'Big Data' techniques to improve learning access and citizen engagement for adults in urban environments

Prof Michael Osborne
Chair of Adult and Lifelong Learning
School of Education
Director of PASCAL Observatory
Director of Centre of Research and Development in Adult and Lifelong Learning
Michael.Osborne@glasgow.ac.uk
Tel: +44 (0)141 341 4

Dr Catherine Lido
Lecturer
School of Education
Corresponding Author
Catherine.Lido@glasgow.ac.uk
Tel: +44 (0)141 330 2733

ABSTRACT
This presentation explores the emerging concept of ‘Big Data in Education’ and introduces novel technologies and approaches for addressing inequalities in access to participation and success in lifelong learning, to produce better life outcomes for urban citizens. It introduces the work of the new Urban Big Data Centre (UBDC) at the University of Glasgow, presenting a case study of its first data product – the integrated Multimedia City Data (iMCD) project. Educational engagement and predictive factors are presented for adult learners, and older adult learners, in a representative survey of 1500 households. This was followed up with mobility tracking data using GPS data and wearable camera images, as well as one year’s worth of contextual data from over one hundred web sources (social media, news, weather). The chapter introduces the complex dataset that can help stakeholders, academics, citizens and other external users examine active aging and citizen learning engagement in the modern urban city, and thus support the development of the learning city. It concludes with

---

1 The work described within this paper was supported by the UK’s Economic and Social Research Council (ESRC) under Grant Number 651920/1, and is a revised version of a forthcoming book chapter, Lido, C., Osborne, M., Livingston, M., Sila-Nowicka, K., and Thakuriah, P. (2016) 'Big Data' techniques to improve learning access and citizen engagement for adults in urban environments. In: Scott, L., Watson, C. and Wu, A.T. (eds.) Quality of Life in Adult Learning. Series: Adult learning in professional, organizational, and community settings. Information Age Publishing, Inc.: Charlotte, NC. (In Press)
a call for a more three-dimensional view of citizen-learners’ daily activity and mobility, such as satellite, mobile phone and active travel application data, alongside administrative data linkage to further explore lifelong learning participation and success. Policy implications are provided for addressing inequalities, and interventions proposed for how cities might promote equal and inclusive adult learning engagement in the face of continued austerity cuts and falling adult learner numbers.

Keywords: Big Data, Lifelong Learning, Urban Inclusion, Learning Cities
INTRODUCTION

This chapter explores the emerging concept of ‘Big Data in Education’, and aims to move beyond using simple administrative datasets, to look toward novel technologies and approaches for examining (and targeting) interventions that address inequalities in access to, participation and success in lifelong learning. The chapter focuses on urban challenges, and provides an exemplar of research into the use of city spaces for learning. We will present the work of the new Urban Big Data Centre (UBDC) at the University of Glasgow, funded by the Economic and Social Research Council as part of the recent UK drive to harness the power of Big Data, to improve urban citizens’ lives. The chapter highlights the Centre’s remit of targeting spatial inequalities in education, and the implications this might have in creating happy, healthy, engaged citizen-learners. The UBDC is currently gathering datasets relevant to the field of education, health and transport in the city of Glasgow, and making such data more open to the public.

The UBDC’s first data product, gathered through the integrated Multimedia City Data (iMCD) project, is available for use by academics, policy practitioners and the general public in order that they may investigate urban issues, and the resilience of citizens in the face of these issues, within their city. The work of this project is outlined as an exemplar of harnessing Big Data and novel technologies to examine the links between forms of formal, non-formal and informal learning and successful life outcomes, such as employment, health and well-being within the city context of Glasgow. Educational engagement, health and employment-related data will be presented for the adult learners in a representative survey of 1500 households, a subset of whom were followed up with mobility tracking using real-time GPS data and lifelogging camera images. This data is set within the context of one year’s worth of social media data capture, focusing on Education ‘Tweets’ within Twitter and other social media postings, as well as other contextual data such as news and travel feeds relevant to the Greater Glasgow area. Preliminary data will be presented here and implications for learning and well-being in the city will be discussed.

This chapter introduces the complex data that can help us examine active aging and citizen learning engagement in the modern urban city. The chapter’s objectives are to offer a case study of Glasgow data for adult learners and illustrate how such data can provide a three-
dimensional view of citizen-learners’ daily activity and mobility. Such data, alongside knowledge of engagement in various forms of learning (including in the workplace, community and family contexts), will provide policy implications addressing inequalities in access to, participation and success in progression into further and higher education, and ultimately, lifelong learning.

The chapter will also present an overview of empirical work surrounding inequality in learning, including the declining engagement of older adults in lifelong learning in the UK. It will then discuss the role of learning in the promotion of active aging, health benefits and cultural and civic engagement in urban settings. It will then present the work of UBDC, analysing preliminary data from the iMCD project, specifically in relation to older adults’ recent participation in various types of learning, cultural and civic engagement and physical mobility in and around the city of Glasgow (using GPS). Finally, it introduces the myriad novel data collected by the Centre which can shed light on three-dimensional movements of such learners around areas of formal and informal learning. Interventions are proposed for how cities might better make use of their spaces to promote equal and extensive adult learning engagement in the face of continued austerity cuts and falling adult learner numbers. The work also explores how this approach can be used to operationalize the key features of learning cities as outlined by UNESCO (2013).

