Urban Informatics and Big Data

A Report to the ESRC Cities Expert Group[1]

Michael Batty
m.batty@ucl.ac.uk

Centre for Advanced Spatial Analysis (CASA)
University College London (UCL)
90 Tottenham Court Road, London W1N 6TR, UK

October 19, 2013

Preamble and Summary

Quite suddenly, it appears that cities have come back onto the research agenda. This is perhaps because it is now clear that most people will be living in cities by the end of this century but it is also because cities appear to be the power-houses of the new economy. There is an emerging view that most creative industries are clustered in cities and that cities promote such creativity because of agglomeration economies which enable increasing diversity to occur. Cities also appear to be greener as they grow in size and in this sense are more sustainable, and it appears that they become more than proportionately richer as they increase in size, all others things being equal. In parallel with all this, cities are being automated as computers begin to permeate public spaces and begin to complement digital information with material movements. These kinds of embedding of computers in the very fabric of cities have given rise to the notion of the 'smart city' where their functioning is being supported by massive data sets resulting from various kinds of sensors embedded in their built form and complemented by individuals using personal sensors to function in the city. This is what we mean by 'Urban Informatics'.

This is a time of very rapid change. Townsend, whose recent book on ***Smart Cities*** *is the first statement and critique of the movement, likens the times in which we live to those when planning was first established a century ago. He says: "The more I think about it, the more it seems like the birth of planning itself in the early 20th century. There are striking similarities to what's happening now - new voices, new scientific ideas, new institutions are emerging on a daily basis. One has to wonder what comes out on the other side for urban studies and planning - a completely new field, or a dramatically changed one."*[1] *In this paper, we will assess these changes from the perspective of the smart city and the new science of cities that is*

[1] This report has not yet been published and if quoting it, note that it as being unpublished and associated with ESRC (The UK Economic and Social Research Council)

proceeding in parallel but is by no means synonymous with it. This is bringing many new approaches onto the scene particularly from large computer and communications companies who see the city in terms of new markets for their technologies which are energy grids, routers, networks, software services and automated control systems. What we will do in this paper is sketch all this in the most general terms entitling it 'Urban Informatics and Big Data', terms that pertain to the rudiments of the smart city in terms of its analytics, software and data.

We begin by defining urban informatics and introduce the smart city movement. We elaborate this by considering the way cities are being instrumented and sensed and the 'big data' – the massive volumes of data being transmitted from these sensors in space and time – that result from this instrumentation. This is changing the perception of what the city is, it is shortening our time scales of interest, and it is having profound effects on the planning systems that have been put in place over the last 100 years that are still largely adapted to the industrial city. We explore urban analytics, models that are being used to simulate routine services and then point to the new actors involved in this debate, the ICT companies who are producing model cities which are highly automated, like the new towns of the last century, to advertise their wares.

We then shift to explore digital participation, in terms of service delivery, the development of citizen science, the construction of Apps, and the emergence of social media and new kinds of network. This takes us into questions of data and privacy and we look to the open data movement which is making cities more transparent at least in terms of public data, while at the same time noting the continually changing digital divide that exists between information rich and poor in cities and their access to technology. All of this is being paralleled by a new 'science of cities' that is somewhat divergent from these technological concerns but is focussed on thinking about the city in terms of networks and flows which of course are coincident with big data and the new analytics. We then draw all this together and pose some pertinent challenges for social sciences research, namely explorations of new patterns of mobility which we encapsulate as part of the 'space of flows' paradigm, changing conceptions of urban planning, questions of data infrastructure in terms of bigness, openness and data integration, instrumented data and social media, digital participation and online communities, the new urban geography of the smart city in terms of the new spaces being created, privacy and confidentiality of data, and finally questions of governance.

This is hardly a research agenda for the social sciences, more a set of pointers to the debate which is beginning. Readers who wish a more detailed and slightly more technical exposition are refereed to the review paper by Batty et al. (2012) 'Smart Cities for the Future' and those who want a readable broader discussion of the smart cities movement should refer to Townsend's (2013) recent book.

An Appendix is included which provides a preliminary list of university research centres and groups that deal with smart cities, urban informatics and big data

Defining Urban Informatics

Urban Informatics is loosely defined as the application of computers to the functioning of cities[2]. In its narrower focus, it pertains to the ways in which computers are being embedded into cities as hardware and as software so that the routine functions can be made more efficient, not only through automated responses but through the data that such computation generates which is central to policy analysis. This narrow focus is on control. In its wider focus, it is concerned with the use of computers and communications to enable services to be delivered across many domains and to enable populations to engage and interact in policy issues that require citizen participation.

Both these foci are currently being articulated through two key ideas. First there is the notion of the *'smart city'*, the idea that cities can become more efficient, hence smarter[3], through the use of computers and computation across wide spatial and temporal domains. The focus is on joining up or integrating operations and services and also disseminating the information associated with these activities to users through a variety of computable devices from regular PCs to smart phones. Second, these systems through their embedding into the built environment and their routine use by populations through hand-held devices ranging from cards to phones, are delivering large quantities of data about the way cities function. This is being streamed and archived in real time, hence providing a detailed spatio-temporal record of all that goes on in the functions that are being automated. These data volumes can be immense, larger than anything we have experienced in cities hitherto and these are currently referred to as *'big data'*.

A generation or more ago, the use of computers for understanding cities was largely focussed on their use for implementing models for policy analysis which in turn were based on a variety of theories about how cities 'worked'. These theories in turn were based on social physics, transportation modelling, urban economics, location theory, social ecology, and related conceptions of how the city was a functioning economic and social system. There was of course no agreement about what constituted the most appropriate theories but insofar as there were distinct approaches to quantitative forecasting of future demographic and economic activities in geographical space, these were based on simulation models, which could only be implemented using computers. These were first developed in the mid-1950s in North America as part of the first wave of transportation studies, due in part to increasing car ownership and the establishment of the interstate highway system, but they continued and matured and still represent a significant approach to urban policy analysis.

The history of urban informatics is however a little different. What was not expected when computers were first used in policy analysis was that those same computers would eventually become embedded in the very city systems that policy was designed to change. In short, it was never expected that the world would be using computers to understand systems that were composed of those same computers, except perhaps in systems remote from human functioning such as manufacturing. There was some inkling perhaps that this was possible for the study of municipal information systems which were first institutionalised around the time when computer models made their appearance. This represented a kind of embedded infrastructure and in the 1970s and 80s, the notion of the 'wired city' first appeared building on these notions (Dutton, Blumler, and Kraemer, 1987). Only in the last decade however have we become aware that cities are now composed of computers at every level and that to understand how populations are interacting with each other and how we can make cities more efficient and more equitable, we must use these very same computers to make sense of systems that mix computable machines with humankind. Urban informatics in general does not quite embrace the use of computers to enable policy analysis through modelling and simulation for it is more geared to making sense of big data and cities more efficient and thus these more traditional activities lie on the edge of the field, although the boundaries are blurred.

There is however a third strand to the current development of urban informatics that does relate to modelling and simulation and this is increasingly being called the *'science of cities'*. This is an even fuzzier term than 'smart cities' or 'big data' but it is gaining some currency from the fact that combined with computation and big data, there is concern that we need much more powerful theories and methods to generate a requisite understanding of cities. This is premised on the development of complexity theory and the notion that cities grow from the bottom up as the product of millions of quasi-independent decisions, yet they hang together in highly ordered ways that tend to defy traditional understanding. The idea that cities change qualitatively as they grow lies at the heart of this science of cities but as yet it has hardly been articulated in any considered way and currently is a mixture of models and theories, some of them computable and policy related, many of them not, that relate back to the foundations of urban economics, regional science, transportation theory, and so on, developed half a century ago (Batty, 2008). To an extent, putting together this science of cities with big data and the idea of the smart city represents a curious juxtaposition of ideas whose unity simply lies in the fact that computation is of their essence.

