

Institution: University of Sheffield		
Unit of Assessment: B-11 Computer Science and Informatics		
Title of case study: Societal and commercial benefits of large-scale data collection and analysis		
Period when the underpinning research was undertaken: 2008–2020		
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:
Ciravegna, F.	Professor of Pervasive Technologies	2000–present
Lanfranchi, V.	Senior Lecturer	2004–present
Kruger, J.	Postdoctoral Researcher	2009–present
Ireson, N.	Postdoctoral Researcher	2004–present
Epton, T.	Postdoctoral Researcher	2009–2013
Period when the claimed impact occurred: 2014-2020		
Is this case study continued from a case study submitted in 2014? N		
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>The Organisations, Information and Knowledge (OAK) group at the University of Sheffield have developed novel technologies to capture and analyse massive volumes of data from mobile phones, sensors, and social media. These technologies have enabled the creation of a successful sports media company with a turnover of £1.8m; underpinned the development of an app to support a national Public Health England (PHE) campaign (800,000 downloads); informed the creation of cycle paths and a Clean Air Zone in Birmingham; underpinned disaster response strategies in Italy (involving 27,000 citizens); and facilitated the development of novel traffic management products that have transformed the field and improved lived experience in busy urban areas of the UK.</p>		
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>The OAK group, led by Professor Fabio Ciravegna, focuses on the development of novel algorithms and infrastructures to address challenges underpinning the acquisition, fusion and modelling of data at very large scales, and their application to generate impact. The underpinning research was developed by the OAK group through leadership and participation in a number of European and EPSRC projects. Ciravegna led WeSenseIt (2012-2016), SETA (2016-2019), LODIE (2012-2015), and RAnDMS (2012-13), and the OAK group was a partner in WeKnowIt (2008-2011) and ReDites (2013-14). The research spans five themes:</p> <p>Theoretical foundation for digital health interventions. In 2013, interdisciplinary work conducted by Ciravegna in collaboration with psychologists at Sheffield and Sussex developed a clear theoretical foundation for digital interventions in the health domain. This was achieved via some of the earliest work for assessing behaviour change techniques (self-affirmation, theory-based messages, and implementation intentions) to target health behaviours in a smartphone intervention [R1]. Subsequent studies extended this methodology to a large-scale PHE digital intervention scheme (the behaviour change mobile application Active 10 with hundreds of thousands of users), in which physical activity was directly monitored using smartphone sensors</p>		

[R2]. The same research also underpins the SETA technology which allows detailed and granular information on user mobility to be collected.

Robust analysis of mobility at a massive scale. To enable participant monitoring over long periods of time and at very large scales, fundamental research was performed to address problems related to the power efficiency of 24/7 mobile tracking algorithms and their robustness across a diverse range of contexts, situations and mobile device types **[R2]**. Algorithms able to continuously and precisely model mobility with limited impact on a phone's battery life were developed for multifarious applications spanning from health and wellbeing to emergency response **[R3]**. These innovations allowed for the efficient and reliable collection of over one billion data points between March 2017 and January 2019.

Computing context. The mobility of citizens cannot be fully understood without also considering the wider context of their sensor data. For example, it is necessary to combine information from current and historical motion data, GPS, and transport network maps with environmental (such as temperature, weather, and road morphology) and external (e.g. the use of walking aids) factors to determine whether a citizen is running, driving, or using public transport. The OAK group developed novel algorithms for determining the context of motion data using sensor fusion and graph-based approaches **[R3]**.

Methodologies for citizen engagement. Working with social scientists at IHE Delft in the Netherlands, the OAK group developed a framework to identify drivers for increased citizen participation in environmental monitoring, and carried out empirical research in the UK, the Netherlands, and Italy between May and November 2013. This led to important insights into the interactions between individual motivations for participation, the objectives and needs of authorities/institutions, and governance structures **[R4]**.

Novel methods for social media analysis. Methods for largely unsupervised, large-scale mining of data from the web were developed, including approaches for populating knowledge bases that addressed well-known problems such as data sparsity and noise (which can be caused by web publishers using inconsistent vocabularies or making errors). These approaches were shown to increase the precision of relation extraction in a number of internationally available corpora and real-world data sets (F1 score increased between 7% and 40% over baseline models) **[R5]**. This research supported the development of novel methods for social media analysis via natural language processing by allowing the context of a micropost (e.g. a Twitter message) to be inferred from a knowledge base. A particular focus was the early detection of rumours on social media, which led to the development of graph-based techniques for rumour detection **[R6]**.

3. References to the research (indicative maximum of six references)

University of Sheffield staff and students in **bold**

- R1. Epton, T., Norman, P.,** Sheeran, P., Harris, P. R., **Webb, T. L., Ciravegna, F.,** Brennan, A., **Meier, P.,** Julious, S. A., Naughton, D., Petroczi, A., Dadzie, A.-S., & **Kruger, J.** (2013). A theory-based online health behavior intervention for new university students: study protocol. *BMC Public Health*, 13(1), 107. <https://doi.org/10.1186/1471-2458-13-107>. Cited by 45.
- R2. Ciravegna, F.,** Gao, J., **Ireson, N.,** Copeland, R., Walsh, J. & **Lanfranchi, V.** (2019). Active 10: Brisk walking to support regular physical activity. *Proceedings of 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*

Impact case study (REF3)

(*Pervasive Health2019*), Trento, Italy, 20th-23rd May 2019, 11-20.

