Project on Social Impacts of Rio2016

Evidencing the Impacts of the Olympic Games - the view from London 2012

Prof Allan Brimicombe, a.j.brimicombe@uel.ac.uk
Dr Yang Li y.li@uel.ac.uk

Centre for Geo-Information Studies, University of East London, UK

Abstract

This chapter is written in the context of the London 2012 Olympic and Paralympic Games. The authors have led four major projects focused on measuring legacy. The Olympic Games are a class of planned mega-event, that is, an event of sufficient size capable of producing rapid or accelerated transformations in the economic, environmental, social and cultural spheres of a city. The Olympic Games Impact (OGI) studies were designed to meet the desire of the International Olympic Committee (IOC) to develop an objective and scientific analysis of the impact of each edition of the Games and are intended to provide a record of both the individual nature of each Olympiad and its host context. London 2012 were the first Summer Games for which this was mandated through the host city contract. But there have been a considerable number of other enquiries, reviews, 'meta-evaluations' and 'supra-evaluations' as well. Much of this is focused on determining the legacy which is defined here simply as: any net impact arising from the Games. The term 'impact' refers to any change or transformation, positive or negative, and which can be attributed to the Games. This implies that both linkage and counterfactual are known. For the most part this has relied on conventional statistical approaches which require all tangible and intangible benefits to be monetised. However, this is likely to overlook key elements of social value which reside in lifestyles, beliefs and culture. For this we may need to turn to Big Data approaches of, say, analysing social media in order to evidence informal impacts to do with individual enjoyment and inspiration derived from a mega-event.

Introduction

On completion of the Olympic Games impact studies (OGI) in December 2015, we held a one-day conference to make the event on the theme of "Debating Legacy". That evening, the BBC London news featured the conference as lead the story[1]. The transcript is as follows:

BBC news anchor: "*The legacy of the 2012 Olympic Games has exceeded all expectations according to a final report commissioned by the International Olympic Committee. Dozens of indicators were used to assess the impact of the London Games although as Sarah Harris reports some east Londoners are not convinced*".

BBC reporter Sarah Harris: "*The physical legacy of the 2012 Games can be seen for miles across the east London skyline, but the International Olympic Committee asked for a more rigorous, scientific analysis of the impact. Today this report was published at the University of East London*".

[1] BBC London News, 18:30 Saturday 12 December 2015

Professor Allan Brimicombe [UEL]: "*We've had to look at 67 indicators over the period 2003 to 2015. What we have now in east London is a 21st Century neighbourhood developing with new housing for east London, park facilities, sports facilities, opportunities for business and employment*".

BBC reporter Sarah Harris: "*But on the streets of Stratford, not everyone's convinced. Crime and poverty levels may have gone down, but some are still campaigning for better local services and are bitter that for them the Games did not deliver*".

Street interviewee 1: "*The thing about the Olympics...it was all always about elite sport. It was never about ordinary people participating in sport. They kind of pretended it was and things like that. I mean, it just goes shows right across the board, people in east London got nothing out of the Olympics*".

Street interviewee 2: "*…almost find us, but yet still the people down at the bottom are suffering a lot*".

Street interviewee 3: "*Local people are actually forced away. Local people are not even local people anymore - they are in Birmingham, Manchester, Hastings. You've got loads of new people, and the people that have been here for a long time, they're being forced away, so...*"

BBC reporter Sarah Harris: "*As far as the Olympic Committee's assessment goes, the 2012 Games will be a difficult one to beat and will leave a legacy in east London to be enjoyed for generations to come*".

This brings into stark relief the results of a more scientific assessment of London 2012 Games legacy juxtaposed against the opinion of the legacy on the street. The impact of the Games and its legacy is thus a debate which includes perceptions and not just a matter of statistics. In this chapter we explore this juxtaposition and consider some novel, Big Data techniques that might be deployed to enrichen that debate.

Defining Legacy

In the context of an Olympic Games (or any organised mega-event), *legacy* is defined here simply as: ***any net impact arising from the event***. The term 'impact' refers to any change or transformation, for better or for worse, that has taken place and which can be attributed to the event. This implies that the causal linkages, direct or indirect, to the event need to be understood. However, the key term in the definition is '***net***', that is, the impact that has occurred over and above what would have happened without the event – what is known as the *counterfactual*. Hosting something like the Olympic Games is rarely context-free or designed on a tabula rasa; rather it is superimposed on existing trajectories of historical development. Indeed, showing how planned legacy fits with city aspirations is now an important consideration in any bid to host the Games. Establishing a plausible counterfactual to measure net impact is therefore critical to knowing what the true legacy of an Olympic Games is.

For London 2012, six legacy promises[2,3] were put forward as a focus of aspirational transformation. These were:

1. To make the UK a world-class sports nation: elite success, mass participation and school sport.
2. To transform the heart of East London.

[2] DCMS (2008) *Before, during and after* London: DCMS; with the addition of the sixth promise in December 2009

[3] The Mayor of London paraphrased the first five promises as:

- Increase opportunities for Londoners to become involved in sport.
- Ensure Londoners benefit from new jobs, business and volunteering opportunities.
- Transform the heart of east London.
- Deliver a sustainable Games.
- Showcase London as a diverse, creative and welcoming city.

www.london.gov.uk/priorities/london-2012/benefits-and-legacy

3. To inspire a new generation of young people to take part in local volunteering, cultural and physical activity.
4. To make the Olympic Park a blueprint for sustainable living.
5. To demonstrate that the UK is a creative, inclusive and welcoming place to live in, to visit and for business.
6. To develop the opportunities and choices for disabled people.