BACKGROUND

The issue of equitable participation in education has been of political and social interest to governments worldwide over several decades, resulting in global attention from policy-makers and practitioners, and even the mainstream media. International bodies such as UNESCO and the OECD, in addition to individual national governments, have consistently advocated for widening access to and participation in further and higher education, and beyond into lifelong learning. Reasons cited for improved equity in HE include both social and cultural benefits of raising aspirations and qualification levels, as well as the more obvious economic benefits (Commission for Employment and Skills, 2009; Leitch 2006; Osborne, 2003; Osborne, Rimmer & Houston, 2015). Therefore, strong arguments can be made for lifelong learning and improved quality of life in older age (Findsen & Formosa, 2012).
However, much recent literature surrounding widening participation (WP) interventions focuses on school-based initiatives and transitions immediately following school-leaving, rather than adult and older adult learners (Harrison & Hatt, 2013; Moore, Sanders & Higham, 2013; Thomas et al., 2005). Examinations of participation from adult learners are increasingly neglected and work that considers transitions from non-formal to formal learning programmes can be difficult to pinpoint (Osborne & Houston, 2012; Moore, Sanders & Higham, 2013; Schütze & Slowey, 2013). Additionally, the socio-demographics of such adult learners are often ignored. Evaluation of the participation of adults (in all forms of learning) is arguably an urgent concern, given that UK adult learner numbers are falling in all sectors, including in HE (Osborne & Houston, 2012; UCAS, 2014). This is despite clear evidence from a range of studies that opening HE for adults yields vast social/cultural and economic benefits, as well as promoting a healthier aging population and potentially leading to the inter-generational transmission of learning (Commission for Employment and Skills, 2009; Leitch 2006; Schütze & Slowey, 2013 & Thomas et al., 2005). However notwithstanding the benefits of increasing mature student participation, it is younger people who over the last two decades have been the focus of most attention on the basis that early intervention produces more beneficial outcomes for all stakeholders (individuals, employers and the state) (Moore, Sanders & Higham, 2013).

Many adult learners are further ‘multiply disadvantaged’ – that is, they fit the definition of under-represented student groups in multiple ways by virtue of their socio-economic class, race, gender and geographical location (ibid, Osborne, Rimmer & Houston, 2015). Therefore, analyses must progress beyond simply analyses of the demographic variables of groups and examine learners more holistically, including by widening notions of engagement, and placing learning within locality and geographic context. There are calls on present interventions to acknowledge the ‘diversity of diversity’ (Moore, Sanders & Higham, 2013, p. x), and to begin to think ‘outside the box’ in examining the complex factors affecting participation.

Examining the effectiveness of WP into Higher Education (or indeed attainment and engagement across the lifespan), without embedding that learning within the context of
place and the physical learning environment, loses a vital dimension. It could be said that of all public services, learning institutions may be the most deeply embedded in place, as schools mainly serve particular communities and draw students from a defined catchment area (Kintrea, Osborne & Lido, in preparation). Bannister, Kintrea and Pickering (2013) have recently illustrated that, although deprivation is not related to aspirations in a simple linear fashion, young people have significant relationships with their local area/neighbourhood, which affect cultural capital, norms, beliefs and expectations about life priorities, including educational attainment and employment opportunities and outcomes. Lupton and Kintrea (2011) further indicate that interventions for effecting change in education aspirations will need to be complex and those targeted at a ‘community level’ will need to consider, amongst other variables, the “quality and experiences of schooling, labour market constraints and employment practices, neighbourhood connectivity, environments and reputation, the availability of information and household financial constraints” (p. 330). In this way, initiatives to increase learning participation aspirations need to be considered within their geographic context, if cities are to achieve significant shifts in community engagement (and aspirations toward) learning.

However, few education researchers prioritise the collection of city-wide indicators that surround the learning engagement of their citizens, from school through to further and higher education into lifelong learning, and to examine this in conjunction with participation in learning. Those education researchers who do harness Big Data research tend to use large administrative datasets solely in the field of ‘learning analytics’, in order to target efficiency and cost-effective delivery of education (Niemi & Gitin, 2012). However in our research we seek to join the emerging body of evidence moving beyond simple linear evaluations of educational outcomes and heed the call of recent reviews (Moore, Sanders & Higham, 2013) to provide more nuanced approaches to assessing whether WP has been successfully embedded within place and the specific geographic context. Therefore, presenting the approach being applied by the UBDC in the city of Glasgow illustrates what we can learn about the characteristics of the city’s learners, their orientations, engagement and experiences (physically, educationally and virtually in online engagement).

