Open data today and tomorrow: the present challenges and possibilities of open data

Dr. Sophia Drakopoulou
Faculty of Arts and Creative Industries, Department of Media,
Middlesex University,
The Burroughs, London NW4 4BT, UK Email: S.Drakopoulou@mdx.ac.uk Website:
www.cybersalon.org

Abstract: This paper argues that because of the present challenges of open data, existing datasets such as governmental open data are limiting the potential possibilities of open data application development. The primary challenges in using open data are the formats in which open data are made available, the digital literacy required to exploit it, and the copyright issues that arise from commercial use. Hackathons provide some creative solutions but are reserved for the 'techno-elite'. In contrast, a new trend is developing in sensor-based purpose-gathered citizen-led open data that can be used to create meaningful interactions with participants and develop open data applications and systems that can serve a particular local area and group of people. This paper draws from the experiences of the members of the London-based collective Cybersalon and the 2014–2015 HyperHabitat series of events, projects, presentations, and hackathons that investigated the changing nature of our living environments.

Keywords: open data; open data hackathons; environmental sensors; smart city; Cybersalon; copyright.

Biographical notes: Sophia Drakopoulou is a Senior Lecturer in Media Communications and Culture and a Director of Programmes in the Media Department at the faculty of Arts and Creative Industries at Middlesex University. Her research explores networked technologies, the city and everyday life. She is a co-founder of Cybersalon.

This paper is a revised and expanded version of a paper entitled 'Data and the city – accessibility and openness. A cybersalon paper on open data' presented at Athens Hybrid City III International Conference 'Data to the People', Athens, Greece, 17–19 September, 2015.

---

1 Introduction

This paper discusses and explores the current challenges posed by and potential benefits of open data to create systems and installations that can serve the needs of a particular group of people in a specific location. The primary challenges highlighted here are the formats in which open data are made available, the digital literacy required to exploit it, and the copyright issues that arise from commercial use. As will be shown, at present Hackathons provide some creative solutions but are reserved for the 'techno-elite'. The paper argues that because of the present challenges of open data, existing datasets such as governmental open

data are limiting the potential possibilities of open data application development. In contrast, a new trend is developing in sensor-based purpose-gathered citizen-led open data that can be used to create meaningful interactions with participants and develop open data applications and systems that can serve a particular local area and group of people.

The paper also draws on the think-tank Cybersalon's experiences of participating in a government-led open data hackathon in London and organising and running a similar event in Athens. The comparison will show that the general arguments about open data match Cybersalon's empirical experience, especially the difficulties posed by the formats in which data is released and the idea that there's a certain dullness to the datasets made available by governments. As a real-world example of the difficulties of accessing and dealing with open data, the paper uses the experience of Open Sensors.io, a London based company supported by the Open Data Institute (ODI) that provides Application Programming Interfaces (APIs) to help businesses, start-ups, and individuals develop applications for real-time data deriving from sensors.

The last section of the paper showcases examples of bottom–up open data and smart city applications and identifies lessons for future such efforts. It also showcases two open data experiments produced by Cybersalon members that use environmental sensors to gather their own data. These are Ilze Black and Nanda Khaorapapong's The Breather, a 'breathing' balloon that uses high-end, sophisticated sensors to make air quality visible; and James Moulding's AirPublic, which measures pollution levels.

The paper highlights the importance of easy, unimpeded access to datasets, open sensors, and clean data that is compliant with established best practices for interoperability and accessibility. The paper also argues for better access to tools that can be used to extract, clean, and visualise data in order to make it meaningful to and usable by general audiences. Lastly, the paper asserts that opening up existing databases of governmental data is simply not enough by itself – instead, more focus should be given to collaborative models between citizens, business, and civil services and to sensor-based gathered data.

2 Open data and the Cybersalon HyperHabitat series

The London-based collective and think-tank Cybersalon's 2014–2015 HyperHabitat series of events, projects, and presentations investigated the changing nature of our living environments. Among other activities, the series included Cybersalon events, participation in the London Hackney Council's "Hack-ney-thon: 24 Hours to Hack for Hackney", and a study of data gathering for the retail industry. In September 2016, as part of the Athens Hybrid City III Conference "Data to the People", Cybersalon organised a two-day hackathon.