Such legacy aspirations are often stated simply so that the public might engage with them and tend to gloss over their actual complexity and multi-dimensionality. Thus in the case of the above promises, exactly what is meant by "to transform the heart of East London" in terms of a measurable legacy? The name 'East London' does not have any standardised (administrative) geographical definition and so where is its 'heart', and furthermore, what is intended to be transformed over what time period? Of course, if such promises are left vague it is easier for politicians and administrators to say this or that aspect was a success (cherry-picking legacy), or if things are not going to plan to say that it is too soon to see an effect. After all, the evidence is that many large public sector projects suffer from both substantial overruns in costs compounded by substantial shortfalls in benefits[4]. But if we are to measure legacy, then these legacy aspirations need to be translated and decomposed into measurable components of outcomes for defined geographical extents and for specified time horizons.

Data evidencing impacts and the evaluation of legacy

Going through the motions of evidencing legacy is often thought of as occurring only after an event is over and the 'show has left town'. Since the very act of preparing a bid, say for an Olympic Games, may have already set in place changes that anticipate some planned legacy (such as the acquisition or agglomeration of land for an Olympic Park), ideally data baselines need to predate the bid preparation. For the London 2012 OGI study the base year is 2003 (two years before being awarded the Games) and extends twelve years to 2015 and thus necessarily includes nine years of data in the run-up to the Games and three years of post-Games legacy data.

Cities rather than countries are contracted to host the Olympic Games and whilst the impact of such a mega-event may be (purposefully) uneven across the city, its impact should hopefully extend well beyond the city itself. Legacies happen in specific places – they are geographically rooted – and therefore location is an important feature in measuring legacy. Nevertheless impact can manifest itself across a number of geographical scales. The IOC OGI studies stipulate three scales of analysis: city, region, and country; and of course these terms are themselves open to interpretation. Thus in the case of London 2012 the 'city' is deemed to be the London Boroughs hosting the Games venues (the Host Boroughs; initially five boroughs but extended to six in 2011), the 'region' is London as a whole and the country is preferably the United Kingdom but could be Great Britain, England & Wales or England depending on availability of compatible data amongst the devolved administrations that currently constitute the United Kingdom.

Most approaches to assessing legacy rely on quantitative indicators to monitor change in aspects of the environment, economy or socio-cultural landscape. A number of issues that arise here. The first is defining the indicator(s) that will adequately reflect change in the phenomena to be monitored. This can be quite hard for soft legacy such as "being inspired" by London 2012. The choice of indicator(s) inevitably needs to be mediated against the availability of data at appropriate temporal and spatial granularity, or, if resources permit, a program of data collection. Inevitably this requires compromise in the choice of indicator(s) which may not themselves directly measure the impact to be monitored but act as good proxy or latent variables, or as a group of variables that triangulate the target impacts. Then there is the issue is maintaining the integrity of a time series over the period necessary to see a legacy effect. As stated above, for the IOC OGI

[4] Flyvbjerg, B. (2016) The fallacy of beneficial ignorance: a test of Hirschman's Hiding Hand. *World Development* 84: 176-189

this is deemed to be 12 years. Published data sets can be discontinued, the method of compilation changed, definitions of variables and the questions asked in official surveys can be revised, and sample sizes (more often than not) reduced. Then there is the lag in official and administrative statistics. It takes time to collate returns, quality assure, ensure disclosure control and finally be approved for publication. This generally means an 18-24 month lag and so data released in 2015 might only be reflecting the situation in 2013.

Approaches to the London 2012 Games

Whilst the President of the International Olympic Committee (IOC) stopped short of pronouncing London 2012 as 'the greatest Games ever', there is little doubt that they were a successful event. But what of the impacts and how they translate into legacy – has the London 2012 Games been living up to its promises? The IOC's official impact report – the Olympic Games Impact (OGI) studies - published in December 2015[5] assesses just that. These studies meet the IOC's desire to develop an objective and scientific analysis of the impact of each edition of the Games and are intended to provide a record of both the individual nature of each Olympiad and its host context. London 2012 was the first Summer Games for which this was mandated through the host city contract. But there have been quite a number of other enquiries[6], reviews[7], 'meta-evaluations'[8] and 'supra-evaluations'[9] as well – over 200 reports and papers on all aspects of (mostly) anticipated and (some) measured legacy of the London 2012 Games. This volume of reports is symptomatic of the government's preoccupation with legacy and a deep-felt need to show that the Games were worthwhile and beneficial (profitable) for the whole country, not just London. This is part of a trend where democratically accountable governments increasingly seek to justify their policies, actions and public sector spending by evidencing their merit, quality and efficacy. Thus formal evaluation of policy interventions and their outcomes have been an integral part of achieving transparency in accountability. The UK Government has a number of documents giving guidance for evaluation, impact assessment and cost-benefit analysis. Key amongst these is:

- *The Green Book* [10]: This provides a framework for the appraisal and evaluation of government policy and projects. The difference between 'appraisal' and 'evaluation' is in the timing. Appraisal is an assessment as to whether a proposed project or policy intervention is worthwhile and most commonly takes the form of a cost-benefit analysis. Evaluation takes place post-implementation to see to what extent the objectives of the project have been achieved and what lessons might be learnt. A supplement to the Green Book has been issued[11] to cover techniques in social cost-benefit analysis. An example of cost-benefit analysis for London 2012 is one carried out at the point of deciding whether or not London should make a bid for the Games[12]. On an estimated expenditure of £1.9bn the income ranged from £1.65bn to £1.98bn, that is, from an 8% loss to a 4% surplus. Much of this variation arose from uncertainty in the tourism income which in many Olympic cost-benefit analyses is viewed as the key wider economic benefit of staging the Games. The decision, as we know, was to go ahead with a successful bid. But by 2005 the estimate for public sector funding had risen to £2.4bn and by 2007 to £9.3bn[13] primarily because the anticipated private sector involvement in the construction did not materialise. Studies of the Sydney

[5] Online version available at http://www.uel.ac.uk/geo-information/

[6] e.g. House of Lords Select Committee on Olympic and Paralympic Legacy (2013) *Keeping the Flame Alive: the Olympic and Paralympic Legacy.*

[7] e.g. Cabinet Office (2015) *Inspired by London 2012: the legacy from the Olympic and Paralympic Games.*

[8] Department for Communities, Media & Sport (2013) *Meta-Evaluation of the Impacts and Legacy of the London 2012 Olympic Games and Paralympic Games – Report 5: Post-Games Evaluation*.

[9] Centre for Sport, Physical Education & Activity Research (2013) *London Legacy Supra-Evaluation.*

[10] Her Majesty's Treasury (2003) *The Green Book: appraisal and evaluation in Central Government*.

[11] Her Majesty's Treasury and Department for Work and Pensions (2011) *Valuation Techniques for Social Cost-Benefit Analysis: Stated Preference, Revealed Preference and Subjective Well-Being Approaches*.

[12] ARUP (2002) *London Olympic 2012 Costs and Benefits*.

[13] Berman, G. (2010) *Financing the London 2012 Olympic Games*. House of Commons Library Standard Note SN3790.

2000[14] Games have shown that many such *ex ante* cost-benefit valuations are over optimistic of the economic stimulus that can be expected from the Olympics.

- *The Orange Book* [15]: This provides a model of risk management. Any Games event has substantial risks (financial, reputational and otherwise). If these risks are identified and monetised at an early stage, then sufficient mitigation and financial contingency can be put in place to cover the risks. In 2007 the National Audit Office (NAO) carried out a risk assessment of the preparation phase for the London 2012 Games[16] and identified six areas of risk that needed to be managed. Prime amongst these was delivery against an immovable deadline; also included was planning for a lasting legacy. By 2010 a contingency of £2.2bn was in place to cover risks. These were of three types: programme contingency (£0.97bn) for the construction of the Olympic Park on a constricted site to a fixed deadline; funders' contingency (£1bn) for changes in scope and wider economic conditions; and a security contingency (£0.24bn). The post-Games audit[17] showed that not all the risks materialised and that an underspend of £0.38bn on the contingency was achieved. The same audit also looked at the immediate legacy benefits which were viewed principally from the point of view of job creation (177,000 job years employment in the construction 2007-2012; 34,500 people in Games-related employment), progress on planned legacy use of the venues, and problems of governance and coordination of legacy delivery. "…. it remains the case that numerous individual organisations are delivering aspects of the legacy and that coordination of this activity remains a challenge".
- *The Magenta Book* [18]: This provides further guidance on programme evaluation and complements the Green Book. The key focus is to identify 'what works', highlight good practice, identify any unintended consequences or unanticipated results, and value for money that can be used to improve future decision-making. "Not evaluating, or evaluating poorly, will mean that policy makers will not be able to provide meaningful evidence in support of any claims they might wish to make about a policy's effectiveness. Any such claims will be effectively unfounded." Meta-evaluation, the synthesis of separate smaller evaluations, is also covered in this volume, and is further discussed below in relation to London 2012.

From the preceding discussion it is clear that measuring legacy can take a number of forms - valuation, evaluation, meta-evaluation, audit - and whilst they tend to be quantitative, can nevertheless be qualitative or a mixture of both. There is also a tendency to include historical analogues, that is, to use previous events as a benchmark or as a comparator to gauge progress or likelihood of legacy emerging. One such study was carried out by researchers at the University of Westminster for the Royal Institution of Charted Surveyors[19]. This study used six case studies of previous events including the Barcelona 1992 and Sydney 2000 Olympic Games, the 1998 FIFA World Cup, and the Manchester 2002 and Melbourne 2006 Commonwealth Games, from which to draw conclusions about the likelihood of a regeneration effect from the London 2012 Games. A scorecard approach was used whereby, having identified from the case studies 15 criteria of good practice, London 2012 was rated and scored 165 out of a possible total of 200, or 82.5%. What becomes clear is that where cities plan for legacy from the moment they start planning for an event, the more likely it is that the legacy will come to fruition. Legacy was embedded in the planning for London 2012.

The IOC naturally has a much broader interest in legacy that just regeneration and has mandated, through its Host City Contracts, for Olympic Games Impact (OGI) studies and, as already discussed above, the baseline is two years prior to being awarded the Games. OGI need to be carried out independently of the local organising committee and so, for London 2012, the main phases of the OGI (pre-Games, Games time, post-Games) were carried out by the University of East London (UEL) funded by the Economic and Social Research Council (ESRC). OGI are based on an IOC Technical Manual which provides the specification for the production

[14] Giesecke, J.A. and Madden, J.R. (2011) Modelling the economic impacts of the Sydney Olympics in retrospect – game over for the bonanza story? *Economic Papers*, 30: 218-232

[15] Her Majesty's Treasury (2004) *The Orange Book: management of risk - principles and concepts*.