<http://eprints.whiterose.ac.uk/145438/>. Cited by 2.

- R3.** Ciravegna, F., Gao, J., Ingram, C., Ireson, N., Lanfranchi, V. & Simanjuntak, H. (2018). Mapping mobility to support crisis management. *Proceedings of the 15th Annual Conference for Information Systems for Crisis Response and Management (ISCRAM)*, Rochester NY, USA, 305-316. <http://eprints.whiterose.ac.uk/130788/>. Cited by 2.
- R4.** Wehn, U., Rusca, M., Evers, J., & Lanfranchi, V. (2015). Participation in flood risk management and the potential of citizen observatories: A governance analysis. *Environmental Science & Policy*, 48, 225–236. <https://doi.org/10.1016/j.envsci.2014.12.017>. Cited by 155.
- R5.** Zhang, Z., Gentile, A. L., Blomqvist, E., Augenstein, I., & Ciravegna, F. (2016). An unsupervised data-driven method to discover equivalent relations in large Linked Datasets. *Semantic Web*, 8(2), 197–223. <https://doi.org/10.3233/sw-150193>. Cited by 11.
- R6.** Varga, A., Basave, A. E. C., Rowe, M., Ciravegna, F., & He, Y. (2014). Linked knowledge sources for topic classification of microposts: A semantic graph-based approach. *Journal of Web Semantics*, 26, 36–57. <https://doi.org/10.1016/j.websem.2014.04.001>. Cited by 37.

4. Details of the impact (indicative maximum 750 words)

The collection and analysis of large-scale data has a wealth of applications. This is illustrated by the diverse fields in which the OAK group's research has delivered impact: health, transport, disaster response, and sports media.

National societal impact

In the UK, Sheffield's research on tracking technology, application efficiency, and server infrastructure [R1-R2] were utilised in the "Move More" initiative of PHE's £3m "One You" campaign. Between 2017 and 2019, the Active 10 app achieved approximately 800,000 downloads. PHE confirms, "*With the support of Sheffield's research, PHE were able to develop and launch the first free-to-use mobile app that provided the user with information on time, intensity and periodicity [of physical activity]. The app played a significant role [...] and made a major contribution to the overall success of the One You campaign*" [S1].

In Italy, the OAK group used technologies developed during the WeSenseIt project [R3-R6] to analyse data from mobile phones, sensors, and social media, managing the successful evacuation of 27,000 citizens from Vicenza in April 2014 following the discovery of a WWII bomb [S2, S3]. This led the Italian government to adopt citizen observatories as support for water and flood risk management (European Flood Directive 2000/60) [S2]. The Special Projects Manager at the Eastern Alps District River Authority states: "*The contribution of the OAK group in this process was key. The WeSenseIt project made the concept real and applicable; the technology developed by OAK provided concrete proof of the power of the citizen observatories as well as a powerful benchmark for requirement analysis and for the development of the final production technology*" [S2].

Impact on mobility and local authorities

As part of the "Big Birmingham Bikes" scheme (ongoing from 2015), Birmingham City Council (BCC) gave a large number of bikes (2,000 initially) to communities historically unlikely to take up cycling. The project initially used a dedicated GPS tracker, but as the number of bikes on offer rose, the cost of such trackers to gather evidence in support of the scheme became

prohibitive. By offering free tracking on recipients' mobile phones from 2017-2019, the Sheffield-designed SETA tracking technology [R1-R2] allowed for significant savings (£130k per year) while collecting even more detailed and granular information on overall mobility. Data showing how cycle use correlated with the routes chosen by riders, combined with vehicle tracking information provided by a SETA partner, caused BCC to revise the proposed locations of new cycle routes to better meet the needs of cyclists and reduce safety risks [S4]. After the Wellbeing Service at BCC became the Active Wellbeing Society (an independent charity), it bid successfully to become one of Sport England's 12 local delivery pilots, being awarded a total of £11m between 2018 and 2019 (of which £420k was devoted to free bicycles). *"The experience, evidence and skills we gained by using the SETA app made a significant contribution to the success of our proposal, allowing us to put ourselves forward confidently as a cutting-edge organisation with a key strength in the use of data gathering and evidence. [...] Our core objective has always been to increase cycling and increase funding for cycling – and Fabio's SETA app has helped us achieve this"* [S5]. To supplement the prior cycling and vehicle data, Sheffield provided tracking apps for thousands of cyclists, walkers, and runners within Birmingham, giving BCC a detailed understanding of mobility within their city, upon which they based the design and operation of their Clean Air Zone [R3].

