in mass real-world datasets of consumer activity.</p> <p>In 2016, the team’s research advances were adopted by several influential organizations in Tanzania. The World Bank (WB) had recognized that previous efforts to address economic development were seriously impeded by (i) poor transport infrastructures and (ii) a debilitating lack of data on citizen mobility and passenger needs. In partnership with WB, this led to the UoN team developing multi-disciplinary and statistical research to develop new best-practices</p>		

in the generation of “Origin-Destination Matrices” [A]. These OD Matrices are fine-grain mobility models, predicting movement patterns across a whole city, reflecting real behaviours embedded in CDR data (revealing fine-grained patterns of mobility, money transfer and call/usage intensity). These advances in analytics research were able to bridge the “data-gap”, and became an “integral” component of the “Dar es Salaam Urban Transport Improvement Project”, a USD 425,000,000 transport investment by the World Bank in Tanzania.

Between 2016 and 2019, a series of multi-disciplinary papers followed, each adding capacity for new consumer, citizen and SDG insights including: poverty estimates, overcrowding, land-use, infrastructure challenges and health risks [3-7]. Supported by the Gates foundation, N/LAB significantly improved the capabilities of AI methods to estimate factors such as poverty at neighbourhood levels, in particular by demonstrating the advances that could be made in harnessing mobile banking data [5]. Further methodological improvements led to the establishment of a visualization platform, exposing a 1km² informational grid of socio-demographic indicators in Tanzania [Figure 1]. This generated interest from the Tanzanian government and, combined with support from Department for International Development (DFID, now FCDO), led the team to develop research-based tools that integrated earth observation imagery with big consumer datasets, extend insights to land-use and infrastructure analysis (SDG9) [4, B, Figure 2]. Most recently, research advances in “variable importance techniques” [7] (methods able to analyse the complex factors which drive outputs of AI models), extended consumer analytics to a health domain. Supported by D-tree international, Zanzibar Ministry of Health (MoH) and EPSRC [G5], the team’s latest work has provided new methods and first quantitative evidence of the key role socio-demographics, living conditions and access to health facilities has in vulnerability to perinatal mortality [7], validated against 40,000 records collected by the MoH’s Community Health Visitors (CHVs).

3. References to the research

1. **Smith, G.**, Weiser, R., **Goulding, J.**, ‘A refined limit on the predictability of human mobility’, *2014 IEEE Pervasive Computing (PERCOM)*, 2014, pp. 88-94, doi.org/10.1109/PerCom.2014.6813948
2. **Goulding, J.**, Preston, S. and **Smith, G.**, ‘Event Series Prediction via Non-Homogeneous Poisson Process Modelling’, *2016 Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2016, pp. 161-170, doi.org/10.1109/ICDM.2016.0027
3. Iliffe, M., **Smith, A.**, **Goulding, J.**, Mobile Money - Towards Understanding Spending Patterns in Emerging Economies, American Marketing Association Conference, 2016. <https://www.ama.org/wp-content/uploads/2019/02/2016-ama-winter-proceedings.pdf>
4. Torres, M., **Goulding, J.**, Valstar, M., and Perrat, B., Automatic Pixel-Level Land-use Prediction Using Deep Convolutional Neural Networks, GISRUK, 2017 https://huckg.is/gisruk2017/GISRUK_2017_paper_115.pdf
5. Engelmann, G., **Smith, G.**, **Goulding, J.**, ‘The Unbanked and Poverty: Predicting area-level socio-economic vulnerability from M-Money transactions’, *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 1357-1366. doi.org/10.1109/BigData.2018.8622268
6. Brindley, P., **Goulding, J.** & Wilson, M. L., Generating vague neighbourhoods through data mining of passive web data, *International Journal of Geographical Information Science*, 2017, 32:3, 498-523, doi.org/10.1080/13658816.2017.1400549
7. **Smith, G.**, Mansilla, R., **Goulding, J.**, ‘Model Class Reliance for random forests’, *34th Conference on Neural Information Processing Systems*, 2020, H-index: 198, <https://papers.nips.cc/paper/2020/file/fd512441a1a791770a6fa573d688bff5-Paper.pdf>

White Papers:

- A. **Goulding, J.**, **Smith, G.**, Perrat, B., Iliffe, M., Engelmann, G. (2017). *Best Practices and Methodology for OD Matrix Creation from CDR data*. UoN and The World Bank.
- B. **Goulding, J.**, Perrat, B., **Smith, G.** (2020). *Remote Sensing and Applied Machine Learning for the assessment of Low Volume Road Condition (ZROADS)*. UoN and DfID.