There are several other themes pertaining to urban informatics that we will address in this paper and it is worth flagging these at this stage. First there is the concern with new varieties of data in cities that hitherto were invisible and implicit or did not exist at all. These pertain to social networks and to social media which relate in general to the proliferation of internet services that are now available for conducting many kinds of human interaction. Big data is as much about services such as Twitter, Facebook and so on as it is about tracking transportation movements or retail purchases using smart cards and smart sensors. A second theme that is currently dominating this debate is the notion that smart cities provide environments for ever greater competition and hence local prosperity for their populations and this ties them very strongly to the development of business. As a rider to this, the largest computer companies such as IBM, Cisco and Siemens, amongst others, are in the business of selling bespoke solutions to those operating large scale facilities in cities such as transport systems, government services and so on. A lot of the hype is based on the momentum created by these companies as Townsend (2013) in his recent book so cogently describes. A third theme relates to ways in which smart cities can integrate diverse services using common keys thus adding value to their products and enhancing competition and prosperity this way. A fourth theme relates to citizen participation and online interactions with public services in this digital world. A fifth theme relates to the rather bizarre notion that there are such things as operating systems for cities just as there are operating systems for computers – computable cities – and these are being proposed as ways of integrating diverse services.

We will review all of these themes in the discussion that follows. We will also note some of the main substantive themes that are currently dominating the structure and form of cities that urban informatics has the potential to relate to. These cover issues such as aging, migration, climate change, energy conservation, housing markets, public health, transportation, regeneration and so on, all of which are related in diverse ways to the idea of the smart city, big data, and the science of cities. To this end, we will first explore the rise of the smart cities movement where the focus is largely on embedding computers into cities to make them more efficient and to deliver services digitally that were once delivered using older technologies. This will take us into questions of big data, how the city is being instrumented which is set against the idea that computable functions should be integrated, joined up. These developments are changing the dynamics of what urbanists who seek to understand and plan the city are interested in and the compression of timescales will in time change our theories of the cities. The real danger in all this is that the city as a social system will be forgotten and the lessons of the last 100 years ignored, and this will be one of our concerns.

We will then examine what a science of the smart city there is so far, first around the theme of urban analytics, then introducing wider notions such as digital participation and social media as new foci generating new data and new communities within cities that are primarily a result of the diffusion of computation into cities and city science. We then shift a little and examine questions of access to data – open data, and then privacy and confidentiality which are perennial issues in this field. There are new digital divides opening up and these we will note. We then explore how urban informatics is being driven by the unholy liaison between the ICT industry and city governments and point to examples of smart cities in practice. All of these provide a context for discussing how urban informatics relates to existing and new sciences of the city, where we note the fact that the smart cities movement is largely ignorant of this wider science, of urbanism in general, and of how computers have been used to think about cities since their deployment in the mid 20th century. We conclude by noting key challenges for the social sciences, mapping these ideas onto the key problems of our times: aging, migration, housing, economic growth, segregation and polarisation etc. illustrating how ESRC might respond to some of these through their emerging cities agenda, thus providing a forum for debate rather than, as yet, a fully worked out agenda.

The Smart Cities Movement

Instrumenting the City

The embedding of computers into cities to control physical systems has happened suddenly. For many years, supply chains have been gradually automated and we have become accustomed to logistics of all kinds working fairly seamlessly to provide just-in-time deliveries that sustain us. Services too that are largely based on information have been delivered almost since the inception of the world wide web (which brought the internet to the masses) in the mid 1990s. But only in the last 5, perhaps 10, years have 'digital' sensors been embedded into the physical infrastructure which provide data for the means to control energy-based systems like traffic but also demand-based systems such as travel, retailing, financial, and other services. Sensors in the physical built environment and those associated with people through portable devices such as smart cards and phones are suddenly complementing one another. There is little doubt that the speeds at which these devices operate are such that routine functions are being massively improved in terms of their delivery and control (Bettencourt, 2013). These developments have been accompanied by a new 'computerese' based on terms like 'big,' 'smart', 'open' and so on which are used to describe everything from search to storage, access to information to software and thence to data. The hype suggests that if the large

ICT companies have their way, it will be a very short time before everything will be connected to everything else in an ‘internet of things’. Several commentators and inventors are now suggesting such things as ‘operating systems for cities’, far fetched as they may seem, but connectivity based on access to services anytime, anywhere is clearly one of the main driving forces in cities.

There is a serious side to these developments in that networked systems in cities do provide access to a wide range of services where information rather than the physicality of the service is important. Location-based services were in the vanguard of this movement a decade or more ago and the first smart cities were in fact composed of largely public services offered across the web (Batty, 1997). Since then, it is companies that are dealing with networks like Cisco and electric grids like Siemens that are making the running with the software companies like IBM providing the intelligence to make all this work. So far not so much of this networked infrastructure has been connected up and the focus has mainly been on systems that supply physical movements for which information is essential. Energy systems are lagging far behind and there is probably much more progress in bespoke systems such as online banking and retailing than in some of the traditional nuts and bolts of infrastructure that a city is composed of

The big quest in all this is joining up or integration. The battle cry of the smart cities movement is to ‘add value’ by merging data sets, particularly by merging big data sets which are believed to contain the mysteries of the city, and in so doing provide a) new insights into the way cities and their population function b) new ways of managing cities and making them more efficient, but more to the point, c) providing new opportunities for business and growth by combining different data and services in new ways. One cannot underestimate the smart cities movement in focussing its mission on new commercial opportunities and it is no accident that the UK government through its Technology Strategy Board (TSB) sees this as a way of making the economy more competitive through new kinds of software services.

The problem in all this of course is that our attempts to build big systems – and in general cities are big systems in terms of ICT – have not been good. In fact the general experience with big ICT in the public sector where the profit motive is not uppermost has been poor. It is very hard to find good examples where databases have been joined up and integrated to add such value, largely because the common keys that enable such integration are usually absent. Data is simply not available to enable the stitching. Take an example. Transport for London (TfL) have massive data on demand for public transport through their Oyster card data where they know where every traveller taps in (and out) at whatever place and whatever time. They know from their APIs where any train or bus is at anyplace and any time. Can they connect them up? No, it is impossible. They do not have any data on what passenger gets onto what train and it is not really conceivable that they will ever get such data for it assumes that the passengers are monitored all the time. This then is likely to be a major but typical problem in automating the city and it has barely been broached.

Big Data: A New Focus on Time and Space

A colloquial definition of big data is 'anything that won't fit into an Excel spreadsheet'[4]. Some define it in terms of the so-called 3 Vs – data that has velocity, volume and variety which implies that the data comes in torrents – in streams like water in a fire hose, that has enormous volume relative to what we can easily handle, and that has great variety although this latter descriptor is hard to pin down. Sometimes a fourth V is added – veracity which means that the data is hard to validate in some way.

Big data is not necessarily any better than little or small data but it is different in many ways. For example, Population Census data that is based on 100 percent enumeration of the population (or as good as) is still the gold standard for finding out what is happening in cities but the problem here is that the data is collected so infrequently (every 10 years) that it misses much of the action that we need to deal with. New and bigger data sets at the level of individuals and at finer time scales is becoming available but the problem often is that this is not comprehensive and that it is hard to relate to other data in that it is collected by those interested in much more partial and specific domains in cities. In the world of cities, the biggest data sets that I have come across so far tend to be related to transport which is sensed continuously and which then produces data on the number of objects that can be in the order of hundreds of millions – persons and/or cars say – but collected second by second and at very fine spatial resolutions. For example the Oyster card RFID smart card data that Transport for London have provided us with has about 900 million records (tap-ins/tap-outs over a period of about 6 months. It is easy to see how this kind of data can explode and in time as long as it keeps being collected, it will lead to enormous data sets, quite unprecedented in our world of cities.

The rise of big data has led to some wild speculations. Anderson (2007), the editor of **Wired** magazine, in an article entitled "The end of theory: Will the data deluge make the scientific method obsolete?" to which he answers 'yes', argues that our search for patterns in data will make theory obsolete. He says "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all." That is a pretty dramatic statement and if we did not have any We thus have a challenge of education as much as one of experience.