**CASE STUDY OF GLASGOW: iMCD PROJECT**
Glasgow proves an interesting case study for presentation when examining ‘learning cities’. Learning cities are those that successfully mobilise their resources to realise learning as a universal human right for all of their citizens, therefore, they “reinforce and create individual empowerment and social cohesion, economic and cultural prosperity, and sustainable development” (UNESCO, 2013, p.2). Glasgow has a long history of engagement with the Learning city agenda (since the 1980s) and can also act as an excellent cross-cultural comparator firstly given its post-industrialist history, areas of centrally located deprivation and also as its status as the largest city in its region (Scotland). The city and its surrounds has higher levels of mortality than other deprived cities, both in Britain and across Europe which persist even when accounting for socio-demographic factors (Walsh, Bendel, Jones & Hanlon, 2010). Given links between health and well-being from Lifelong Learning engagement (Gray & Leyland, 2009; Gray et al., 2012), it is logical to apply urban Big Data approaches for imaginative solutions to such urban issues within this city.

The UBDC’s first data product has emerged from the iMCD project, which consists of a large-scale, representative survey of 1500 households within the Glasgow area (including a 24 hour travel diary), combined with a sub-sample with sensor data collecting – including GPS trails for one week’s worth of travel as well as ‘Lifelogging’ camera images from 48 hours of travels (which included an additional 24 hour travel diary to match with the GPS and images). This data is set within a large-scale social media capture (including 12 months’ data capture and cataloguing of online news archives, social media posts, user blogs [Tumblr] and photo sharing sites [e.g. Flickr]). Education and Learning City keywords have also been digitally crawled and catalogued by the UBDC. This purposive, survey-based approach, with the wider information extraction from sensors and social media data capture, allows researchers to better understand the context and motivations driving social aspects of this urban environment, illuminating the context that may affect people’s attitudes, lifestyles and beliefs, and their mobility and time use patterns. Each strand will be addressed separately below, but it should be noted that the Centre is additionally acquiring satellite images and other transportation-related data, to be housed alongside administrative data for the city such that urban indicators will be linked with the present survey data below.

*Survey*
The survey assesses individuals’ patterns of travel activity and daily tasks, values and priorities in the domains of education, transport, sustainability, technology, and cultural and civic engagement and was developed following an exhaustive review of (largely UK) national survey questions in the domains of interest, and content validity was assessed by a team of eight subject matter experts (SMEs) from inter-disciplinary backgrounds. The draft survey content was compared against the 42 UNESCO (2013) features of Learning Cities developed in Beijing in order to ensure that it could directly measure (or allow indirect comparisons at the city/country level) those key concepts for a successful Learning City amenable to collection through this method.

The survey collected a rich variety of demographic information- including age, ethnicity, nationality, migrant status, relationship status and religion, as well as household demographic information, such as number of children and income. The survey also collected extensive data on learner engagement over the previous 12 months in formal (structured/leading to nationally recognised qualifications), non-formal (structured but not leading to national qualifications) and informal (self-led unstructured or experiential) learning (consistent with definitions from the Adult Education Survey [AES, 2014] and Organisation for Economic Co-operation and Development [OECD, 2015] but operationalised in line with government Office of National Statistics harmonised guidelines [ONS, 2014]). For instance, the formal learning question read:

‘During the last 12 months, have you taken part in any course or apprenticeship intended to lead to a recognised qualification, regardless of whenever you completed the course or obtained the qualification?’

Additionally, ‘show cards’ were used to give examples of various potential forms of learning. These were followed by questions concerned with mode of this learning, time committed to the learning, funding and motivation. In addition, the survey collected information on people’s literacies in the domains of finance, health, numeracy, politics, information technology and ecology, as well as foreign language skills.
A representative sample was taken of households in the Glasgow Clyde Valley Planning Area (Glasgow City Council area, contiguous local authorities plus Inverclyde), targeted using Royal Mail Postcode Address File (PAF) with a stratified random sampling design to ensure a representative sample of the population for age, gender and ethnicity. The sample was also stratified by deprivation for a roughly even number of respondents in each deprivation decile. The preliminary iMCD data is from n=1,037 adult interviews from 710 households. The average age was 50.73, with a standard deviation of 18.72, ranging from 16 to 93, which is slightly older than would be expected from 2011 census data (National Records of Scotland, 2015), but is largely in line with other surveys, such as the Scottish Household Survey (2015). The sample consisted of 45.2% male respondents and 54.8% female. They are largely of White British or European ancestry (95.7%), and are Scottish (82.1%) or British born (45% of those not Scottish born). The average household contained 2.23 persons, of whom 1.93 are reported to be adults, and of whom we successfully obtained responses from 1.5 adults per household.