Grant Details

Funding body	Investigators	Title	Dates	Amount
--------------	---------------	-------	-------	--------

G1. EPSRC	Smith., A. Goulding, J. Meng, X, et al.	Neo-demographics: Opening Developing World Markets by Using Personal Data and Collaboration	2014-2016	GBP612,744
G2. ESRC	Goulding, J. Ilfiffe. M, Smith, A.	Big Data for Flood Resilience in East Africa	2015	GBP9,972
G3. EPSRC	McAuley, D., Goulding, J. Smith, A. et al.	From Human Data to Personal Experience	2015-2019	GBP4,062,954
G4. Bill & Melinda Gates Foundation	Goulding, J., Smith., A., Smith., G.,	Financial Data Mapping in East Africa	2017-2018	GBP70,533
G5. EPSRC	Goulding, J. Dryden, I., et al.	Risk prediction for Women's Health and Rights in Tanzania: novel statistical methodology to target effective interventions	2019	GBP 553,446
Approx. TOTAL:				GBP 5,300,000

4. Details of the impact

The impact of Goulding and the N/LAB team’s research is reflected in the (i) range of real-world development projects which have applied its outputs; (ii) continued investment in its new methods for deriving socio-demographic indicators from private-sector consumer data; and (iii) the contributions to local government systems and best practices that it has made.

In 2016, N/LABs initial research on mobility prediction [1, 2] came to the attention of the World Bank, who approached the team seeking help to inform their USD425,000,000 investment in the “Dar es Salaam Urban Transport Improvement Program” (SDG11). Correctly understanding the movement patterns and transport needs in the city was crucial. However, with 70% of the city’s 6m inhabitants living in unmapped slums/informal settlements, combined with poor infrastructure and severe congestion, traditional road-side surveys were unviable - logistically complex and prohibitively expensive. Research [A] addressed this gap, demonstrating how complex consumer-datasets (CDR) could be transformed into the “Origin-Destination matrix” for the city, for the first-ever time. The resulting insights into traffic, movement and commuting patterns became ‘an integral and substantive component’ of the World Bank’s Dar es Salaam transport masterplan [C]. The World Bank’s Lead Transport Economist also confirms that they consequently used results from this research ‘in direct support of the large-scale investment in a Bus Rapid Transit (BRT) system in the region, which has made real, long-term impact to both the city and its local population’ [C].

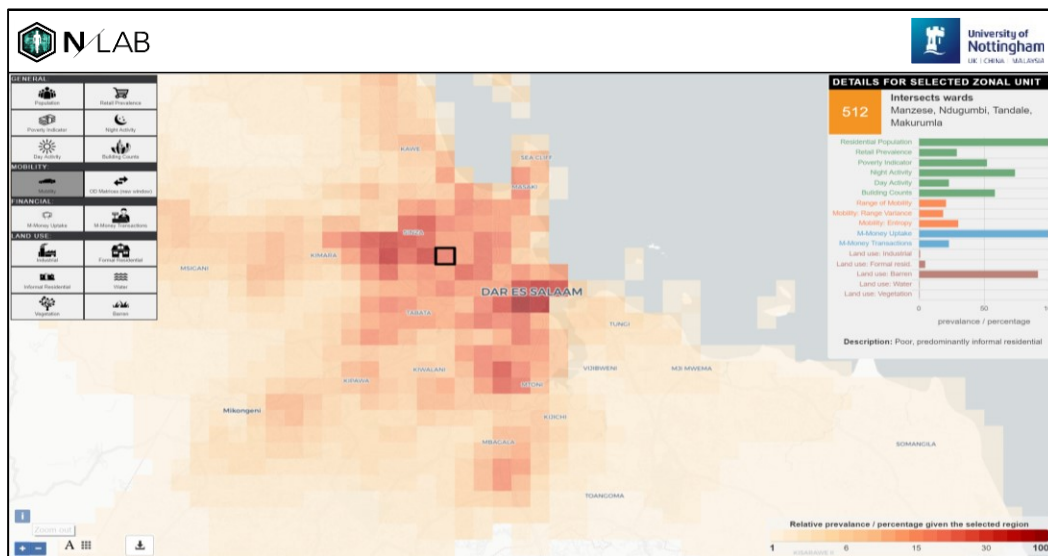


Figure 1: Insight Tiles Visualization, showing mobility of neighbourhoods in Dar es Salaam