New Perceptions of What Cities Are

Many of these new approaches to both thinking about smart cities and enabling new technologies to make them more efficient imply very different concerns from those that have traditionally been ascribed to urbanists, architects, planners and those concerned with urban life. In short, these new technologies are changing our focus on the city from that based largely on space to one based on time. In the past, our concern for cities has been on relatively long time scales, as much because what happens in a city on a daily or hourly basis has been beyond our systematic observation and control. Moreover, data and information in any considered way has not been available and hence our understanding of cities on very short time scales has been rudimentary, individualistic and largely subjective. Instrumenting the city is providing systems of control, systems to deliver services and big data on a second by second, locationally precise basis, and this is shortening the timescales for which we have data and information about the city quite radically. In the past, the focus of planning in an institutionalised sense was in terms of years and decades. Now the concern is with the next 5 minutes, 5 hours, 5 days, … and this is changing our perceptions of what cities are and what they are for.

There is a tendency for those involved in smart cities who do not have a background in urban thinking and policy to see the city more in terms of technology and engineering than in terms of social structure. There is an emerging mismatch between urban technologists who see the city in terms of the delivery of services from transport to water to electricity, and those concerned with questions of social polarisation, the location of land uses, public participation, in short the age-old questions that many constituencies involved in cities address. Moreover the notion that urban problems are simple to solve should by now have been dispelled for the experience in everything from garden cities to green belts, from the provision of public housing to the provision of transport systems over the last 50 to 100 years, has been salutary and sobering. Problems in cities are 'wicked' in the terminology of Rittel and Webber (1973) in that they are more likely to get worse than better if you attempt to address them in directly obvious ways which seek simple solutions. The smart city movement has to yet address this question (Townsend, 2013).

The irony of the current obsession with smart cities largely motivated by the biggest computer companies developing new markets for their software services, is that because of the focus on technology, there is a strong push towards developing a science which legitimises these endeavours. In fact, the tenor of the smart cities movement is rather different – more pragmatic, less abstract and more about engineering the built environment, than the sciences which have been developing over the last 50 years or more which are strongly social and economic based. We will return below to this 'science of cities' which should be and in fact is a major concern of our Research Councils but it is worth noting that at present, there are enormous inconsistencies between this science and the pragmatic and mainly business oriented message of the smart cities movement.

Urban Analytics

In fact, for more than 50 years there has been a rudimentary science of the smart city built around the scheduling of emergency services such as police, fire and ambulance. In the 1960s with a concern for developing efficient responses to urgent incidents, operations research techniques began to be applied to the services that were charged with such response (Larson and Odoni, 1981). In particular ways of scheduling these services based on location-allocation models in the widest sense began to be developed and these applications have continued ever since. These models are based on very well defined responses which although involving highly risky and occasionally extreme events are in another sense routine.

Since the rise of big data and routine sensing, these kinds of models are being extended and to an extent are merging with those that deal with the longer term location of activities and movement, such as standard transportation model analyses, and other varieties of location-based services. In particular, urban analytics is focussed on understanding patterns in big data and in this sense, a good deal of its focus lies on new ways of data mining, although to date, the discovery of new insights into the way cities function has been minimal. What is clear however is that the focus on data is revealing enormous heterogeneity in cities, considerably greater than was ever perceived in the past and in this respect, the standard model of urban structure in which activities are clustered and segregated into specific spatial locations is continually being undermined. In a sense, this is no more or less than the notion that cities are places of incredible diversity at their most individualistic level but that patterns and structures that appear more homogenous arise from this heterogeneity in ways that are puzzling and complex. Urban analytics thus provides the essence of analyses that reinforce the view that cities are complex systems *par excellence*.

To an extent, urban analytics and urban operations research are currently expanding at a rapid rate due to the application of various computational techniques, some of which were developed a decade or more ago for supply chain analysis. These kinds of logistics have been widely applied to automate supply chains – which incidentally tend to be more routine and well behaved than people movements in cities – but the logic of these methods is being developed now for a range of location-based services. Urban analytics is slowly extending to the analysis and modelling of public services delivery although in all these domains, efficiency is not necessarily the prime driver of these systems and progress in this domain will clearly be slower. Some of the software involved which is based on data mining is being adapted to the locational realm and in this sense, this domain is the closest in the smart cities movement to the development and applications of geospatial modelling and GIS (geographic information systems) (Murray and Grubesic, 2007).

The New Actors: Software, ICT, Company Towns

Before we change tack and talk about digital participation and citizen science, it is worth noting that the actors and agencies in urban informatics are very different from those that have been traditionally associated with understanding and planning the city. In particular the largest ICT companies dealing with software services such as Microsoft and IBM, communications and routing such as Cisco Systems, and energy transmission such as GE (General Electric) and Siemens have all recently established divisions which roll out their products to the city in general, and to city municipalities in particular. This is their new frontier: the provision of bespoke solutions for cities involving energy, communications, services and the data that supports them. The idea of an operating system for the whole city has also been raised and there are even companies that are floating the idea although one has to take all this with a pinch of salt. One needs to peel back the hype before a clear picture emerges. Here we cannot attempt more than a cursory glimpse of this domain. There are literally hundreds of smart city web sites and initiatives, at least one for every significant city on the planet but the fact that so little has yet been accomplished and so much discussed means that the vision of the smart city is more myth than reality at present. But things are happening and it is thus important to get some sense of the momentum.

The large ICT companies are demonstrating their wares in what are the modern equivalents of the 20th century new towns which in turn were based on the idea of the garden city (http://en.wikipedia.org/wiki/Smart_City). These new towns of which the most prominent are Songdu in the Seoul metropolitan area, Masdar in the United Arab Emirates, Paredes near Porto in Portugal with Santander and Barcelona touted as being the best examples of smart cities based on existing towns. Songdu is funded largely by Cisco in terms of its connectivity with many homes and offices wired for the delivery of services that enable transport and information to be delivered immediately and seamlessly to the residents and workers. Although the connectivity is impressive, it is not possible to say how this is being used and doubtless the users will adapt these systems to their own requirements. Townsend (2013) gives a pretty damning account of the experience to date and notes that there has been much backtracking by Cisco and others with respect to the notion that this will be a model for the 'future city'.

Masdar in the UAE, near Abu Dhabi, is designed to be self-contained in terms of the generation of energy from solar power. Paredes is more like Songdu and the company Living Planit originally designed its idea of an operating system for the smart city in this context. This company has proposed the "Urban Operating System' (UOS™) which provides the essential platform for Machine to Machine Communication (the "M2M market")" which they argue will generate a market of US$1.2 trillion by 2020 (http://www.living-planit.com/ . This seems far fetched and it is not clear at all what is to be connected up. Smart Santander on the other hand is a consortium which consists of companies, universities, and research institutes from different sites in Europe and Australia and the local government of the area of Santander. It is an EU project that is experimenting with innovative " … research and experimentation of architectures, key enabling technologies, services and applications for the Internet of Things in the context of a city" (http://www.smartsantander.eu/). It is early days yet and it is difficult to know what has been done but the intention is to build such connectivity from the bottom up. Telecoms like Telefonica are major players. In fact there seems to be a particular fascination with smart cities in the Iberian peninsula with Barcelona being one of the most progressive in terms of its focus on big data, smart meters, and automated transport (http://cityclimateleadershipawards.com/barcelona-barcelona-smart-city/).

Digital Participation and Social Media

An instrumented city with many social interactions being based on, or at least complemented by flows of information has the potential to engage its citizenry in many different ways. To an extent, these flows of information now exist on very short time scales and at relatively precise locations concerned with information that has traditionally been imparted to citizens through personalised transactions. Public participation prior to the digital age at least in planning the city was highly formalised at different stages of the plan-making process which was conducted over time scales in terms of months and years rather than interactively in terms of days and weeks that are now possible using web-based delivery of information. All of this is part of what in general we might term electronic service delivery which can now be of any form, in a world where citizens are empowered with devices that enable them to pick up information and act on it anywhere, at any time.