**Sensors**

Upon completion of the survey participants were asked if they wished to take part in further research, and were followed up with a request to participate in the iMCD sensor project. A sub-sample of 404 respondents agreed to carry a GPS sensor for one week, yielding n=358 valid GPS trails. They were also asked to complete a written activity/travel diary for the first 24 hours for comparison, and n=230 agreed to use a wearable camera as well for 48 hours. Along with each of the images, additional information was collected by the Lifelogging camera: acceleration (Accelerometer), the direction in which the camera is facing (Magnetometer), brightness and the colour bands for each of the taken images (Colour sensors) and movement (Motion Detector). The GPS device saves the coordinates of all locations a person visits during the week, and the life-logger captures images throughout the first two days of the same week. The GPS tracker collects points at a 5-second interval and the wearable camera takes images every 5 seconds, which allows for linkage of the GPS data with images and the coded travel/activity diary. Finally, these were linked to the survey participants using a unique ID.
This layered methodology was proposed by Thakuriah and Geers (2013), and is based upon using data from infrastructure-based, vehicle-based, mobile, portable and wearable sensors alongside background data from censuses, administrative data sources and other data programs to form a variety of contextual socio-demographic, built environment and other place-based attribute data for analysis. An earlier example of such work stems from the Chicago area Spatial Decision Support System (Minocha et al. 2008; Cottrill & Thakuriah, 2010). All survey and sensor data collection was subjected to University of Glasgow ethical approval.

Social Media

As stated, the survey (including a travel diary for 48 hours), alongside the follow-up sensor data collection (GPS for one week and lifelogging images for 48 hours), is set within a large-scale social media capture (including over 12 months’ data capture and cataloguing of online news archives, social media posts, user blogs [Tumblr] and photo sharing sites [e.g. Flickr]). The ‘Glasgow memory server’ has been available in searchable ‘dashboard’ form, since early 2016. It consists of social media capture of public data available from both news websites (e.g. BBC, Scotsman, Guardian, Daily Record, Evening Times, Evening Express and Courier), and social media sites (e.g. Twitter, Flickr and YouTube). New sources were searched, indexed and stored periodically throughout the project and therefore include key events such as the Commonwealth Games in Glasgow, the Scottish Independence Referendum both in 2015 and changes to the Curriculum for Excellence (Scottish Executive, 2004) initiative within the Scottish School system.

For data ‘crawling’, the team used Application Programming Interfaces (APIs) provided for accessing each social media type, which only permit access to content that has been published into the public domain by users. For Twitter, they collected tweets by users and other tweeters, in two manners: those public tweets mentioning Glasgow or related events (e.g. by geo-location and hashtag); and a < 10% random sample of all tweets, which were later analysed for more Glasgow-related tweets. Fewer than 10% of such tweets are geo-located, but those that are allow geographically searchable analyses and visualisations. Education data was searched and collected in the domains shown below.
**Textual media retrieval**

Tweets *geo-located* in Glasgow used a polygon around Glasgow as input for the Twitter API to filter the global stream of Twitter in order to capture these tweets, for instance from the following users:

Table 1: Twitter users captured

<table>
<thead>
<tr>
<th>GTC Scotland (General Teaching Council) @EducationScot</th>
<th>Confederation of British Industry (CBI) Scotland</th>
<th>FoE Glasgow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education Scotland</td>
<td>SCDI Glasgow</td>
<td>What’s On Glasgow</td>
</tr>
<tr>
<td>EducationHUB</td>
<td></td>
<td>Health Education Art</td>
</tr>
</tbody>
</table>

**Tweets**

Tweets that contained the following terms or hashtags (case-insensitive) were related to known events in and around Glasgow during the time period (e.g. CWG [Commonwealth Games] and indyref [Independence Referendum]).

Table 2: Tweets captured

<table>
<thead>
<tr>
<th>Education Glasgow</th>
<th>Learning City(ies) Glasgow</th>
<th>Lifelong Learning Glasgow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills (share/ exchange) Glasgow</td>
<td>Smart City(ies) Glasgow</td>
<td>Widening participation Glasgow</td>
</tr>
<tr>
<td>Open Learning Glasgow</td>
<td>Unesco Glasgow</td>
<td>Glasgow University (maybe worth searching Tweets for Strathclyde as well and School of Art?)</td>
</tr>
</tbody>
</table>

**Facebook**
For Facebook, only posts and comments for open Glasgow-related Facebook groups were analysed.

Table 3: Facebook data captured

<table>
<thead>
<tr>
<th>Key local Further Education Colleges (FECs)²</th>
<th>Inclusive Education at the University of Glasgow</th>
<th>Glasgow Skills Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Strathclyde Centre for Lifelong Learning</td>
<td>Glasgow School of Art</td>
<td>Skills Exchange Scotland Ltd</td>
</tr>
<tr>
<td>Centre for Research in Lifelong Learning, Glasgow Caledonian University</td>
<td>First4Skills Scotland</td>
<td></td>
</tr>
</tbody>
</table>

**Preliminary iMCD Project Findings**

The preliminary data was analysed using SPSS, and initially linear regressions were run to identify key drivers for engagement in each form of learning: namely, obtaining a higher education qualification, as well as present/recent engagement in formal courses, non-formal courses and other less structured forms of informal learning. Demographic factors were entered into the regressions, followed by area-relevant factors. ANOVA was then used to further unpack emerging geographic (local authority) variations in learning engagement across the city. These are in no way exhaustive findings, but rather indicative to suggest the types of analyses that can be done and linked to our GPS data, and future datasets for emerging ‘3D learner stories’.