Becoming fully operational in 2018, the BRT now has over 300 buses carrying a projected 400,000 passengers each day [K, I]. Previously, public transport was provided by 5,200 privately-owned "dala-dala" mini-buses, contributing to extreme traffic congestion and an estimated 416,451 tons of CO₂ emissions generated by the city yearly [H, L]. The team's technical developments in analysing consumer/user data, and methods for transforming results into user-friendly data products, has contributed to an average reduction of travel times of 10 minutes per commute (and a 30-minute average reduction for journeys to the city centre) [H]. An independent study, published in 2017, reported passenger waiting times having reduced by more than 50%, delays in journeys have reduced 60%, and savings of fare cost to passengers of 28% [J]. While the BRT has targeted improved labour market mobility and inter-firm trade (SDG8), the environmental benefits are also a key output (SDG13), with a UN Environment report estimating the reduction of Green House Gas (GHG) emissions to be 'in the order of 70,000 tonnes annually' [K]. This led to a 2018 ITDP (Institute for Transportation and Development Policy) sustainable transport award, which stated that the project had 'taken a big step in improving the quality of its citizen's lives' [I], with the daily users not only spending less time travelling, but being 'more satisfied and feeling safer' (SDG11.2) [H].

Following this study on transport, the team expanded methods so as to capture indicators for poverty and overcrowding (SDG1); land-use (SDG9); financial activity (SDG8); informal neighbourhoods; social vulnerability [3-7]. In 2018 the research was presented to the Tanzanian Minister of State, leading to a collaboration with the Department of Roads, Zanzibar (ZDoR). Road conditions in Zanzibar are dangerously poor, which impacts SDGs across the economy, tourism and health services alike, and manual surveying of the region's unpaved roads are precarious, time consuming and infrequent. Therefore, N/LAB, supported by DfID, developed first-ever AI models capable of analysing road conditions from drone imagery with ZDoR [B]. These advances were combined with the socio-demographic insights afforded by the team's research on consumer data, resulting in the 'introduction of a new software platform into the Zanzibar Department of Roads to help monitoring and assessment of surveyed roads, linking drone imagery with demographics and go-pro camera footage' [E] (see Figure 2).

According to ZDoR, this has transformed their 'road quality reporting and maintenance planning in a way that would otherwise be unviable' [D]. N/LABs research has improved this government agency's 'day-to-day operations, delivering at no cost an ongoing increase in capacity for information acquisition and utilization within the process of road quality reporting and maintenance planning' [D]. FCDO (previously DfID) state that the work 'has both changed best practice and provided efficiency gains at this government department' and supported 'key resource allocation decisions in a way that was not previously possible' [E].

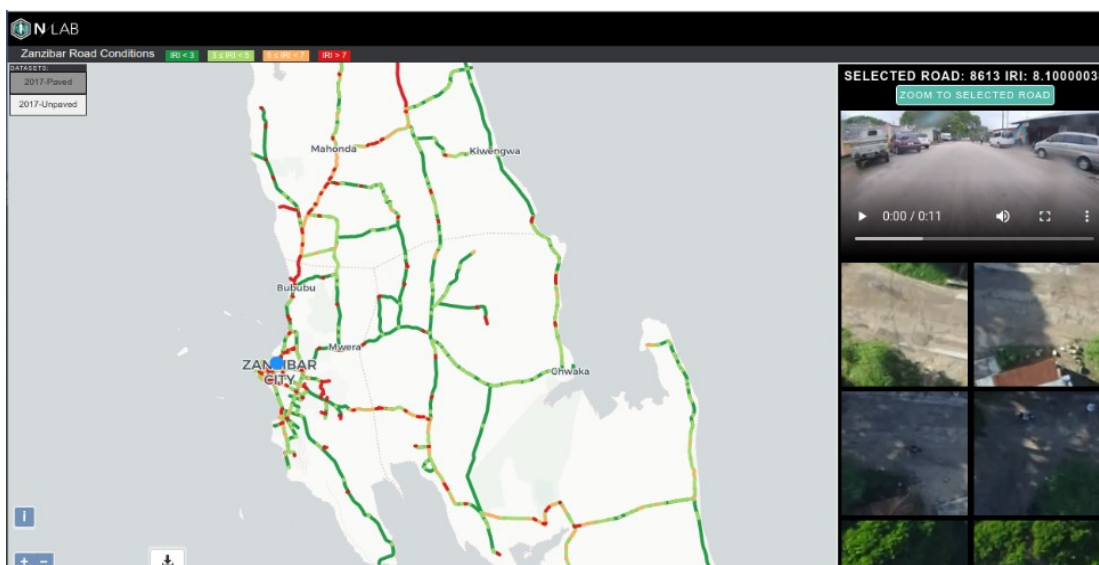


Figure 2: Software interface now in daily used by Department of Roads for resource allocation, here illustrating detected road conditions in each tile, along with linked drone and go-pro footage