Service Delivery

The very earliest applications of web-based technology were focussed on service delivery of routine services ranging from information about transport and public facilities to simple environmental information and out of this milieu came online mapping. It is an open question as to whether these services constitute the essence of the smart city but in principle they are different from the large data sets now being captured by various urban sensors which in turn enable the services that are being sensed to be delivered more efficiently. In this, online mapping and open mapping is crucial as is the entire constellation of geo-positioning technologies. So far however, most of this service delivery is focussed on passive data rather than interactive although with the advent of Web 2.0 technologies which encompass interactivity, users are able to write as well as read the information that is the subject of the delivery.

Public participation in the traditional sense however has not really been massively augmented by these developments. There is little doubt that the potential now exists for various kinds of interactivity and online planning but so far, the ability to mobilise is still limited by the ability to actually bring people together even in a web-based context. Moreover there is no real sense as yet about the extent to which these new online services are changing practice although with the existence of email, Skype and related interactivity, there are likely to be profound shifts taking place. How these are affecting the structure of activities in the city is an enormous question. We see bookshops and related retail facilities disappearing and changing due to patterns of demand that take place online and we see transport interactions being substituted and complemented by online activities. But the bigger picture of what all this is doing to the city is extremely unclear, largely, we think, because so much of this activity is bottom up, inspired by individual actions, and hence intrinsically and necessarily uncoordinated.

Citizen Science

In fact the web, which we are using here as a shorthand for any kind of online interactivity, is now enabling its users to create data as well as engage in its manipulation. The idea of crowd-sourcing although in principle not directly based on a digital online world, has been massively enhanced by the growth of individual interactivity through the web. Again much of this activity is in near real time although some of the more impressive crowd-sourcing activities are geared to solving problems collectively in an online manner over more protracted periods than minutes or hours. The ability to collect data systematically by broadcasting questions across the web and then soliciting responses is in its infancy but there have been some impressive developments of interesting data sets so far. Linking these to locations within the city is one important development where responses to broadcast questions provide information about their location (Hudson-Smith, Batty, Crooks, and Milton, 2009; see http://www.surveymapper.com). One problem is always the representative nature of these responses because like social media data, there is no control over who might answer such questions and it is even difficult to filter the data once it has been collected, given that some of the key structuring variables are often missing. In fact, it is in the discovery of new ideas that crowd-sourcing has produced some really impressive answers to well defined questions for which the power of the crowd and its ability to think laterally have made enormous contributions (Nielsen, 2011).

Hackers and The Grassroots

We could elaborate in considerable detail the way individuals are creating software through small programs called Apps and disseminating these either in free or open source form or for profit. Invariably the model for these Apps is a light version which is free and a heavier version which costs the user, is true for many web services. The same comments about their impact on the city as that pertaining to all these kinds of service delivery apply. It is hard to figure out how these are changing people's behaviours in the city and there are enormous social science challenges involved in developing such assessments. These may have profound effects, or not, on urban structure. It is clear for example that knowledge of prices for example for real estate in the city are now widely informed by online web sites and this must be altering demand and supply. But in terms of Apps which deliver small software programs that are often mobile, the impact of these on behaviours is extremely difficult to figure out for the same reasons already stated. We have not been able to stand back as yet and look at their impact in a considered way but they are changing the face of the city. Townsend (2013) implies that these kinds of development are as important to the idea of the smart city as the bigger software protocols and applications being developed by the major computer and communications companies and that these are much more people-centric in their impact and relevance.

Social Media: Social Networks

Our last foray into the way populations are responding to the smart city involves the development of social networks which bring individuals together across the net. The web-based search engines such as Google were in the vanguard of these developments but more specific developments grew up around them relating to interactions through web pages such as Facebook, short text messaging such as

Twitter and related sites that enable users to log in and share their data and personal details with one another. As such media are online, then it is possible to collect this data in real time and to archive it for analysis of interaction patterns which pertain to a variety of social networks. Anything that can be tagged in space and time can be examined for patterns of location and their dynamics but constructing networks from such data is quite problematic. Moreover the variety of purposes of such interaction tends to be difficult to unravel. At first sight, much social media lacks structure in that we are recording raw data feeds without supposing any preconceived structure on the data. This is the problem of big data in general where there are often no guidelines to what to look for in data that is not structured to any particular purpose when it is collected. This has very different potential from data pertaining to censuses of transport and physical movement although it is not dissimilar from email traffic. It is very different from accessing web pages which are highly structured in the first instance but in any event, these are new sources of data that in principle have the potential to provide new perspectives on the functioning of the computable city. So far although there is considerable interest in social media data, there have not been many particular insights and what has been produced in locational terms seems to mirror existing urban structures as we know them.

Data, Privacy and Confidentiality

Open data

In parallel to the development of smart cities, the idea of open data has come swiftly onto the agenda. Open data goes back to the mid 1950s in terms of access to scientific data but in terms of public data, it was propelled by an agreement in 2004 by the OECD's Council of Ministers that all publically funded archival data should be opened up and made free to a wider citizenry. In particular, this has led to the open data movement, like open source in software and open access in scientific journals, and it has been particularly driven by the US and UK governments through their www.data.gov and www.data.gov.uk initiatives. In the UK, this has led to quite widespread release of data in various public archives but the real power of the movement has been the government's transparency agenda which is a good deal wider than data and the acceptance by many public authorities that they should make their data open in many ways. This of course has been made possible by the move to entirely digital data representations and it cannot be separated from digitisation in general. In the UK, this has been also spurred on by the formation of the Open Data Institute (www.theodi.org) which seeks to popularise and initiate the wider move to open data as well as providing leverage for opening up data in this way and training for enabling less specialist constituencies to access this data. This supports the idea of digital participation in urban planning and the wider politics of municipal government, and it is influencing the provision of new data centres for government data in particular which the ESRC are currently building so that academics can gain access to various sorts of data pertaining to administration, business and social media.

Big Data, Little Data

We have already introduced the idea of big data but it is important to note that big data is unstructured and often hard to use in comparison with much smaller data sets (which nevertheless might still be inaccessible in terms of processing and analytical understanding). There is also the issue of multiple data sets and whether or not a large number of smaller data sets in concert raise the same issues as big data, particularly when the notion of integration and adding value by merging data sets using common keys is invoked. In fact, it is likely that relatively small structured data sets which are collected with very specific purposes in mind either digitally using sensors or through a combination of digital and manual methods contain fields that enable different data sets to be stitched together quite easily, thus producing much larger data sets. Geospatial data such as Census data in particular which is address-coded can often be used to provide a basis for matching to many other data sets and in this way can provide data which begins to meet the tenets of the methods needed to explore big data through new techniques of data mining. In fact, in terms of the social domain, point of sale data related to geo-demographics, house prices data, employment types by size and location and their attributes such as wages and such like all have incredible value in understanding the city. Increasingly it is these data sets that are traditionally small but becoming bigger that will constitute the focus of urban informatics. The frontier in this kind of research will certainly lie at the interface between well-sourced traditional data such as the Population Census, which is increasingly being automated in terms of its collection, and big data pertaining to the sorts of streamed data that is coming from transport demand (such as that from RFID cards). Merging this together and then adding attributes from geo-demographic data sets collected by private agencies is in principle possible but the real hurdles lie in ownerships as well as privacy concerns rather than in the technical feats of actually producing such integrated data. These are very much questions that pertain to the social context, not the technology itself.