2.1 What is open data?

Open data are widely presented as a new movement, but it's not really new. In various ways, academia and others have been accessing each other's data for centuries, and much of the movement's ethos derives from the 35-year-old open source community. However, today open data has a centre-stage role to play in developing successful applications that involve

citizens, the civil service, and the commercial sector. Web inventor Tim Berners- Lee and University of Southampton computer scientist Kieron O'Hara (Berners-Lee and O'Hara, 2013) have argued that the 'web of linked data' is the key to preserving the open character of the web and to using it as a resource to create new value and knowledge. For them, the web is fundamentally a read/write medium, and keeping it editable by all is essential for sustaining online openness. In their vision, social media platforms and mobile apps pose a problem by walling off the data their users generate and not sharing it with the web. They proposed a new mode of sharing data: open data.

In principle, what makes a dataset 'open' is a licence that states it may be used freely, subject to conditions such as giving credit to the publisher (attribution) or releasing the results of combining it with other datasets under the same open data licence. Open data may be linked to, freely shared, and freely discussed; ideally, it's available in a standard format so it can be readily processed and reused. To make it reliable, open data needs guaranteed availability, consistency over time, and traceability to its original owner.

The idea behind opening up data collected by public bodies that historically kept it inaccessible is to enable informed decisions of interest to the public as well as local officials, government and state agencies, and services (Black, 2012). Raw data from research trials is beginning to be shared in academia and the sciences. For example, the Incorporated Research Institutions for Seismology (IRIS) provides seismographic data from hundreds of stations to inspire new collaborative projects and research initiatives (Tananbaum, 2008), and the All Trials campaign pushes for the publication of raw data from clinical trials. Swivel, a 2006–2010 attempt to create a "YouTube for data", failed to get beyond the prototype stage, but showed the potential.

3 The current challenges of existing open datasets

As will be shown in the paper, currently, the main challenges in open data are: the specialist knowledge and skills: the digital literacy required to use the data; the delays caused by the lack of a set standard for circulating, sorting, and using open data, and; the ethical and copyright issues when combining open data with other datasets. As will be discussed below, currently, it is commercial companies that are benefiting from the use of open data rather than the community. Providing training in the necessary skill set and improving the quality of published data (Scassa, 2014) are some suggested solutions to the challenges posed by existing open datasets.

Although the idea of open data are commendable, and governments worldwide are embracing it, the practice brings new problems. For example, many of the datasets governments are now opening up were never intended for external use, and figuring out what they are and how they're structured is hard work. A second is that there is no set way to handle and create open data or applications made with it. Organisations observe different principles (Kitchin, 2014); computer scientists practice different methods of data mining (Shelton, 2015). In other parts of the world such as Nigeria, where road maps and other standards do not yet exist, this mode of open data can present solutions (Okwuone, 2015). However, even in developing countries open data initiatives meet the same challenges and

difficulties as they do in the Western world. A recent study of open geographic projects in Ulaanbaatar, Mongolia (Wang, 2016) reveals similar issues: inaccessibility of data due to the lack of necessary infrastructure and inadequate digital literacy (Wang, 2016).

There are two main elements that stall and complicate working with open data: copyright and the requirement for specialist knowledge and skills. These two elements are the biggest reasons why the most significant beneficiaries from open data at present are commercial companies rather than individuals and communities. As will be shown, despite these challenges, there is real potential in open data to enable citizen-centric approaches by creating interactive installations and systems that benefit niche groups of people. The improvements made to services, and the communities involved, all of which turn using and creating with open data into a creative process that can be modified as the situation demands. There's a developing trend in sensor-based, purpose-gathered, citizen-led open datasets, and citizens collecting environmental data is a growing field (Gabrys et al., 2016). As opposed to existing datasets and governmental datasets, citizen data offer alternative and collaborative ways of collecting data and creating applications and systems that serve a particular local and a particular group of people.