The preliminary analysis of n=1,037, with a sub-sample of n=377 older adults (aged 60+), were examined quantitatively in order to better understand older learner engagement in formal, in-formal, non-formal and family-learning contexts in the City of Glasgow. Findings soon to be published (Lido et al. 2016), revealed that all forms of older learning participation in Glasgow were lower than younger and middle-aged counterparts. However, there was a subset of ‘actively ageing’ learner-citizens. These older adults were socially and

² These are Vocational Education Colleges offering largely non-advanced post-compulsory education.
technologically engaged, participating not only in education, but also physical, cultural, civic and online activities (e.g. online political discussions and boycotts). In line with previous findings (Marmot Review 2010; Osborne, 2014), these older learners reported better overall mental and physical health, but interestingly they were also more likely to be working full or part-time and caring for others. Long-term disabilities impinged upon participation in learning for Glasgow citizens and were associated with less engagement in non-formal learning activities, though had no effect on formal and informal levels of participation.

With regard to preliminary GPS sensor findings, an examination of the engaged older learners’ GPS trails showed more city activity than their matched non-learning-engaged counterparts. Place-based variables, such as feeling safe and belonging to the local area, moderated adult, and particularly older adult, participation in learning all types of activities. The full dataset will be accessible for analysis by researchers and the general public in late 2016, providing a complex data source to explore demographically diverse learners’ journeys within an urban context.

**BIG/ NOVEL DATA FOR GLASGOW: COMPLEMENTARY UBDC DATA**

There is presently a push for the use of Big Data by governments in order to cost-effectively examine urban issues, and this has included the push for using Big Data to examine inequality in education and the promotion of lifelong learning (see Osborne & Lido, 2015 for a review). The primary purpose of UBDC is to act as a rich research resource promoting the use of innovative methods and complex urban data to address global city challenges, including education as well as transportation, housing, technology and ecological issues in urban settings. Therefore, it is important not only to examine the implications of the iMCD data collected above, and the fruitful findings of preliminary analyses for adult and older adult learners in the city, but that this data be made available to external users, including policy-makers, citizens and other academics who may wish to explore learning in the city of Glasgow. Furthermore, the data described above in the iMCD has much potential for linkage to wider administrative and technologically novel data. This section will outline the centre’s progress toward acquiring and linking data in the Greater Glasgow Area, including administrative educational data (Scottish Pupil Census, Higher Education Statistical
Authority [HESA] and Higher Education Statistical Authority [UCAS]), Strava data (an app measuring cycling and active travel⁢), mobile phone, LiDAR data.

**Administrative education data and linkage**

The acquisition of Scottish Pupil Census data (2007 to present) from the Scottish Government relates to a linked research project with UBDC on educational disadvantage and place (Kintrea, Osborne & Lido, 2015), which implements the linkage of innovative ‘Big Datasets’ in order to examine a range of social and educational challenges within cities. It aims to better understand the relationship between place and educational disadvantage in the Glasgow city region, in order to identify the drivers of disadvantage, as well as academic success in school and beyond.

In particular, the research aims to answer two main questions:

- How does place shape educational disadvantage and, alternatively, learning success?
- What is the ‘spatial school system’ in the Glasgow city region and how does it work in relation to the production and reproduction of educational (dis)advantage?

The project intends to inform policy options for narrowing the gap in educational attainment experienced by young people from disadvantaged backgrounds. It is directly relevant to the Scottish Government’s overarching ‘Wealthier & Fairer’ and ‘Smarter’ objectives (see report by Arnott & Ozga, 2010). In particular it addresses the same concerns about the ‘attainment gap’ (the difference in educational attainment between children living in disadvantaged and non-disadvantaged areas) that have informed the development this year of the Scottish Attainment Challenge and the Attainment Scotland Fund (see report by Sosu & Ellis, 2014).

The data acquired by the Centre, for use on this project, includes both contextual administrative data for the Glasgow city region, drawn from pupil, school and teacher census records, as well as pupil attainment within the public examination system of the Scottish Qualifications Authority (SQA). This data allows us to address questions about

---

³ See http://labs.strava.com
spatial inequalities in educational outcomes across the Glasgow city region. Additionally, we can provide linkage for a sub-sample of pupils in the iMCD household survey (described above) to their school census and attainment data. While the broad relationship between educational disadvantage and place in Scotland is established by existing research, there is little detailed understanding beyond the observation that young people who live in disadvantaged areas have lower educational attainment. Access to administrative data is needed to understand the patterns of disadvantage (and advantage) that exist in order to inform interventions in greater depth.

The purpose of such data linkage with iMCD data (outlined under the methodology section above) is to extend the understanding of the neighbourhood, and familial and cultural contexts of educational attainment by associating educational outcomes and school-related variables obtained from official data with extensive household survey data about school children’s family and living arrangements, housing, family leisure activities, travel behaviour and parents’ educational experiences, attitudes and behaviours.