Privacy and Confidentiality

Privacy and confidentiality is an enormous area that dominates any discussion of the digital society and all the fears and concerns of identifying individual personages and invading their privacy through data that is being collected as part of instrumenting the city are central to this whole debate. It is worth stating that a lot of instrumented data such as that associated with RFID cards and embedded sensors is intrinsically non-personalised and only when individuals begin to add their own attributes to the data – which they may well do where registration is available (or required) – is there the possibility of invasion of privacy. An example is in order. The data we have from Transport for London which is all tap-in and tap-outs using Oyster cards which we have for 6 months in 2012 simply gives us position and time of access and the status of the card used (child, free card and so on). Not much can be added to this data and it is hard to see how any individual trip could be sourced to an individual in terms of more personal attributes. It is not impossible but the keys to data merger are not really there. In fact it is not even possible to tie a traveller to a particular train or bus (as we also know where these are in time and space from a quite independent data set collected by TfL), and thus even the operators of the system cannot tie demand to supply.

However, it is reported that various companies have data sets on individuals that suggest that if one knows address, age and sex, then in the US 85% of all individuals can be identified in terms of much more personal characteristics. These kinds of issue abound and will increasingly be of concern in terms of the data being generated through many kinds of digital record which are increasingly part and parcel of post-industrial society. There is an urgent need for considered research with respect to big data of this kind, how it is being merged with other data of a more specific variety, and what the limits should be on the procedures for adding value to data in this way. The current debate about what government agencies (e.g. the NSA, GCHQ etc.) have on electronic mail and related data as well as general surveillance in the city through cameras and other sensors illustrates how important these issues are and are likely to become (Baumann and Lyon, 2012).

The New Digital Divide

30 years ago when the personal computer began to proliferate, there was an important concern that the poor and generally disadvantaged would not be able to access such technologies thereby excluding them from learning anything at all about the digital world. To an extent this has disappeared or rather these divides which can be spatial with respect to who has access to what, have become more muted but now the concern is that several groups in many different kinds of society – but mainly the poor and the elderly – still do not have access to the internet. This is particularly severe with respect to online banking, the delivery of public services such as welfare benefits and such like, and there is little doubt that these divides are of great concern to the functioning of cities on all levels. New patterns of segregation are appearing. But of even greater concern are the divides that are taking place with respect to the access to new technologies that enable people to be online continuously. Smart phones, for example, are almost a prerequisite now if one's business is dominated by email traffic but more than this, access to information across the web is instantly available anytime, anywhere. This requires users to be not only internet and smart phone savvy but also to be able to operate the devices – and this tends to be more difficult the older the user. Moreover, new forms of information system that are essential to

everyday life such as social network sites like Facebook are no longer just based on friendship links. There are many services that are not accessible unless one is signed up to such media. New and subtle information/digital divides are appearing and these threaten to dominate discussion of the smart city, derailing perhaps the move by the large ICT companies to automate and instrument everything in sight. The key issue then is not whether such instrumentation and information is available, but who uses it and how? These are deep questions that relate not to technology *per se* but to social life in the city.

A Science of Cities

Fifty years ago if you had asked the question "what can we do with computers with respect to cities?" the answer would have been we can build computer models of cities – abstractions – that can then be used to pose conditional questions such as 'What If …..' The predictions from such models would then inform planners and policy makers about the future. In fact, the message at the very beginning of this transition was even stronger in that people believed that model-based predictions were much firmer and certain than ever we would expect today. In fact, there has been a retreat from this hard science to the position where most of us do not consider we can predict anything but the shortest term futures (and even this is debatable).

The first change in focus with respect to computers and cities came very early too with the development of graphics. The notion of a model of the city was largely symbolic and mathematical in the early days but graphics pushed the field into thinking about how one might generate digital versions of the architects' analogue models and with the advent of computer-aided design in the early 1980s – as much on the PC as on any other machine – 3-D geometric models became possible. Through the 1980s and 1990s, these models were massively improved and even began to merge with GIS in the late 1990s. With the advent of Google Earth, they have become routine and this arsenal of digital modelling tools for cities now embraces the symbolic and the iconic as well as various representations that have ported these models to a variety of devices and environments ranging from hand held to virtual realities.

With the advent of mobile phones in the late 1980s and with the roll out of optical fibre networks – wide area networks, for the first time the notion that computers might be embedded into the city's fabric became significant (Batty, 1989). Of course computers had been embedded in cities from the beginning but they only became visible outside the enterprises and services that powered the city, when they moved into more public realms. With the advent of the web, all kinds of devices were linked to the internet and for the first time the prospect of smart cities based on the wired city metaphor appeared possible. These early developments were and continue to be largely invisible to direct observation and scrutiny. This is one of the most significant differences between the industrial and post-industrial city in that electronic and digital flows are replacing and complementing material flows (Batty, 1990, 1997). These tend to have a degree of invisibility that remains problematic and it is almost as if cities have become more complex due to this invisibility. In fact big data does not help much with this problem. Much of business and a lot of public administration is now done by email and considering the sheer volume of such traffic, there is very, very little knowledge of what this all means spatially and temporally to the economy and the city.

The notion that cities are composed of the very instruments that we use to understand them poses a conundrum for urban theory. So far there have been few extensions of our symbolic models to the new world of the smart city. Our models are still built on old theories of social physics and urban and regional economics, the physical transportation of materials and people, urban housing markets of a kind that now seem archaic compared with the way global flows of capital are distorting markets, and so on. In fact the current concern with these theories and models is more to making them accessible as web services in and of themselves rather than extending them to deal with electronic flows such as email, online retailing and commercial services etc. Parallel to this is the emergence of new views of cities that have little or nothing to do with the smart city *per se* although they are founded on the notion that ICT is essential to the future economy of cities. Coming from complexity theory and old ideas about agglomeration, as well as from the view that big cities are no longer 'evil' as they were seen in the 19th and much of the 29th centuries, there are new theories emerging that deal with the power of networks to distribute energy and information around in economical ways in cities that drive the forces of creativity and agglomeration. Networks are of course central in all of this and the smart city is built around networks as much as it is around locations. In fact this is the message of commentators such as Glaeser (2012), and Jacobs (1961) before: first and foremost cities are places where people come together, where people connect. The best examples of this can be seen in the work of the Santa Fe group (Bettencourt, 2013) and perhaps my own (Batty, 2013)

There is thus emerging an unholy liaison between smart cities, big data and a new science of cities that reinforce one another only partially but are all themes that are changing the nature of the way we think about cities. This is partly a marriage of convenience as new actors and voices come onto the scene (as Townsend implies in the quote I used in the summary to this paper. It is partly due to the fact that we are groping for new theories of the city and these are likely to be founded on new data sources that have changed our perspectives immeasurably on what we might be able to understand. In **Physics Today**, there is an interesting report by Kramer (2013) which quotes Steve Koonin, the Director of CUSP, the new smart cities centre in New York City. He argues that new skills are needed to look at cities in the way we have portrayed here and he singles out physics. He says: "The kind of skills physicists bring to thinking through complicated situations, data driven and so on, are not all that common in urban science and technology at this point. Physicists have a lot to bring to the table here."

Smart Cities: Challenges for the Social Sciences

We cannot produce a definitive list of research topics here for the purpose of this paper is to raise the debate and it is more meant to point to matters for extensive discussion. We need to relate ideas about urban informatics and big data which we have discussed under the rubric of the smart city to other papers meant to inform the discussion, in particular Urban Society, Governance-Government, Urban Resilience, Economics and Finance, Civil Society, Inequality and Exclusion, and Environment and Sustainability.

A good paradigm in the social sciences under which to discuss challenges that emerge from our domain here is based on Castell's (1989) idea of the 'space of flows' whose Wikepedia entry defines the idea as " … a high-level cultural abstraction of space and time, and their dynamic interactions with digital age society" (http://en.wikipedia.org/wiki/Space_of_flows). This is a concise way of thinking about the change in emphasis from industrial to postindustrial cities where the focus now is on how movement is being facilitated by information flows that complement and/or replace and/or add to traditional material flows involving people and energy. It represents what Negroponte (1995) called the transition from 'atoms' to 'bits'. Much of what we discussed in the early part of this report dealt with how cities are being instrumented to control such flows and how online services articulate these interactions. Moreover this pertains to the development of a new science of cities based on flows rather than locations and the unholy liaison between the smart city and this new science noted above provides a coincidence of effort and interest that is significant to any research agenda (Moses, 2013)

We will divide the challenges into five sets of ideas involving theories about the space of flows and the new mobility, changing conceptions of urban planning, data infrastructure which involves big data, integrating data, relationships to traditional media and open data, instrumented data and social media, digital participation and community, the new urban geography of the smart city, privacy, confidentiality, and surveillance, and governance. There is also a research agenda with respect to the physical and engineering sciences and it is essential to be aware of the fact that a lot of work on smart cities, urban informatics and big data is being pursued under these different disciplinary perspectives. Indeed the agendas that involve big data, complexity and energy are quite widely developed elsewhere and any effort by ESRC for example in these areas needs to take all this into account. Moreover the large ICT companies and the UK Government's Technology Strategy Board have very specific agendas in this domain and we will note these by way of conclusion.