3.1 Benchmarking

Benchmarking can be interpreted as establishing a standard, drawing focus away from the practical experience of artists, entrepreneurs, and academics, among others. An analysis of government-led open data initiatives (Susha et al., 2014), including the World Bank's ODRA-Vistula Flood Management Project, found many attempts to create benchmarks for evaluating the success and impact of open data projects: ODB (from the Open Data Institute and the World Wide Web Foundation), ODI (Open Knowledge Foundation), the PSI Scoreboard (ePSI Platform), and Open Data Economy (Capgemini Consulting). Such efforts typically include examining aspects of the projects such as policies, readiness, implementation, adoption rates, impact, and ecosystems, which may include publication, organisational transformation, community building, user support, and feedback loops. After studying the effectiveness of, differences between, and limitations of the different efforts, Susha et al. conclude that such efforts require the development of administrative levels from global to local as well as the provision of support for improving the status quo, and that benchmarks must be a continuous process of measurement, research, and updates in order to meet the needs of rapidly-changing development (Susha et al., 2014).

3.2 Commercial use of open data

Another critical issue is the fact that currently open data made available by governments are disproportionately used by commercial companies in order to make profitable smartphone and other kinds of software and web-based applications. Slee (2014) explains that corporations have the right set of skills and access to open data – and therefore are best placed to profit from its application and use, in contrast to community groups, which lack technical skills and resources.

Kitchin (2014) quotes Jo Bates (2012) to argue that in the UK Open Government Data (OGD) initiatives gained 'political traction' only when the business sector realised there was

commercial potential in accessing free data. In that sense, it can be said that the UK government's embrace of open data was not a well-intentioned political act aimed at enhancing participatory democracy, but rather a commercial move intended to generate more capital, aid businesses, support economic growth, and stimulate the start-up sector.

Many of the processes that determine how datasets are collated and collected imply social exclusion and social privilege (Johnson, 2014). A frequently-quoted example (Slee, 2014; Gurstein, 2011) is Benjamin's study in Bangalore (Benjamin et al., 2007), where digitising the land registry favoured those with financial resources. Gurstein argues that this does not mean that all digitisation exacerbates the digital divide; instead he makes the point that the methods of accessing and processing open data need to be made more egalitarian. For Gurstein, important factors in accessing open data include: "cost and availability of internet access, the language in which the data is presented, the technical or professional requirements for interpreting and making use of the data, and the availability of training in data use and visualisation" (Gurstein, 2011).

As an example, Gurstein mentions the State of California, which pioneered the Ca.gov website and where communities are being trained on how to use open data to create visualisations. However, the main issue central to the critique of the commercialisation of open data are the fact that although open data are released into the public domain, commercial profit-making companies are using open data to create closed systems and claiming copyright protection. Combining open data with other datasets can compromise privacy and bring up ethical and legal issues (Slee, 2014; Scassa and Campbell, 2009).

3.3 Copyright and open data

A significantly growing area for litigators and legislators is the ownership of open data and the copyrighted representations of this data (such as roadmaps, transit maps and data visualisations). There have already been several US legal cases in which courts have been asked to decide who owns the representation of the data and not the dataset itself. The reason is that copyright law applies differently to open data that are part of a series and representations produced based on that data such as transport maps and transit timetables. Under US copyright law, maps are "[p]ictorial and graphic works of authorship" (Scassa, 2014, p.1782). The result is that even though the data may be free of copyright, its representation can be copyrighted (Scassa, 2014, pp.24–26). Copyright law surrounding original interpretations and representations of open datasets is still in development. For this reason, many municipalities are reluctant to release data as open, especially to app developers, since the developers do not share their profits and the communities and local citizens do not benefit directly from the release and use of that open data.

Large broadcasting corporations such as the BBC have been quick to realise the potential of open data to enhance audience and citizen participation. However, again, the complexity of copyright law has added to the difficulty of releasing content and/or platforms, even when they are based on content sourced from the public domain. The BBC's open data experiment 'Backstage' was intendent to "encourage data reuse, knowledge exchange, social networking, community engagement and outreach". (Lin, 2015, p.150). The initial idea of Backstage was to 'get the BBC's data out' (p155) connecting people both inside and outside the organisation,

and building a developer community'(p.149). As it was being developed, copyright issues emerged in audience participation and in the reuse of content and data. In a typical example, Littledale a researcher funded by the BBC and the Arts and Humanities Research Council, says that many of the search engines and systems he developed could not be offered to external audiences because of the difficulty of complying with the licences controlling the copyrighted material he used (Littledale, 2010). This example shows the problematic that emerge when using existing open datasets for commercial purposes and the copyrighting representations and patenting of the systems that emerge from it.