**Higher Education Data: HESA and UCAS**

The Centre is also in the process of acquiring higher education administrative data, both for contextual data regarding applications and HE destinations (performance), and skilled employment upon leaving for pupils from the Glasgow area. In addition, we seek to link this to a sub-sample of iMCD survey participants (for whom we have permissions) in order to examine higher education trajectories for our household participants. This is post-millennial data from UCAS student applications and HESA student records, and will be licensed for third-party use through UBDC.

We are presently acquiring fine grained data (some of which is individual and other requires minimum cell sizes of n=3 or n=5). We are seeking such data for full and part-time Scotland-domiciled student applications for application years of entry to Higher Education (HE) study from 2000–2015 (which maps onto the cohort for whom we are acquiring pupil records for the iMCD). Also we seek their subject of study, performance, degree outcomes and initial employment once in study. This will allow the Educational Disadvantage and Place team, as
well as third-party users wishing to conduct academic analyses, to examine the impacts of policy shifts post-millennium for HE participation in Glasgow (and beyond into the UK).

We will acquire all the main demographic fields/variables (includes age, gender, SES) and study demographics (institution, subject), along with variables describing institution 'type', to develop predictive models for applications, acceptances, performance, achievement and destinations within UK Higher Education for Glasgow-based students. The data will contain home residence area indicators within Scotland to map and spatially model flows of Glasgow-student HE applications and allow a detailed examination alongside deprivation indices.

Therefore, such HE and early employment data completes a 'lifelong learning' dataset for the Centre unlike any found elsewhere, one that follows Greater Glasgow area pupils from a specific time period, from their school records to school leaving, higher education applications, higher education performance and even into initial employment. Such lifelong learning administrative and household survey-linked data is of interest to government stakeholders, academic users and the general public interested in educational opportunity and achievement in the local area. Research questions include those around barriers to lifelong learning; inequality in attainment and progression; and ultimately place-based effects on educational achievement and life outcomes. These are key urban big data priority areas for our data sourcing and acquisition.

Government stakeholders may be interested in this data in the context of debates around standardised testing in schools, progression to skilled employment, and the socio-economic disadvantage of young people and how this is related to a household cycle of exclusion. Higher Education Institution stakeholders may be interested in topics surrounding widening participation, recruiting under-represented student groups from specific Glasgow areas, and engaging with these students successfully. Other researchers might be interested in mapping and visualising deprivation alongside flows to different HEIs in Glasgow, Scotland and the UK, and such visualisations might inform the general public when directing decisions surrounding young people’s progression to further study.
Strava

There is increasing interest in research on cycling not only as a mode of travel, but also in understanding its role in sustainable and healthy lifestyles. Academics from public health, transport and urban geography want to understand the role this mode of transport might play and what factors might impact on the numbers choosing cycling. There have been a number of studies which have considered the public health impact of people choosing cycling as a means of transport (Garrard, Rissel & Bauman, 2012; Rojas-Rueda, de Nazelle, Tainio & Nieuwenhuijsen, 2011). However, research in this area is hampered by the poor level of data available on cycling. The development of GPS technology and emergence of companies providing apps and websites which allow people to record their cycling activity raises the possibility of new sources of data. For instance GPS data has been used to demonstrate the relationship between cycling and positive health outcomes (Dill, 2009; Calise, Dumith, DeJong & Kohl, 2012; Edwards & Mason, 2014).

Strava is the dominant app for millions of cyclists in the UK. The data it provides potentially represents the best of its kind on cycling currently available. While cyclists are a small proportion of any given area’s population, and with a possible demographic bias towards young and middle-aged males (Griffin & Jiao, 2015), there is evidence that there are very high correlations between Strava recorded journeys and traditional measures of cycling prevalence (e.g. counting cyclists using enumeration methods). For instance, counts at a number of sensors throughout Seattle were found to have correlations of over 90% with Strava data.

Datasets created through Strava have the potential to provide insight on cyclists’ revealed preferences for roads with cycling lanes, paths, terrain characteristics, in and near populated places with businesses. This data can help planners understand how cyclists use current infrastructure and plan changes more effectively. The data can be used to evaluate future or past changes giving better understanding of the more positive interventions. More importantly for our work, such data provide context to the existing iMCD household survey, which examines not only the learning engagement and attainment of the household members (formal, informal and non-formal and family learning) alongside the transportation attitudes and activities (in the form of a travel diary). Strava data can further
inform the preliminary findings of iMCD that those older adult learners who are presently engaged in learning activities are also more likely to engage in active travel around the city.

**Mobile phone**

In the literature there exist a number of examples of monitoring a population’s geography and mobility using data collected by telecommunications companies. Such mobile phone data represents very detailed information regarding individuals’ movements aggregated to the level of various geographic units to prevent identification and movements of specific mobile phone users. Researchers can use such data to infer the most significant places that a mobile user visits frequently, and to predict the future location of the user, which is central to understanding individuals’ mobility and social patterns (Gonzalez, Hidalgo & Barabási, 2008; Wesolowski, Eagle, Noor, Snow & Buckee, 2013; Bengtsson, Thorson, Garfield & Von Schreeb, 2011; Deville *et al*. 2014). Much like the GPS, the Centre’s acquisition of such data will allow questions to be asked regarding how individuals (with certain demographics) are moving around the city over a longer period of time than our GPS methods, and will offer a more diverse sample from various areas of deprivation in and around the Greater Glasgow area.