- The Space of Flows and the New Mobility: there is an urgent need for new theories of the smart city that replace the locational paradigm which served the idea of the industrial city rather well in the past with notions about flows and mobility, particularly about digital flows, flows of information as they link and dovetail with physical flows and people flows. This is a deep and broad agenda and it is not possible to specify this in any detail here but it is the backcloth against which ideas about urban informatics and big data must be seen. It is essential to see the field in the perspective of social and economic issues rather than simply of technological and computer engineering. In fact it is of utmost importance in setting the smart cities agenda in the wider social context without which we are destined to repeat the physicalist mistakes in the past planning of our cities

- Changing Conceptions of Urban Planning: urban informatics shortens the time horizons of understanding and planning cities quite dramatically. Most of our prior understanding has been for the relative long term but we now have a much more immediate sense of how cities are changing and this is leading to new ideas about resilience, disruption, extreme events and a rapidly changing dynamics that can only be observed and understood using new data sources. Urban planning as established in practice as well as the public domains that support it, needs to adapt to these new prerogatives.

- Data Infrastructure: Big Data, Data Integration, Traditional Media, and Open Data: we require a new grasp of what big data streamed from many sensors either embedded in the physical city or attached to individuals or other mobile objects are able to bring to our understanding of the city. We need to contrast this understanding with the types of control that are being instituted in the city by a diverse collection of operators and users and we need to look at the limits of such control. We need to grapple with the extent to which we can add value to data by fusion and integration and we need to relate new physical to social data and new big data from sensors to traditional data which is still the gold standard of our knowledge about the city. We need to have some conscious view of how good and how accessible private data sources are, coming from commerce and business as well as private data from companies devoted to collecting and selling such data. In particular, price and income data as well as other economic flows are important in this context. We need to explore the extent to which open data is useful for understanding the city and the extent to which it is truly available for cultivating new businesses and new ideas about the city. Last but not least, the smart city gives us the opportunity to begin to think about new data sources from crowd-sourcing and it is important to gauge what is possible here and how such crowd-sourced data is integrated with more conventional sources.

- Instrumented Data and Social Media: a lot of new data is being generated either from instruments embedded in the environment or attached to objects that are fixed or mobile including persons. There are a variety of new networks that are being formed spontaneously for social and commercial data transmission and we need to make sense of how these new networks are linking to existing ones and possibly replacing them. In short we need a new social network science and we need to be very clear about the representativeness of the data that is being collected. We need targeted studies of these new data sources, how they are generated, how good they are in terms of their representativeness and comprehensiveness. The question of the meaning and semantics in all these new media is an urgent issue as much of it seems largely irrelevant to our knowledge of the functioning of the city. If we cannot interpret it, we have no chance of saying any meaningful about its impact. Moreover the construction of networks from this is difficult as most digital flow data tends to relate to where one is located and when, not to whom one relates. Finally the quality of this new data is a major issue. It is clear that much of it is unstructured and there are many gaps in it in terms of its comprehensiveness. This goes for even very well defined data sets from RFID cards for there any many instances where such cards do not function due to sensors being switched off due to human or technical error. The long lasting nature of these data is also an issue of relevance for as the technology changes so does the data and this means that its archival integrity can never be assured.

- Digital Participation and Community: in principle the smart city should be able to deliver services – passive and interactive – easily and efficiently notwithstanding any digital devices that affect access. As yet apart from the delivery of routine public services and of course massive penetration of online retailing, most public services are not automated. The singly biggest provider of such services such as the NHS in the UK for example has relatively primitive online access at least in terms of patient interaction. Public participation in planning too is primitive in terms of positive actions even though there are strict forums for such engagement. As yet there has been little interactivity with respect to problem solving and plan making and much interaction remains at a direct personal level. This is perhaps surprising seeing that there is a long history of community action in urban planning in many western countries but its automation is difficult and progress has been slow.

- The New Urban Geography of the Smart City: there is little doubt that automation and instrumentation of retailing, transport, health, house buying and a variety of other traditional spatial behaviours is changing the way in which the city is structured spatially. We do not know quite what this means except that life in large cities seems faster than in the past and the speed of change is greater. This means we urgently need online data systems which record traditional location and movement data at a finer temporal scales, yearly rather than ten yearly or even finer. In time this may occur as big data sensed in real time and space is archived over longer periods but all this is necessary to figure out new patterns of segregation, new digital divides, new areas of deprivation as well as the extent to which populations are being driven into different locations by the new economics of the smart city. A lot of this relates to globalisation as well and of course it all depends on the space of flows.

- Privacy, Confidentiality, and Surveillance: This is a massive area. We need to figure out the extent to which the data sets that are now coming on stream change the nature of privacy and confidentiality if they are merged with other data. Adding value is hard through data fusion but not impossible and there are many *ad hoc* instances of where privacy can be invaded. New studies of these with respect to online data are urgently required. Moreover some sort of catalogue of the extent to which people are being surveyed in the city is required so that we can get a sense of where and when such surveillance takes place. This is wider than urban informatics *per se* but it is certainly part of the agenda.

- Governance: this too is a wide topic. One of the key issues is how urban planning needs to fit into this new conception of the city and how other public and private services inform questions of equity and efficiency. To an extent this is part of parcel of how the city needs to integrate its management and control functions and to engage its citizenry and this is an issue that clearly pertains to other aspects of the ESRC Cities agenda.

Postcript: Urban Informatics, Big Data and Other Research Agencies

We need to note that this area is highly interdisciplinary in that there are features of the smart cities problem that are physical, engineering as well a social in tenor. These are also strongly methods-based and thus a wide combination of disciplines are needed in researching these matters. This is already recognised in that the Engineering and Physical Sciences Research Council (EPSRC) have as much a stake in these issues as does ESRC and in the last ten years there have been several initiatives that cut across the research councils in transport, complexity, planning urban design as well as in social analysis. Big data typically is part of traditional science and computer science but new methods in statistical and in economic theory, and large-scale social data are the domains of ESRC.

The other features that complicates these questions is the focus of the Technology Strategy Board which has launched its Catapult projects in Cities and in Transport with the catapult in London (https://futurecities.catapult.org.uk/) and the demonstrator projects in Glasgow (http://futurecity.glasgow.gov.uk/) significant to any research agenda involving research into cities. The mission of the TSB appease to be to generate new businesses in providing software for smart cities as well as big data analysis in the geospatial domain. The role of a science of cities in this is not yet very clear and urgently needs to be clarified. The Government Office for Science has a new Future Cities Foresight (http://www.bis.gov.uk/foresight/our-work/projects/current-projects/future-of-cities) project that is relevant. How all these fit together is something that the ESRC Cities agenda should consider, as in the last analysis, the problems of cities are social and economic rather than engineering based.

Bibliography

Anderson, C. (2007) The End of Theory: Will the Data Deluge Make the Scientific Method Obsolete? **Wired Magazine**, July 16, accessed 13/10/2013 available at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

Batty, M. (1989) Technology Highs, **The Guardian Newspaper**, Thursday 22 June, p. 29

Batty, M. (1990) Invisible Cities, **Environment and Planning B**, **17**, 127-130.