3.4 Openness critique

Open data development and policies are at the forefront of government decision-making in both the US and the UK, where open data are promoted as an enabler of government transparency. Both governments emphasise the potential benefits to society and the economy (Susha et al., 2014). However, making, storing, releasing, reusing, and distributing open data all present challenges that detract from the idea of 'open'. Who should be responsible for making data formats accessible remains an unanswered question. The open source movement has paved the way for negotiating between open and free use of software and data and commercial interests and applications. Open data uses similar rhetoric to that of open source software movements such as that surrounding GNU/Linux (Slee, 2014, pp.115–116; Ritzer and Jurgenson, 2010, pp.23–25). Both are based on community-driven systems that should be free for anyone to use, including commercial companies in order to create marketable products.

Open spatial data is spatial because it's been collected at a particular location and whilst mobile (Lauriault, 2017). Spatial data can be generated by governments, business and citizens and most datasets are meant to be open (Lauriault, 2017). Currently, an entire business sector is being developed that capitalises on 'free' government data and also benefits from tax relief and government support (Lauriault, 2017, p.99). The sharing of governmental open data is made possible by entrepreneurs, media creatives and techies who are expected to innovate and make profitable applications. So even though governmental data are shared with citizens, in the context of governmental openness, the citizens' political engagement with that data is limiting in scope (Birchall, 2016, p.2, p.8). Compared to existing datasets from governments and other bodies whose "data volumes were much smaller", new datasets created by "collecting data from the crowd" are bigger and allow for more flexible and creative solutions and applications (Batty, 2017, p.213). As will be shown in the paper, the possible applications of sensor-generated, citizen-gathered open datasets can combat the present challenges of existing open datasets.

4 Hackathons and open data

As discussed previously, much needs to be done to create standards of practice and promote the necessary digital media literacy so that all can access such material. Davies (2010), who has conducted a comprehensive review of the UK's Data.gov.uk site, warns about the risk attached to focusing on data-for-developers. Similarly, Johnson (2014) differentiates between 'enterprise open' and 'citizen open' data. Because of the technical skill required to process

open data, Hackathons are seen as the best current option for creating innovative open data applications. Hackathons highlight the challenges and present some of the potential governmental and other datasets can offer and facilitate existing civil services as well as create new ways of experiencing the urban environment. However, they also highlight issues of techno elitism and inaccessibility of formats that current open datasets pose.

Many of the best-known open data initiatives began with hackathons, including the BBC's Backstage (Lin, 2015). Data.gov.uk began with a large 2009 hackathon-style event, tapping into the hacker community to stage the first "Rewired State", where eighty developers built applications by scraping government data (Davies, 2010). Hackathons regularly use common technology industry techniques such as rapid experimental prototyping and engineering to create practical applications. In the public sphere, Hackathons can kickstart individuals' understanding of code and provide general-purpose spaces for focused innovation efforts. For example, New York City's Big App organises regular events at which programmers, developers, designers, and entrepreneurs use open data to devise solutions to civic problems.

In mid-November 2014, with the cooperation of the UK's Government Digital Service and MiniBar Labs (an agency connecting tech startups, industry, and communities), Hackney Borough Council organised the weekend-long "Hack-ney-thon: 24 Hours To Hack For Hackney". The brief was to get software developers and cultural industry practitioners living in and around Hackney and its Tech City to collaborate intensively to create local solutions and make Hackney's council services more efficient and user-friendly. As part of the event, the Cybersalon team developed Hackney Treasures, an app that links location data to historical Wikipedia papers to celebrate the work of all the creative and interesting minds that live or have lived in the Hackney area. A variety of technology partners provided API access and mapping tools to augment the government's open data sources. Simple ideas like Hackney Treasures demonstrate that technologies can be modified to meet the requirements, needs, and recreational activities of a particular group as well as represent an area in a specific context. The event's winning applications included: an app that enabled citizens to review and assess the council's response rate in dealing with complaints and repairing street lights and potholes; another that simplified the process of registering a new birth; and a third that enabled interested parties to get easy, personalised notifications regarding city planning applications.