[16] National Audit Office (2007) *Preparations for the London 2012 Olympic and Paralympic Games – Risk assessment and management*.

[17] National Audit Office (2012) *The London 2012 Olympic Games and Paralympic Games: post-Games review*.

[18] Her Majesty's Treasury (2011) *The Magenta Book: guidance notes for policy evaluation and analysis*.

[19] Royal Institution of Chartered Surveyors (2011) *The 2012 Games: the regeneration legacy*.

of standardised data on 120 possible indicators in the environmental, social-cultural and economic spheres. Some of these indicators, such as *So9 Health*, are themselves baskets of indicators capturing many dimensions. Not all 120 indicators are expected to be reported on for all Host Cities, but an appropriate selection is negotiated between the IOC and the local organising committee. Thus for the final post-Games London 2012 report, 67 indicators were deemed necessary (15 environmental, 27 socio-cultural, 25 economic) to adequately assess the legacy.

In relation to OGI, no official counterfactual was established. Instead the trend in the data from 2003 onwards are analysed and interpreted as to whether the changes reflect a net impact that can be attributed to the Games. In some cases, such as the transformation of the Olympic Park site or improvements in the transport infrastructure, this is straightforward. In others it has been difficult to disentangle regional and national trends and government policy changes and interventions from the effect that the Games have had *per se*. A case in point would be crime rates which have fallen in East London but then been generally falling nationally since 1997 as part of the 'great crime decline' affecting many western countries[20]. Since London 2012 was touted as the sustainable Games, a method was devised to calculate a sustainability rating using all the indicators where a minus score has introduced unsustainability, zero is the status quo and a score of one achieves full sustainable. London 2012 achieved an overall score of 0.63 (environmental 0.56, socio-cultural 0.68, economic 0.61) which is up from 0.37 in the 2010 Pre-Games OGI report and is because the nature of the impact of the London 2012 Games have become clearer in the years following the Games.

The OGI, as stated, shows change in relation to a baseline, but does not have a coherent counterfactual. To establish such a counterfactual for each and every indicator would be a daunting task. The counterfactual is more easily achieved where there is a single or small group of indicators such as when valuing the contribution to GDP of the London 2012 Games. The meta-evaluation study[21] carried out by DCMS[22] provides one such example. The direct spending on the preparation for the Games by the public sector – the Public Sector Funding Package (PSFP) - totalled just over £8.9bn. This included the land purchase, infrastructure and venue construction. The economic calculation of the *gross* GVA[23] impact of this spending for the period 2007-2012 is £11.5bn. Regarding the counterfactual for calculating the *net* impact, the report states "*The modelling compares the impact of the Olympics with the counterfactual assumption that the Olympics weren't awarded to London and therefore there was no construction or operational spending. There is no counterfactual assumption related to spending the public money on anything else*." This, in the authors' view, is a politically convenient assumption to make because it means the gross benefit is also the net benefit. Some displacement was accounted for (movement of production from other parts of the economy to the Games preparation) giving a net GVA impact of about £10bn and therefore the Games make an economic surplus. However, about £3bn of the PSFP came from sources such as the National Lottery and would have circulated in the economy and contributed to GVA even if London had not won the Games – so the counterfactual cannot be zero. The other £6bn of central government spending, a miniscule amount of total government spending, may well have been spent on other projects in the boom years prior to the recession. A more realistic counterfactual would have been to model the value of PSFP as government consumption (spending on goods and services e.g. more doctors and teachers). The net impact would then more properly reflect the difference between spending on consumption versus spending on infrastructure and thus value the true legacy.

Moving beyond the purely quantitative

Quantitative studies of impact and legacy can appear quite clinical and the public can remain unconvinced. Intangibles such as "being inspired" are difficult to capture convincingly in monetary terms. Also, time moves on, the issues and emphasis change and as we saw above from the BBC news, broader social and

[20] Zimring, F. (2007) *The Great American Crime Decline*. New York: Oxford University Press.

[21] A meta-evaluation is an over-arching synthesis of the findings of individual project-level evaluations in order to provide a comprehensive understanding of outputs and impacts associated with a mega-project.

[22] Department for Communities, Media & Sport (2013) *Meta-Evaluation of the Impacts and Legacy of the London 2012 Olympic Games and Paralympic Games; Report 5: Post-Games Evaluation, Economic Evidence Base*.

[23] Gross Value-Added (GVA) measures the economic contribution of each producer and contributes to the calculation of Gross Domestic Product (GDP); GVA + taxes – subsidies = GDP

economic changes can be conflated with legacy in the minds of the public. In a democracy everyone's opinion counts. So it is desirable to be able to:

- capture shifts in debates, and
- understand impacts from a more socially discursive perspective that employs Big Data approaches that tap into the 'chatter' around events and therefore peoples more frank opinions about an event.

This we now illustrate, first of all with an analysis which mines debates around London 2012 in the House of Lords and therefore amongst senior legislators, and secondly an analysis of Twitter data for the opening and closing ceremonies for the PyeongChang 2018 Winter Olympic Games.