Following this influence on the department's practices, N/LAB worked with Zanzibar's Ministry of Health (MoH), and their implementation partner D-tree International (a digital health NGO),

to shed light on local population health and, in particular, the serious problem of perinatal mortality which, despite significant efforts, remains at 4% in the country. D-tree and MoH had previously established their “Safer Deliveries Program” in 2015, employing ‘over 400 Community Health Visitors (CHVs)’ who meet expectant mothers across Zanzibar but, according to D-tree, ‘quality-of-data challenges still impede numerous health outcomes’ and, if this was solved, ‘better targeted interventions/education during and after pregnancy’ could avert 75% of infant deaths [F].

To this end, in 2018, Goulding and Smith worked with D-tree to produce first-of-their-kind automated “risk models” for perinatal mortality, linking the socio-demographic features (e.g. age, living conditions, poverty, social isolation, access to health facilities) exposed by their consumer- analytics, with clinical data from 40,000 historical births [7]. The research, which augmented established variables (age, prior birth, no. abortions) with unexpectedly powerful socio-demographic indicators (relating to living conditions, affluence and isolation), ‘helped underpin development of predictive models, used directly by CHVs in their data collection apps’ which contribute ‘directly to CHV performance’ and ‘10% uplift in risk assessment accuracy’ [F]. Furthermore, D-tree state that N/LAB’s research has moved ‘analysis from research to practice, with the approach now being used within the Ministry of Health via their National Community Health Program’ [F]. The expansion of this program continues, with D-tree stating ‘the research by University of Nottingham has not only already influenced best-practice in assessment of health statuses but has advanced on-the-ground approach, helping CHVs to plan resource allocation, visitation schedules and reduce perinatal mortality occurrence’ [F].

These examples demonstrate the impact of N/LAB’s research between 2016 and 2019, both in terms of measurable contribution to SDG goals and in influencing best practices at parastatal organizations and government agencies in East Africa. However, the application of transformative consumer research continues and, via EPSRC support [G5], in 2020 work extended to support vulnerability modelling for girls at risk of Female Genital Mutilation in north Tanzania. The founder of the Anti-FGM NGO “Hope for Girls and Women”, has emphasised the value already generated by this citizen driven work: ‘With drone images and cell-tower data, we now have maps to help rescue girls in the Mara region of Tanzania. The work of [Dr Goulding] and the team has changed the way data is collected [in] 87 villages. We can now see risk at villages, even isolated ones, for the first time. This is giving us new ways to [...] decide which villages to target education with and how to focus our work for the most effect’ [G]. “Hope for Girls and Women” goes on to say that this programme, generating otherwise unattainable socio-demographic indicators via consumer data, is “already having a real effect, making changes to our country and putting pressure on local leaders for finding long-term solutions” [G].

5. Sources to corroborate the impact

- A. Best Practices and Methodology for OD Matrix Creation from CDR data [White Paper]
- B. Remote Sensing and Applied Machine Learning for the assessment of Low Volume Road Condition (ZROADS) [White Paper]
- C. Testimonial letter from World Bank
- D. Testimonial letter from Zanzibar Department of Roads
- E. Testimonial Letter from FCDO (Formerly DfID)
- F. Testimonial letter from D-Tree International
- G. Testimonial letter from Hope for Girls and Women in Tanzania
- H. Morten, M., Evaluating impacts of the Dar es Salaam Bus Rapid Transit System, 2020. [pp. 1, 14, 25]
- I. The Institute for Transportation and Development Policy, Sustainable Transport Award, <https://staward.org/winners/2018-dar-es-salaam-tanzania/>
- J. Chengula, D.H. and Kombe, K., 2017. Assessment of the Effectiveness of Dar Es Salaam Bus Rapid Transit (DBRT) System in Tanzania. International Journal of Sciences: Basic and Applied Research, 36(8), pp.10-30. [p. 10]
- K. Wong, R., 2018. Terminal Evaluation of the UN Environment-GEF Project. [pp. 10, 60]
- L. Ahferom, M.T. and Svensson, M.G., Sustainability assessment of a bus rapid transit (BRT) system: The case of Dar es Salaam, Tanzania, 2009. [p. 23]