Open data are widely established in Europe in general and in Greece in particular. In September 2016, the Cybersalon team organised and ran the two-day hackathon "Wave Your Open Data Magic Wand" as part of the Athens conference Hybrid City III: Data to the People. Aimed at participants of all technical levels, the brief was to mix creatives and developers and build innovative data visualisations and concept mobile apps and engage with the raw data in order to show Greek contemporary culture in the best possible light. Datasets for the hackathon came from Eellak.gr (Εταιρεία Ελεύθερο Λογισμικό/Λογισμικό Ανοικτού Κώδικα), Greek Free/Open Source Software Society (GFOSS), and Geodata.gov.gr, and ranged from data collected about Greece's frequent forest fires and the locations of preservations areas to the national census. As noted above, the technical expertise required to access to open data partly depends on how it's formatted. Our own experience at the Athens hackathon confirms this. Most of the datasets were formatted as comma-separated values (CSV files). Because of the data's 'rawness' in terms of meaningful values, participants needed to be

trained on sorting through and using them. It proved difficult to make these files readily accessible to the less digitally literate of the hackathon attendees.

There were many Geodata.gov.gr datasets, and we sorted them into categories and picked about 25 for the hackathon. Even this smaller number of datasets proved too many for our participants to grasp and engage with. Given that most of the audience was practitioners of the creative industries rather than developers, the project focused on the 'spirit of open data' rather than active use of the datasets we provided. The winning app was 'Find my Oasis', an app citizens can use to rate parks in terms of cleanliness, dog walking, meeting space, and so on. The experience taught us that without specialist technical knowledge open data files are difficult to handle, manage, and use by people lacking technical knowledge.

4.1 Open data formats

Most literature on the subject, whether it's written by the ODI and other practitioners or academics agrees that the format in which data is made available is crucial in making it accessible to all. The US's Data.gov site offers its datasets in multiple formats – XML, JSON, RDF, and CSV – in an effort to make them as widely accessible as possible. All of these formats can be turned into spreadsheets, a format many can read more easily. However, each has its strengths and weaknesses. RDF (Resource Description Framework) is recommended by the W3C and is the most versatile format; it allows easier sorting. XML/XSD (Extensible Markup Language Schema Definition) has structure and helps define elements and categories; in that sense it is more easily usable. JSON (JavaScript Object Notation) is primarily used for live data that's updated in real time via a web browser. Finally, CSV, as already noted, is hard to work with because it has no structure such as headings and columns, and requires a key in order to make sense of it. As our developer said during the Athens hackathon: "Files are boring. We want APIs!" (Application Programming Interfaces). Most open data researchers and advocates agree that easy-to-use interfaces such as search engines and APIs are essential to enable citizens to make use of these datasets. For example, the World Bank-funded Nigeria.opendataforafrica.org attempts to make its data easily readable by citizens by displaying data such as infant mortality rates, oil exports, and unemployment rates in familiar and traditional visualisations such as pie charts and bar graphs.

4.2 Banality of open data

In advocating the semantic web, the World Wide Web Consortium (W3C) argues for the importance of linked data. In order to make the data valuable and meaningful to people, it needs to be combined and contextualised. In both Athens and Hackney, the government datasets we accessed included the census, land registry, natural areas of preservation, the register of births and deaths, and statistical data about fires, thefts, and health. You could call this type of data dull, an accusation that gains some support from the Data.gov.uk site, which showcases the apps companies have made using governmental open data. With the exception of the 'Flood Alerts API' application, which obviously benefits an at-risk community, all the apps are transit-oriented community. They help find walkable routes, retrieve bus arrival times, or calculate salary in relation to housing market and NHS data. Most of these APIs and apps require payment and present commercial opportunities. So far, transit and wellbeing data attract particular commercial interest, and there has been much focus on the

copyright of these apps (Scassa, 2014). However, it's still unclear whether other kinds of governmental data present the same commercial opportunities (Johnson, 2014).