House of Lords Debates

All debates in the Houses of Parliament, both the Commons and the Lords, are recorded verbatim in Hansard[24]. The House of Lords has formally debated the London 2012 Olympics nine times between November 2008 and November 2015. Taking the first and last of these, the objective was to see how the focus of the debates had changed. To assist in this task, machine-based text mining was used. The platform was open source R[25] using the tm library[26]. The 2008 debate contained 21,518 words and the 2015 debate contained 15,469 words. A small extract from Hansard is given in Box 1.

Box 1: Small extract from Hansard for the 2015 debate[27].

Olympics 2012: Regeneration Legacy

05 November 2015
Volume 765

Motion to Take Note
2.52 pm
Moved by
Lord Mawson
That this House takes note of progress made in the regeneration of East London since the 2012 Olympic and Paralympic Games and the remaining challenges.
Lord Mawson (CB)
My Lords, a great deal has happened in east London since the summer of 2012, when the world marvelled not only at this country's ability to put on such a successful Olympic and Paralympic Games and to arrange the weather for it, but also its bold promises to create a legacy from the Games in east London second to none. While there have been some challenges, the legacy promises made in east London in terms of regeneration are on track and developing at quite a pace. I declare an interest as a director of the London Legacy Development Corporation and as chairman of the Communities and Regeneration Committee.

A document such as this requires substantial cleaning prior text mining. For example, each participant in the debate is named which needs to be removed as their names are not part of the debate. Also many words such as articles ("a", "the") and conjunctions ("and", "but") detract from the main substance in the nouns and verbs and can also be removed. There are automated techniques for doing this. The result is a cleaned corpus which can be machine analysed. Figure 1 provides a summary of the most frequent terms in each debate as wordclouds where the size of each word reflects its relative frequency (Figures 1(a) and

[24] https://hansard.parliament.uk/
[25] https://www.r-project.org/
[26] http://tm.r-forge.r-project.org/
[27] https://hansard.parliament.uk/Lords/2015-11-05/debates/15110539000344/Olympics2012RegenerationLegacy

1(b)). Furthermore, it is possible to visualise in this way the commonalities between the two debates (Figure 1(c)) as well as the differences (Figure 1(d)).

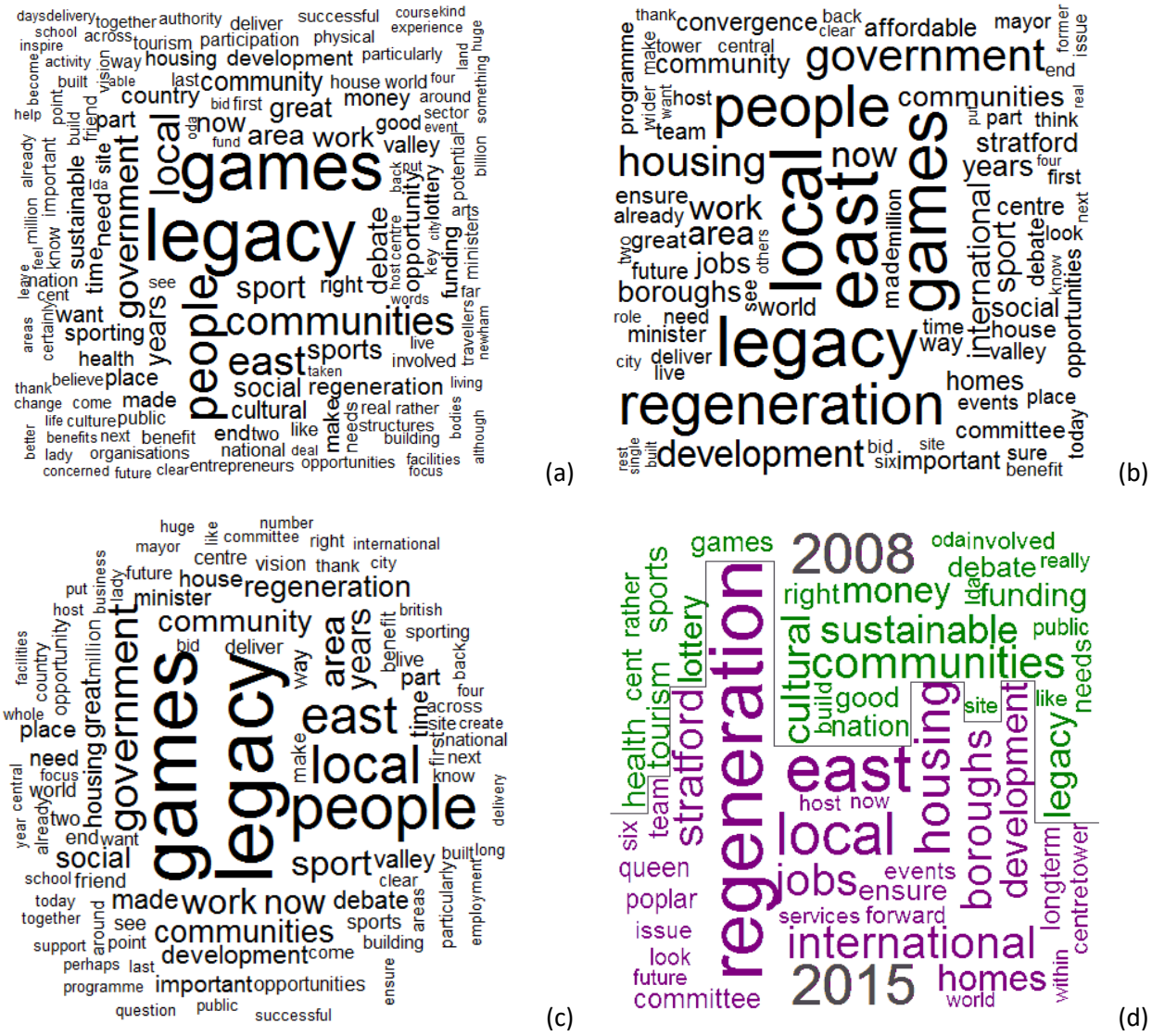


Figure 1: Wordclouds of the two debates in the House of Lords: (a) 2008, (b) 2015, (c) commonalities between the two, (d) differences between the two debates.