The Floow Ltd, a company that collects second-by-second data on vehicles in motion to provide risk analysis services to the motor insurance industry, used their participation in the SETA project to develop six new data products for traffic management within the REF period. According to their Chief Innovation Officer: *"The University of Sheffield was the driving force behind the SETA project [...] Sheffield's expertise in active travel and complex visualisation systems [...] helped The Floow understand the needs of traffic managers and the potential of products that might help support them"* [S4]. These products have enabled local authorities to more effectively manage transport for the benefit of their residents. The Floow's contract with Greenwich has transformed the council's priorities for traffic management by overturning previously false assumptions with precision data, e.g., it was revealed that the proportion of traffic from outside the region was 75% and not 20% [S4]. This approach has been expanded and repeated, providing new insights to make streets safer and less polluting across London and the UK [S4].

Commercial impact

As a result of these new data products, **The Floow** has been able to win commercial contracts in new sectors. Working with Sheffield and the other SETA partners *"ensured that emerging new to market products better met wider needs of end users including a broader overview of 'mobility'. [...] These gains were fundamental to our success in these areas and are thanks to SETA"*. The new products have generated agreements in excess of £400k, over the financial year 2019-2020, and led to contracts in new markets. These markets include governmental, with supply to the Department for Transport, highway agencies and local authorities across the UK; autonomous, with new aspects of support helping to set routes for robo-taxi deployments; and traffic management, with supply of traffic data to large consultancy organisations and monitoring teams [S4].

Football Whispers, a start-up aiming to create an online platform for football fans and sports journalists to analyse and discuss the latest football news and transfer rumours, approached Ciravegna in October 2015 for help with developing the technology needed for their launch in January 2016. The OAK group's social media-based technology [R6] was developed for the prediction of football transfers by analysing messages in social media across multiple languages and 36 international leagues [S6]. According to the Founder and former CEO of Football Whispers: *"Thanks to the work of the OAK group at the University, we were able to launch on*

time in January 2016 and with our full service offering – something that we would not have been able to accomplish without their input. [...] In that time our business grew from 0 to 2,500,000 unique monthly users. [...] It was adopted as a tool by Sky Sports, ESPN (the world's largest sports television company) and football magazine FourFourTwo, and was also used by the sports pages of The Sun and The Daily Mail. During the period of my Chairmanship (2015-2017) Football Whispers generated substantial income from advertising, and from the separate professional services we offered to media organisations, football team press officers and sports journalists, alongside their service to fans" [S7, S8].

The company expanded into other sports and now operates as All Sports Whispers to reflect its involvement in the NFL, NBA, and other organisations. All Sports Whispers employs 35 people and has generated over £1.8m in revenue since its founding in 2017 [S9].

5. Sources to corroborate the impact (indicative maximum of 10 references)

- S1.** Confidential email from the Product Lead – Marketing at PHE (2020). Corroborates user download data for the Active 10 app and how it has helped the PHE campaign.
- S2.** Confidential testimonial from the Special Projects Manager at the Eastern Alps District River Authority (2020). Corroborates the role of Sheffield's technology in the evacuation of the city of Vicenza and its adoption by the national and regional governments of Italy to meet the requirements of the European Directive 2007/60 and the Water Framework Directive (2000/60) of the Eastern Alps District.
- S3.** The Times article "Italian city is evacuated for British bomb", 26 April 2014 (subscription required). Reports number of citizens evacuated in Vicenza whilst a recently discovered Second World War bomb is defused (paragraph 4). (Accessed 24 Feb 2021). <https://www.thetimes.co.uk/article/italian-city-is-evacuated-for-british-bomb-q2wgxtprqg9>
- S4.** Confidential testimonial letter from the Chief Innovation Officer, Director and Co-founder at The FLOW Ltd (2020). Corroborates how Sheffield's research was used to a) influence the BCC's cycle routes, b) provide a number of UK councils with precise data to allow for the development of improved local traffic management plans and c) develop new to market products for The FLOW. The end users of The FLOW's products and the economic impact of these products are presented.
- S5.** Confidential testimonial letter from the Director of Insight and Knowledge at the Active Wellbeing Society (2020). Corroborates the critical role that Sheffield's research played in achieving the charity's objective and allowing them to become one of Sport England's 12 local delivery pilots.
- S6.** Football Whispers brochure (scan of hard copy only) confirming the importance of the University of Sheffield's involvement in the Football Whispers product (2016).
- S7.** Confidential testimonial letter from the Founder and former CEO of Football Whispers (2020). Corroborates the critical role Sheffield's research in Football Whispers product launch.
- S8.** Various media pages reporting on Football Whispers:
- Sky Sports (2016) (Accessed 7th Aug 2020). <http://bit.ly/3exKl3C>
 - Football Whispers award nomination (2017) (Accessed 2nd Oct 2020). <http://bit.ly/3qFGlWo>
 - ESPN (2018) (Accessed 2nd Oct 2020). <https://es.pn/3bH898s>
 - The Sun (2018) (Accessed 2nd Oct 2020). <http://bit.ly/3tiPu9j>
- S9.** Company accounts filed at Companies House (2017-2019). Corroborates employment and financial figures for All Sports Whispers. (Accessed 29th July 2020). <http://bit.ly/2Na3oLh>