Although the BBC began the Backstage initiative as a way to open its databases to the public, it was mostly the 'techno-elite' who engaged with it as they were the ones with the necessary specialist knowledge and skills (Lin, 2015). Backstage fell afoul of the same challenges that need to be addressed today in the context of open data: digital literacy and copyright. Hackathons highlight the current issues that arise from accessing multiple formats, the digital literacy required to process open datasets and the issues with copyright deriving from the reuse and representation of Open data. In that sense then, it can be said that Hackathons may present some creative solutions to the current challenges of open data, but also affirm that they required highly developed technical knowledge and are therefore reserved for the 'techno- elite' rather than the general public.

5 Sensor-based open data

As discussed, existing datasets that derive from governmental data and administrative data (Lauriault, 2017, p.103) such as the census and demographic data on unemployment, income or education provide limited ways in which these can be made meaningful in everyday life. This paper argues that beyond existing open datasets and the challenges posed by their commercial use, sensor-generated data that can be appropriated to benefit a local community and appropriately design APIs and user interfaces can enable citizens to access, use, and make sense of open data. Instead of dull government data, data generated by sensors deployed throughout a city can provide real-time information about footfall, traffic, and other urban rhythms. In this model, cooperation among citizens, businesses, and civil services can create new ways of generating and using open data for the benefit of the people of a particular locale. What's more, this kind of data, also called crowd-sourced, or citizen-led data (Lauriault, 2017, p.103) can be made meaningful to everyday life and activities by providing models and services that facilitate everyday life and serve a community and/or a particular local area. In that sense then, spatial data can be place-specific and gathered for the specific aim and purpose of servicing the needs of that community and place.

Indeed, new models are emerging that combine efforts by the state, government, companies and citizens, in order to build real-time dynamic systems and representations of data that can serve a particular group of people and/or a particular local area. The potential in the use of open data gathered by citizens is currently being explored in a variety of projects. Citizen Sense, a project led by Dr Jennifer Gabrys, uses environmental sensor data to facilitate citizen engagement providing crowd-sourced datasets. Projects and applications range from measuring air pollution in the urban environment, to monitoring the flora and fauna of natural habitats. Furthermore, initiatives such as the Programmable city (Dublin) provide examples where civic services, citizens, academia and businesses work together to produce graphs and interactive systems based on purpose-gathered open datasets. Community-led sensing networks such as the Air Quality Egg allow individuals to take air pollution readings of their environment and collectively represent that data on the website's map. Although the Air Quality Egg sensor readings are not considered valid data for scientific research (Verrilli,

2013), the representation of collated readings around the world has a political underpinning towards the climate/ecology debate.

New crowd-sourced open data innovation models are developing, where the user is an active participant in the gathering of the data and will directly benefit from the application of that data. User innovation is an example of new models emerging in which the consumer is actively involved in the making of the product or service they are using, either by being an active participant in the gathering of the data or by emitting and sharing their own data (see von Hippel, 2005). Initiatives such as Citizen Sense, Programmable City and Air Quality Egg are part of the growing trend in 'citizen data' and provide models of open data gathering where the data gatherers directly benefit from the application of that data. The data gatherers (be it citizens or companies) are actively engaged in the gathering of that data, by using open data sensors and other sensor-based devices such as the smart phone, and will directly benefit from the application that will arise from that gathered data. This data can be more interesting and present far more opportunities for innovation than the existing governmental open datasets, as the data is gathered for a specific aim and purpose and in order to develop a specific application and/or system that can benefit users and companies alike. This paper will now explore ways that open datasets can be made meaningful to and have a direct impact on citizens' everyday life.

5.1 Using environmental sensors and making open data meaningful and tangible to citizens.

What follows presents the work of artists and practitioners who are experimenting with the forms that sensor-generated data might take and ways the public might engage with and use them. The projects discussed here showcase bottom-up approaches and demonstrate the possibilities offered by existing open data sensor-based technologies. Two projects – The Breather and AirPublic – have experimented with both the interaction between sensors and the public and the way data is generated. They offer solutions, but also raise questions such as how best to apply sensors in public spaces and which responsibilities citizens might take on in everyday life to make measuring, distributing, and exploiting open data more meaningful.