Figure 1 represents an encapsulation of the two debates in a way that would be difficult in just listening to the debate or reading the transcripts and becomes by-and-large self-evident. In both debates the words *games* and *legacy* quite rightly loom large given the topic of the debates. There is also a focus *east* and towards *people* and *communities*. Notice though how *sport*, given that this is the Olympics and features in Promises 1 & 4 above, is not so important in the debates. However, it is the differences in emphasis as shown in Figure 1(d) that are perhaps the most revealing. In 2015 the debate was much more focused on *regeneration*, in *east* London *boroughs* and particularly around *Stratford* where the Olympic Park is situated. There is a stronger focus on *local*, *jobs*, *homes* and *housing*, but much less so on *sustainability* and *community*. To a broad extent this reflects the concerns of the people in the street in the BBC news piece above. The debate over legacy has changed its emphasis from 2008 to 2015 and can be evidenced in this way.

Monitoring the Opening and Closing Ceremonies at PyeongChang 2018

On-line social media has a global reach with Facebook having over 1 billion users and some 500 million tweets being transmitted daily. These, together with others such as Instagram, Tumblr, SnapChat and WhatsApp, have been recognised as new and important sources of social data[28]. Given their typical daily volume and heterogeneity of data, they can be characterised as Big Data. The possibility of anonymous, a-synchronous and non-face-to-face contact has allowed freer expression in writing and images of personal opinions, thoughts and behaviours on a scale that was not possible before. This has also lead to new deviant behaviours such a cyberbullying, revenge pornography and sexting which are finding their ways into the statute books as indictable crimes. Cyberbullying, for example, has been found to be contagious[29]. The same characteristics of social media exchanges also allows difficult to discuss topics to be aired, monitored and analysed such as stigmatising attitudes to mental illness[30]. There is therefore considerable potential in using social media to access and analyse people's frankly expressed opinions, say about an event. We have conducted an exploratory exercise in collecting and analysing Tweets for the opening and closing ceremonies of the PyeongChang 2018 Winter Olympics to see what value it might have.

To access publicly available tweets, we currently use the R platform[25] which allows easy access and use of the streamR[31] API to tap into the instantaneous Twitter traffic. Once handshake and authorisation are achieved, requested data are downloaded in JSON or XML formats and parsed to a CSV spreadsheet format for analysis. In accessing tweets for PyeongChang 2018 three separately specified requests were made in a loop so that a whole ceremony could be evenly sampled in real-time. These were: tweets worldwide in English, tweets in the Korean language, tweets about Team GB. We will be focusing here on the first two. Each tweet carries with it additional information in a header such as date and time, owner ID, number of followers and geographical coordinates of message origin. The latter are particularly useful in mapping tweets though inclusion of geographical coordinates needs to be activated by each user and is only present in about 10% tweets, nevertheless given that streamed tweets being captured are high volume and are themselves a sample rather than the totality of tweets useful analyses can be carried out. The number of tweets accessed is given in Table 1. The mapped distribution of tweets in English are in Figure 2 and respective wordclouds in Figure 3.

Table 1: Specification and numbers of tweets collected.

Tweets	Opening ceremony	Closing ceremony
Worldwide in English @pyeongchang2018, #PyeongChang2018, @Olympics, #Olympics, @iocmedia, @bbcolympics, #bbcolympics, @NBCOlympics, #WinterOlympics, @testevents_pc, #HelloPyeongChang, @OIS_PC2018, @AroundTheRings	127,042	183,653
Korean language @pyeongchang2018, #PyeongChang2018, #2018평창 동계올림픽대회 및 동계패럴림픽대회 공식 트위터, @Korea_Olympic, @OlympicKorea	18,764	47,246

[28] Kalampokis, E.; Tambouris, E. & Tarabanis, K. (2013) Understanding the predictive power of social media. *Internet Research* **23**: 544-559
Lansley, G & Longley, P.A. (2016) The geography of Twitter topics in London. *Computers, Environment & Urban Systems* **58**: 85-96

[29] Liu, C. & Sui, D. (2017) Exploring the spatiotemporal pattern of cyberbullying with Yik Yak. *The Professional Geographer* **69**: 412-423

[30] Reavley, N.J. & Pilkington, P.D. (2014) Using Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. *PeerJ* **2**: e647