Various clustering methods allow differential analyses of mobility patterns by gender and various other demographics, such as age (Arai *et al*., 2015). Such data may be used in the prediction of the crime hotspots (Bogomolov *et al*., 2015) and even estimating food consumption and poverty indices by different groups of users (Decuyper *et al*., 2015). All of these contextual issues may affect an individual’s or a household’s ability to participate in lifelong learning, active citizenship and to obtain their concomitant health and economic benefits. For instance, the mobile data can provide number of internet connections, phone calls, texts along with the location for specific users and we could potentially identify so-called ‘silver surfers’ and digitally engaged older adults to assess technological development among older learners in Glasgow.

Finally, the mobile phone data may be successfully used to construct proxies for origin destination-matrix (Hayes & Wilson, 1971; Fotheringham & O’Kelly, 1989; Sila-Nowicka *et al*., 2014). Therefore, we can examine mobility overall for older adults in and around the city
but also look for specific pattern of journeys for certain demographic groups, say around Glasgow University in the west end of the city (and not just where they go, but how fast and how often). Such derivations could be used as a base for urban planning and aid in future road network development (Wang et al., 2013) improving quality of life of the city users, and access to and use of a city’s learning spaces. Much of this data can be used to increase the mobility for those with restricted mobility due to disability, age or economic restrictions.

**LiDAR**

LiDAR (Light Detection and Ranging) is an active form of remote sensing (Whitworth, 2015) which uses lasers (from aeroplane and satellites) to create 3D virtual models of the environments around us (including the urban environments in which we learn). Using laser pulses LiDAR records accurately the distance between a sensor and a target object. In the bid to ‘open up’ data the Environment Agency of the Scottish Government is making such data available freely and openly. The UBDC will also play a role in making such data available and usable to those interested in further examining a ‘bird’s eye view’ of Glasgow, Scotland and the wider UK. Such data can “help organisations, businesses, and individuals to do everything from manage forests, discover hidden archaeological features, and even create virtual reality worlds for gaming” (ibid).

In the modern world there are various applications for LiDAR data to improve management of different processes or areas such as:

- Digital Elevation Models (DEM), which are used to plan the park and tourism area. Highly accurate land surface modelling helps, for instance, to find the best location within an area to construct a playground, plant trees or plan walking paths (Fjørtoft and Sageie, 2000; Sulebak, 2011);

- Surface models created from LiDAR that can be used as a graphical value to maps. DEM is added underneath all layers that show the 3D view of the land (Shan & Aparajithan, 2005; Matei, Vander Valk, Cheng & Sawhney, 2013).

LiDAR data can be added to aerial photography to develop a 3D view, which makes it easier to plan roads, buildings, bridges and rivers (Abayowa et al., 2015). By making such data available, UBDC can allow urban planners to address inequalities in learning, attainment and
income, for instance identifying areas of high concentrated populations (and high rise buildings) but low levels of participation. Such 3D models allow planners to implement interventions to ease travel, mobility or safety with areas of deprivation in Glasgow.

**Ethical Implications of novel big data approaches**

As technological approaches are advancing, and social scientists move closer to capturing real-time, real-world data about how their citizens are engaging with their city-spaces (such as learning engagement), attention must be given to the nature of the ethical issues specific to each of these strands of data described above, but also to an overall project (such as iMCD) or even to research institutions themselves. One of the core ethical issues for the present project stemmed from the function of the Urban Big Data Centre, which is to make data such as these available to a number of different potential users, for example academic, policy makers and wider public users. Potentially one of the major issues for Big Data is that by combining data you undermine confidentiality of either data set allowing respondents and there data to be identified so it is in its combination that problems might arise. Therefore, not only must researchers ensure projects meet subject and institution specific research ethics, but consideration is given to freedom of information and the ethical treatment based upon its disclosive nature. Open data is by definition non-disclosive, however, UBDC deals with safeguarded as well as controlled (highly disclosive data). It is important that citizen participants themselves are aware of the nature of the data collected, and informed consent is documented, especially as the nature of Big Data means it is complex and the threats to the privacy may not at first be apparent... In the course of this project, for instance, written consent was required, but additionally participants were given the opportunity to review (and withdraw) the GPS and lifelogging camera data. Additional ethical and data management protocols (including the treatment of social media data) can be found at ubdc.ac.uk.