Batty, M., (1997) The Computable City, **International Planning Studies**, **2**, 155-173.

Batty, M. (2008) Fifty Years of Urban Modelling: Macro Statics to Micro Dynamics, in S. Albeverio, D. Andrey, P. Giordano, and A. Vancheri (Editors) **The Dynamics of Complex Urban Systems: An Interdisciplinary Approach**, Physica-Verlag, Heidelberg, DE, 1-20.

Batty, M. (2013) **The New Science of Cities**, The MIT Press, Cambridge, MA.

Batty, M., Axhausen, K., Fosca, G., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., and Portugali, Y. (2012) Smart Cities of the Future, **European Physical Journal Special Topics**, **214**, 481–518.

Bauman, Z., and Lyon, D. (2012) **Liquid Surveillance: A Conversation**, Polity Press, London.

Bettencourt, L. M. A. (2013) The Uses of Big Data in Cities, SFI Working Paper: 2013-09-029, Santa Fe Institute, NM, available at and accessed 13/10/2013 http://www.santafe.edu/media/workingpapers/13-09-029.pdf

Castells, M. (1989) **The Informational City: Economic Restructuring and Urban Development**, Blackwells, Oxford.

Dutton, W. H., Blumler, J. G., and Kraemer, K. L. (1987) **Wired Cities: Shaping the Future of Communications**, G. K. Hall, New York

Foth, M. (Editor)(2008), **Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City**, Information Science Reference, IGI Global, Hershey, PA, 270-290.

Foth, M., Choi, J. H-J., & Satchell, C. (2011) Urban informatics. In **Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work**, ACM, Hangzhou, China, 1-8, available at http://eprints.qut.edu.au/39159/1/39159.pdf

Glaeser, E. (2012) **Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier**, Penguin Books, New York.

Hudson-Smith, A., Batty, M., Crooks, A., and Milton, R. (2009) Mapping for the Masses: Accessing Web 2.0 Through Crowdsourcing, **Social Science Computer Review**, **27**, 4, 524-538.

Jacobs, J. (1961) **The Death and Life of Great American Cities**, Random House, New York.

Kramer, D. (2013) Smart Cities Will Need Big Data, **Physics Today 66**(9), 19-20.

Larson, R. C., and Odoni, A., R. (1981) **Urban Operations Research: Logistical and Transportation Planning Methods**, Prentice Hall, Englewood Cliffs, NJ.

Moses, M. E. (2013) Information Technology: Slouching Towards Utopia, **Nature 502**, 299–300 (17 October) http://www.nature.com/nature/journal/v502/n7471/full/502299a.html

Murray, A. T., and Grubesic, T. H (Editors)(2007) **Critical Infrastructure: Reliability and Vulnerability**, Springer, Berlin, DE.

Negroponte, N. (1995) **Being Digital**, A. A. Knopf, New York.

Nielsen, M. (2011) **Reinventing Discovery: The New Era of Networked Science**, Princeton University Press, Princeton, NJ.

Rittel, H. W. J., and Webber, M. M. (1973) Dilemmas in a General Theory of Planning, **Policy Sciences 4**, 155-169.

Townsend, A. (2013) **Smart Cities: Big Data, Civic Hackers and the Quest for a New Utopia**, W.W. Norton & Co, New York.

Appendix: University Centres Dealing with Urban Informatics, Smart Cities, Big Data and Related Urban Science

The process of attempting to identify the key centres world-wide is quite tortuous largely because so few initiatives in this area have any longevity. Most have been started within the last 2 years and do not have anything other than a web presence. In short, the volume of writings about urban informatics is still quite small and not well focussed, and established groups are few and far between (Foth, 2008). This is a time of very rapid change insofar as people who are working on smart cities and a science of cities are concerned and my trawl revealed a hornet's nest of semi-formed and new initiatives with very few having been around for more than 5 years.

My trawl is far from complete but it does identify the main centres which have an established presence; there are not that many. They are groups that began in urban modelling and/or GIS and in this sense are largely tied historically to transport, planning and regional science. We have cast our net as wide as possible but at the onset, we must make very clear distinctions between different types of centre. We define a centre as a cluster of 4 or more significant individuals working in the domain of computer applications to cities that we have defined at the beginning of the report. We can classify these into 1) established centres that are clearly identified with urban informatics and urban science 2) emerging centres in urban informatics 3) GIS labs (geographic information systems) that have a strong urban science component 4) centres focusing on urban modelling and simulation 5) media centres that focus more on visualization, art and social media in the urban realm, 6) computer science labs that focus largely on big data and urban mobility, and 7) complexity centres with a focus on urban science. We will not attempt here to deal with the vast array of groups working with mobility and transport data that come to the field from computer science. In time, we will extend this list to include these, and in this sense, this appendix is a living document.

Our trawl is by no means comprehensive and as this area is rapidly expanding, this is a work in progress. There are many people working in this area but most do not have a background in urban scientific research and insofar as this is a reasonably definitive list, it is only so with respect to the fact that most if not all urban science groups which have track records are identified. Moreover it is largely based on centres which have a web presence in the English language. What we will do is list each centre, indicate its web address, note its key focus, and approximate the date at which it was established. We will not define which type the centre or cluster is because there are many overlaps which makes this tricky but enough information is included here to establish the nature of each group.

Centre	*Web Address*	*History and Current Focus*	*Established Date*
CASA (UCL) Centre for Advanced Spatial Analysis	www.casa.ucl.ac.uk www.blogs.casa.ucl.ac.uk	Established as a GIS centre with strong urban focus, now orientated towards simulation, spatial data and visualization	1996
UCL-Imperial: ICRI	www.cities.io	Intel Collaborative Research Institute between UCL and Imperial – computer science oriented, mobility, big data	2012
Imperial Urban Systems Laboratory	No current one-stop-portal but see www3.imperial.ac.uk/cts www3.imperial.ac.uk/digital-economy-lab/partnernetworks/dce/	Recently established centre largely drawing together Digital Cities Exchange, the Cisco Future Cities Centre, ICRI, NEC smart water lab, and BP urban energy systems. Centre for Transport established 1960s.	1965, 2000, 2012
CUSP (Center for Urban Science and Progress) New York University	www.cusp.nyu.edu http://marroninstitute.nyu.edu/	Established from an initiative by NY City Mayor to establish science campuses in the City; this is one of several initiatives at NYU . The Marron Institute acts as the portal	2011
IIT Bombay – Center for Urban Science & Engineering	http://cuse.iitb.ac.in/?page_id=117	Recently established partner institution with CUSP, focussing on engineering applications of informatics	2013
University of Warwick Urban Studies Initiative	www2.warwick.ac.uk/research/priorities/sustainablecities/ & http://www2.warwick.ac.uk/fac/cross_fac/cim/	Recently established partner institution with CUSP, focussing on social applications of informatics	2012
Cornell Tech Built Environment Hub	http://tech.cornell.edu/	Established, it seems, as part of the NY City Science Campus Initiative, partnered with Technion	2012
Columbia University Center for Smart Cities	http://idse.columbia.edu/smart-cities/	Physics, engineering and data analysis for urban energy, transport and infrastructure	2010
Columbia University Urban Planning Spatial Information Lab	www.spatialinformationdesignlab.org/	Lab supporting visualisation and GIS related to projects in New York City and urban research	2008
Urban Center for Computation & Data (UCCD) U of Chicago	http://urbanccd.org/	Focus on Big Data and Computing spinning off from Argonne National Labs	2010
Future Cities Lab: ETH Zurich Singapore Centre	www.futurecities.ethz.ch/	Part of Architecture and Planning at ETH Zurich, established to pump prime urban planning through National University of Singapore	2008
Senseable Cities Lab Singapore-MIT Program	http://senseable.mit.edu/livesingapore/	Part the SMART Programme between Singapore and MIT through various departments in particular DUSP (Urban Studies and Planning)	2004
A*STAR Singapore Complexity Group	www.ihpc.a-star.edu.sg/	Initiative of the Singapore Government to develop urban science for public agencies based on state of the art research	2009
Singapore U of	http://lkycic.sutd.edu.sg/	A focus on urban innovation and	2009