[31] https://cran.r-project.org/web/packages/streamR/streamR.pdf

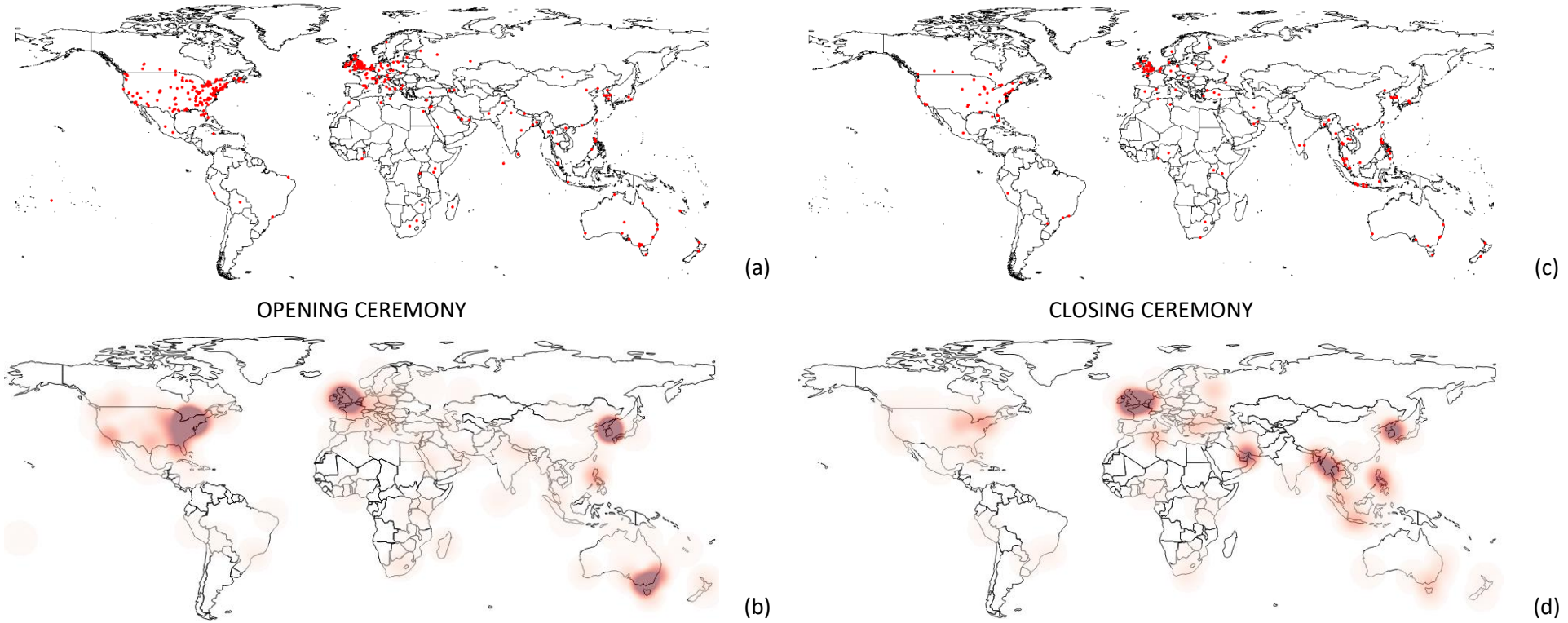


Figure 2: Geographical distribution of tweets in English for opening and closing ceremonies showed as point maps (a) and (c) and as heat maps (b) and (d)

Figure 3: Wordclouds for tweets in English for the opening and closing ceremonies.

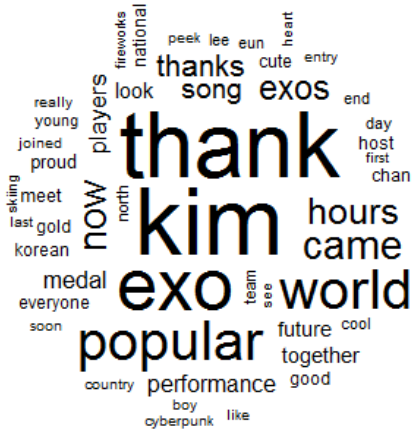
The points in Figures 2(a) and 2(c) give the location of origin for geographically tagged tweets, one point per tweet. It must be remembered that on this scale of map or where one individual sends out many tweets, the points can be superimposed. To overcome this, heat maps have been produced as shown in Figures 2(b) and 2(b). As would be expected the Korean Peninsula is a hotspot and so too is the Great Britain (despite coming 19th in the medal table) at both opening and closing ceremonies. But for the USA (4th in the medal table), tweet-based interest in the closing ceremony is much less than for the opening ceremony. The same is for Australia (23rd in medal table). New 'hotspots' open up for the closing ceremony around southern Iran and the Gulf, in Thailand and the Philippines. It must be remembered that many countries in Africa and the Middle East do not compete in the Winter Olympics, also these tweets are in English only – a more comprehensive mapping of interest would include all languages.

The wordclouds in Figure 3 summarise the tweet contents. Both ceremonies are extravaganza shows rather than a sporting event and the wordclouds reflect this. *EXO* is a Korean boyband; *BTS* is another and one of their hit songs is *DNA*. Certainly music was the main chatter. Athletes get some mention in the opening ceremony but much less in the closing ceremony.

For the Korean language tweets, we took a random sample of a thousand from both opening and closing ceremonies and bulk translated them through Google Translate. Figure 4 gives an analysis using text mining to produce wordclouds.

(a)

(b)



(c)

(d)

Figure 4: Wordclouds of the sampled tweets translated from Korean: (a) opening ceremony, (b) closing ceremony, (c) commonalities between the two, (d) differences between the two.