**RECOMMENDATIONS FOR LEARNING PROMOTION IN CITIES**

Learning has direct implications for improving quality of life, particularly in older adults (Findsen & Fermosa, 2012), and therefore, the Learning Cities agenda seeks to promote learning engagement for all citizens (UNESCO 2013) in an effort to produce ‘happy people’ in ‘happy places’. Urban research, such as that of the UBDC, seeks to address urban
challenges (such as urban inequality) through citizen engagement. The preliminary analyses of the iMCD survey and sensor data indicates that participation in learning for older adults and educational engagement, and attainment levels are indeed linked with geographic place, neighbourhood space, school choice and transport in Glasgow. This will most certainly moderate the effectiveness of any widening participation initiatives, be they outreach, informational or economic. Yet, the exact relationship between these variables is not yet clear. Linked data of the sort being collated and held by the UBDC Glasgow, which will ultimately be accessible by academics, practitioners and the public alike, are needed to explore in some depth the factors that impinge on application and admission to HE. The UK government has presently called for such transparency from agencies offering services to university such as UCAS, requesting regular reports on acceptance rates for those from minority ethnic groups and those from economically disadvantaged backgrounds to further its commitment to ‘openness’ and ‘widening access’. It might be added that true evaluations of the success of such initiatives would need to visualise such data to better examine the link from postcode to institution (and ‘type’ of institution and course), whilst maintaining a focus on demographic variables such as age and ethnicity.

For Glasgow specifically, this data will lead to mapping postcodes for learner engagement, and using GPS data alongside photos, satellite data, travel diaries and even Google street view to allow researchers, even remotely, to investigate areas identified as having 'less engagement' in various types of learning. This allows the researcher/policy-maker to immerse themselves in this environment (including virtually) and identify further necessary interventions. For instance, in terms of diversifying the locations of education provision, the University of Glasgow is presently undergoing a £135m expansion of an urban-based university campus, and additionally expanding across the city to open academic space in previously marginalised communities (known as the ‘East End Hub’, it is a substantial investment to develop a satellite campus in a traditionally deprived area of Glasgow, illustrative regeneration, community outreach and tackling inequality in university access). The techniques we have described provide future opportunity for evaluating the success of such outreach initiatives including whether new patterns of travel emerge (and as a consequence new types of learners) emerge as new centres of learning develop, and help the university shape its relationships with the community learning engagement.
More generally data should be systematically gathered over a longer timespan on all forms of publicly funded interventions to widen access, promote attainment and retention and beyond to employment, postgraduate study and lifelong learning. In this way, we will have a better picture of ‘active learner citizens’ engaging with and contributing to their cities/local environments (versus those who feel marginalised, lack a sense of belonging and safety within their physical environs). There has been a plethora of funding initiatives over the years, but little attention to their sustainability and ‘success’. Although the data presented is not yet complete, nor integrated with larger administrative datasets, the present survey embedded within the context of other UBDC data should yield a holistic picture beyond ‘the student lifecycle’ to the student ‘life course’ and allow the city to better target interventions promoting learning within and outside of formal higher education institutions.

This chapter summarised the iMCD project (‘Understanding Glasgow’ survey, sensors and social media data capture), with emphasis on how it can be used to assess adult (and older adult) engagement in learning, as well as health, social, cultural, political engagement and overall quality of life. It then discussed the contextual data UBDC has gathered, to illustrate how active travel (Strava cycling data), mobile phone (apps showing GPS) and LiDAR and satellite data can be used to examine and improve access for adults in and around the city and overall improve the quality of life of older adults. This data, when fully analysed, has considerable implications for policy-makers in education, as well as cognate sectors, for improving the lives of citizens, particular in regions committed to the Learning City Agenda. Preliminary analyses contribute to the evidence base that links participation in learning (formal, non-formal and informal) with benefits in terms of health and well-being, and de facto concomitant economic benefits to the state in terms of the support that it would need to offer for an aging population. In addition, initial analyses illustrate the engagement in learning is associated with cultural and civic engagement in older adults, as well as more digital engagement and physical mobility in and around the city, arguing for 3D approaches to examining active aging. However, we have further to go in the analyses of actual behaviours of older adults in urban environments and how their engagement in education, cultural and civic activities is impacted by important infra-structure determinants. In the present study we have assembled a means of linking data concerned with housing, travel
and demography with behavioural outcome in relation to educational participation (in all forms of learning across the lifespan).

The opening up of such data allows policy-makers to consider customised interventions at an area level that could make a real difference to areas of deprivation showing disengagement, for instance with lower levels of adult engagement in formal and non-formal learning. For instance, participatory methods could allow citizens to identify areas of the cities in which they do not feel safe travelling at night, and areas which are not actively used at all (so-called ‘dead urban spaces’). These areas may act as deterrents for physical engagement in learning for area residents. Methods of online learning engagement might also be more highly promoted in these areas, and investment in IT infrastructure targeted to meet the needs of those disenfranchised from learning. Furthermore, the methods employed here not only have potential impact for the city of Glasgow, but the survey, GPS and lifelogging methods can be adopted for cross-cultural comparisons with other cities and regions in the world. The UBDC will act as resource to open us the use of such data, as well as to provide support and training for those wishing to examine learning and mobility in their own learning regions.

The methods used in Glasgow and the operationalization of learning city indicators can be transferred to other settings, and be used to allow cities such as Taipei to measure its progress objectively against reputable international standards.
References


UNESCO (2013) *Key Features of Learning Cities- Introductory Note*, Hamburg, UIL.