Technology & Design Lee Kuan Yew Centre for Innovative Cities		future cities with interests in city growth and urban networks	
Amsterdam-Delft Senseable Cities Lab	No web site as yet but see http://amsterdamsmartcity.com/	A proposed spin-off from MIT's Senseable Cities Lab based on various cooperation with Dutch agencies and universities	2014
MIT DUSP – Institute for Advanced Urbanism	http://cau.mit.edu/	Emerging focus on urban issues informed by ICT applications to multiple social and physical urban problems	2012
MIT Media Lab City Science	http://cities.media.mit.edu/	A synthesis of work being done on sensing, participation, visualisation and urban economics through the Media Lab	2012
Harvard Data-Smart City Solutions	http://datasmart.ash.harvard.edu/	A centre dedicated to exploring urban data with respect to government	2013
SMART Infrastructure Facility, U of Wollongong	http://smart.uow.edu.au/	Smart cities focus with respect to energy, transport, infrastructure, wider than cities per se	2009
AURIN Program, Uni Melbourne	http://aurin.org.au/	Spatial data infrastructure and analytical methods for urban and regional research and planning	2009
QUT Urban Informatics Centre	www.urbaninformatics.net/	Urban informatics centre built around visualisation, social media, creative industries and networked participatory interaction	2006
Tel Aviv Uni Geosimulation and Spatial Analysis Lab	http://geosimlab.tau.ac.il/	Complexity modelling of urban systems with a focus on agent based modelling and visualisation	2000
Santa Fe Institute	www.santafe.edu/	The centre that 'invented' complexity theory whose urban science focus is on city size, migration, networks, all pertaining to city systems	1985 urban group 2005
Wuhan LEISMAR	www.lmars.whu.edu.cn/en/List.asp?ID=338	Geomatics, visualisation and big data centre built around virtual reality and 3D modelling and sensing of cites	1990
OII Oxford Internet Institute	www.oii.ox.ac.uk/	Longstanding initiative to explore the socially networked world with a significant urban focus on information geographies	1995
NUI Programmable City Group	http://communications.nuim.ie/240113.shtml	Recently established centre relating to programmable cities relating to the way software is changing urban behaviour.	2013
Eindhoven University Built Environment Group Urban Science and Systems	www.tue.nl/en/	Transport, retailing, and agent-based modelling centre with large scale applications to Dutch planning	1990
VU (Vrie University) Geodan/Spatial Economics Group	www.feweb.vu.nl/gis/spinlab/	Regional science and GIS group with extensive research into smart cities, transport, and spatial visualisation	1990
Planning Support	www.uu.nl/faculty/geosciences/en	Transport and GIS group with	1995

Systems U of Utrecht	/research/institutesandgroups/	strong interests in planning support systems and application of models to plan-making	
Martin Centre, U Cambridge	www.martincentre.arct.cam.ac.uk/	One of the traditional urban modelling groups with focus on urban energy, resources, urban design	1967
OPUS Group U Cal Berkeley	www.urbansim.org/	Open Platform for Urban Simulation developed as part of city planning and the UrbanSim modelling group	2008
CSAP U Leeds	www.geog.leeds.ac.uk/research/csap/	Urban modelling group with strong focus on GIS, ABM, social media and open data	1970, 2005
Spiekerman-Wegener U Dortmund	www.spiekermann-wegener.de/	Long standing urban modelling group affiliated to U Dortmund	1978+
CSIS Centre for Spatial Information Science U Tokyo Urban Engineering	www.csis.u-tokyo.ac.jp/english/	GIS centre for extensive applications in 3D modelling, sensing, spatial analysis and traffic flow modelling	1997
IEIES, Chinese University of Hong Kong	www.iseis.cuhk.edu.hk/eng/	GIS group with strong applications to virtual urban environments and remote sensing	2002
Tsinghua University Spatial Design Lab	http://designbeijinglab.com/	A new lab concentrating on smart cities, design, and urban data	2012
NCG National Centre for Geocomputation National University, Ireland, Maynooth	http://ncg.nuim.ie/	GIS centre established to work on urban and spatial problems but also focussed on multimedia and spatial analysis	2005
Centre for Geoinformatics St Andrews University	www.st-andrews.ac.uk/geoinformatics/	New lab specialising in spatial analysis, GIS and quantitative urban modelling	2011
IMI (The Institut Municipal d'Informàtica) Barcelona	http://guia.bcn.cat/institut-municipal-d-informatica_92086026158.html	A 200 person initiative to support various smart city initiatives in Barcelona funded by the municipality	2009
City Science Group UPM, Madrid	www.citysciences.com/	A smart cities centre at the Polytecnnic University of Madrid building on transport, VR, computing, visualisation and architecture (CeDInt -Centro de Domótica Integral – Research Centre for Smart Buildings and Energy Efficiency)	2012

There are many other groups dealing partially with city science and there is no easy classification of these. Many of them exist in computer science, physics and engineering departments such as those at Cambridge-UK, Pisa, Bologna and so on. I have purposely not included the very significant set of GIS centres in North America, most of which have urban applications, and in our group at UCL, we know these well. A good site listing these centres is the UCGIS site where the focus is more on teaching but this is a good resource, see http://ucgis.org. There are many smart city sites in cities across the world and some of these have research foci. Townsend's (2013) book **Smart Cities** and the paper by Batty et al. (2012) 'Smart Cities of the Future' provide as comprehensive a summary of these kinds of initiative to date in this vast array of examples. There are many smart cities projects in the EU and the following web sites are those which appear most important but these are by no means all.

http://www.smart-cities.eu/
http://www.eu-smartcities.eu/
http://ec.europa.eu/eip/smartcities/
http://setis.ec.europa.eu/implementation/technology-roadmap/european-initiative-on-smart-cities
http://ec.europa.eu/energy/technology/initiatives/smart_cities_en.htm
http://www.smartcitiesineurope.com/
http://www.eera-set.eu/index.php?index=30
http://connectedsmartcities.eu/
http://www.epic-cities.eu/content/smart-cities
http://urbanixd.eu/about/

It is not possible to do justice to this proliferation of groups. I will try and keep this list updated, so that those reading this appendix should contact me at m.batty@ucl.ac.uk to tell me what I should include and what I have left out.

End Notes

[1] Anthony Townsend, personal communication, October 10, 2013

[2] There are many definitions of urban informatics and one which focuses more on small scale computable devices and the media that comes from them – which is somewhat different from our emphasis here is from Foth, Choi, and Satchell (2011) where say: "Urban informatics is the study, design, and practice of urban experiences across different urban contexts that are created by new opportunities of real-time, ubiquitous technology and the augmentation that mediates the physical and digital layers of people networks and urban infrastructures." Their approach is writ large in their work at QUT Urban Informatics http://www.urbaninformatics.net/

[3] The term 'smart' is widely used in North America in a variety of everyday conversations. Its appendage to the word city appears to have come from it generic use to refer to intelligent devices such as smart phones but also from usage with respect to policy such as 'smart growth' which is the term used for policies to contain urban growth and sprawl.

[4] I cannot attribute the source of this definition other than pointing to a post by Ed Ramsden on Linked-In at http://www.linkedin.com/groups/I-usually-think-Big-Data-3866594.S.125341031. There are many other definitions relating to ways of mining/searching/visualizing/analysing the data, all of which relate to the difficulty of doing so due to its size. One suggests that as data is usually part of the solution, any data that is 'big' becomes part of the problem. Moreover, it is a constantly moving target and what is big data yesterday is little data today. Wikipedia has a good entry about it and basically it relates to data that requires specialist skills in its analysis and data whose analysis often involves searching through massive data sets simply to do basic classification. It is often unstructured, and one of the main problems is finding structure in such data sets as they are so large, that even the simplest query operations lead to analytical problems and massive processing time: see http://en.wikipedia.org/wiki/Big_data.