Two words stand out in the commonalities wordcloud – *thank* and *Kim*. Typical of the thanking is the tweet: "*Thanks for revealing the flame of hope Thank you for lighting up our wishes and dream of peace*" in the opening ceremony, and "*Thank you for your love to the 2018 PyeongChang Winter Olympics*" in the closing ceremony. Of course *Kim* might be thought to reflect the surprise diplomacy by North Korea, but it is a very common name in Korea and many of the tweets specifically refer to Kim Yu-na (a retired figure skater) during the opening ceremony, and during the closing ceremony refer to the South Korean women's curling team (all of whom are named Kim) – the so-called Garlic Girls who won silver. The boyband *EXO* features strongly more so in the closing ceremony when they performed live. In Figure 4(d), the term *growling* features as a big difference between closing and opening ceremonies and refers not to animal noises but to one of EXO's music video dramas. There is obviously a strong, local cultural content to the Korean tweets.

Tweets, along with other text such as emails, can be subjected to sentiment analysis. This is where content of the tweets are compared to lexicons of positive and negative words and each tweet given an overall net score. Emoticons can be difficult to process, so were removed before the analysis. All tweets from worldwide in English and the 1000 translated random sample of Korean tweets were run through a sentiment analysis, the results being given in Figure 5.

Row Labels	Count of Open_score	
-4	15	0.01%
-3	163	0.13%
-2	798	0.63%
-1	8637	6.80%
0	76054	59.87%
1	30957	24.37%
2	8379	6.60%
3	1605	1.26%
4	392	0.31%
5	13	0.01%
6	29	0.02%
Grand Total	127042	

(a)

Row Labels	Count of Close _score	
-5	2	0.00%
-4	8	0.00%
-3	82	0.04%
-2	941	0.51%
-1	7726	4.21%
0	99701	54.29%
1	53563	29.17%
2	14898	8.11%
3	6343	3.45%
4	322	0.18%
5	47	0.03%
6	9	0.00%
7	6	0.00%
8	1	0.00%
10	1	0.00%
11	3	0.00%
Grand Total	183653	

(b)

OPENING CEREMONY

CLOSING CEREMONY

Row Labels	Count of Open_score	
-3	1	0.10%
-2	4	0.40%
-1	13	1.30%
0	316	31.60%
1	247	24.70%
2	202	20.20%
3	216	21.60%
4	1	0.10%
Grand Total	1000	

(c)

Row Labels	Count of Close _score	
-3	4	0.40%
-2	1	0.10%
-1	102	52.20%
0	522	52.20%
1	277	27.70%
2	71	1.90%
3	19	1.90%
4	2	0.20%
5	2	0.20%
Grand Total	1000	

(d)

Figure 5: Results of sentiment analysis of tweets: worldwide in English (a) opening ceremony (b) closing ceremony, sample of translated Korean language tweets (c) opening ceremony (d) closing ceremony (*Row Labels* is the net sentiment score of each tweet).

Looking at the net sentiment rating of the worldwide tweets in English in Figures 5(a) & 5(b), they approximate a normal curve with the majority of tweets tend to be overall neutral. There is not much change in the distributions between opening and closing ceremony except in the tails, a few tweets clearly being effusive about the closing ceremony. Strongly negative tweets tend not to be publishable. An example of a negative tweet (score -4) is: *"#BOYCOTT #PyeongChang2018 #Olympics #SouthKorea NATION OF UNIMAGINABLE CRUELTY 3 MILLIONS OF DOGS TORTURED EVERY YEAR*". Examples of positive tweets (rated 5) are: "*It's good to see N. KOREA and S. KOREA are ready to work have fun together. Let PEACE prevail*" and "*what a beautiful moment, amazing how much the olympics do for peace & unity. so happy to see this actually happen*".

Turning to the sentiment rating of the sample of translated Korean language tweets in Figures 5(c) & 5(d), there is a marked difference in overall sentiment between opening ceremony (average score 1.28) and closing ceremony (average score 0.38). In other words there seems to have been, overall, a considerable local cooling in sentiment. Although the sentiment distributions of the Korean language tweets do not have such long tails, this may well be due to having a much smaller sample. An example of a positive opening ceremony tweet (rated 4) is: *"@pyeongchang2018 It was the best opening ceremony of peace, harmony, excitement and impressions*"; whilst for closing ceremony (rated 4): "*Our exo is so cool!! We are so cute! They are so cool, and Kyungsoo is so cute!*".

Conclusions

There have been a considerable number of different ways to assess the legacy impact of the London 2012 Games. As with any long term project that is intended to be a catalyst for long term change and transformation, the real extent of legacy may not yet (in 2018) be apparent. The urban transformation of the Olympic Park is not expected to be complete before 2030. Cultural changes towards, say, more healthy and active lifestyles can take a long time, may even be generational. In this chapter we have outlined the many approaches taken to evaluating the impact and legacy of the London 2012 Games. Over time, the debates around the legacy of a mega-event are likely to evolve and it is important to have ways of capturing that alongside more traditional quantitative approaches. We have illustrated this by comparing the House of Lords debates in 2008 and 2018 using text mining techniques. We have also for the first time explored Big Data approaches to capturing and analysing both interest and sentiment towards an Olympic Games, specifically and the opening and closing ceremonies of PyeongChang 2018. This taps into the potential in using social media to access and analyse people's frankly expressed opinions about an event and provides richer data with which to assess some of the softer intangibles beyond conventional attendance and viewing figures. Social media data better reflects individual enjoyment and inspiration gained from a mega-event in the context of local social and cultural norms. These Big Data techniques should be further explored and developed for assessing impact and legacy of future editions of the Olympic Games.