5.1.1 The breather

The Breatherdeveloped at Queen Mary University London by Ilze Black and Nanda Khaorapapongin, was the ODI's first art commission. It is a prototype large-scale, interactive, data-driven balloon-like object intended for future city environments and uses the air quality egg sensor (see airqualityegg.com). The project's main objective was to make real-time data readable in the physical world via an interactive interface that responds to changes in air quality as measured by networked sensors, rendering visible what was previously invisible through the perceptible movement of a breathing object. The Breather project intended to install a large balloon on every street corner that signalled the air quality at that spot via the depth of each 'breath': a living, breathing informer that helps city dwellers choose the best route to navigate ever-more-dense city streets and see their pollution levels. However, The Breather's public reception at a busy London roundabout told an unexpected story. The playfulness of its movement and softness of its surface invited passers by to stop and hug it, or tune into and listen to its breathing.

5.1.2 AirPublic

In a second prototype using environmental sensors, James Moulding's AirPublic measures London's air pollution by deploying mobile air quality sensors. These piggyback on existing transport networks – chiefly the municipal bike-hire scheme – in order to improve the availability of air quality data. Moulding found that even in the most-monitored locations current information on urban pollution is patchy at best. In the many locations across London where air pollution is not monitored, the only available information is often-unreliable inferences from computer modelling. While this basic level of information is enough to draw government attention to the issue of dangerous urban air quality, the data presently available is insufficient to help individuals. Moulding's mission is to build reliable, detailed, organic, real-time information about air quality that relates to people's everyday lives and movements. The knowledge AirPublic develops will allow those who call these cities home to learn how pollution affects them as individuals and how they can reduce their exposure, plus suggest how, together, we can create healthier environments. Increasing the resolution of air quality data and creating accessible real-time measurements are key to making air quality data relevant, increasing awareness of air pollution, and encouraging wider society to take action. Both these examples showcase forms that data gathering by sensors and data representation could take in the interests of engaging people directly with specific goals.

A way to combat the banality of existing open datasets is to create purpose-gathered datasets using environmental sensors and to provide representations and systems as ways to make open data more tangible and meaningful to citizens. Examining a model where making open data useful to a particular local, requires modes of collaboration between private initiatives, government intervention and citizen gathering and participation. The next session discusses the modes and processes that are available to assist small start-ups and other businesses to develop using open data sensors and creating solutions for particular locals and areas in London.

5.2 Collaboration between citizens, business and civil services.

This section looks at examples of cooperation between government companies and citizens with the case study of OpenSensors. Drawing from Yodit Stanton's experience as presented at the 26 February 2014 Cybersalon event "UnSmart Cities: The New HyperHabitat". And also highlights how the aforementioned issues cause delays in even the most open models of collaboration.

5.2.1 OpenSensors

This example highlights the need for entrepreneurial initiatives to open up communication between communities, the public and private sectors. Software engineer Yodit Stanton is founder and CEO of London-based OpenSensors.io. This ODI-incubated start-up provides software and platforms for the Internet of Things with the goal of creating a 'commons' around real-time data for all kinds of internet-connected devices. Stanton's Cybersalon presentation detailed the circuitous route she had to follow to acquire publicly-owned datasets from government agencies and companies when she was commissioned by

Westminster Council to analyse its parking and footfall data. After three months of trying, neither those who had commissioned the study nor those who had provided the systems that generated and collected the data could provide access. Incomprehensible outsourcing contracts, a chain of custody involving eight separate companies, and tortuous chains of conversions between file types as the data travelled from sensor device to data analyst all helped create the blockage.

While Stanton's experience showed her that local authorities are keen to open up their data and appreciate its usefulness in delivering insight and improving the civic space, she noted they often do not know how. Further, they fear liability for future abuse or misuse. Stanton challenged us to forget the hype surrounding future smart cities and focus instead on tangible problems we can solve today by simplifying the costly, increasingly complex digital infrastructures that councils must manage. She cited Bristol's efforts to create a 'programmable city' as an example of getting it right. Open Programmable City Region (see www.bristolisopen.com) is collaboration between industry, academia, local communities, and local and national government that is intended to upgrade the city's bandwidth and wireless connectivity. Stanton's talk highlighted openness and cooperation as essential for solving problems in civic systems and services. Finding solutions that serve all parties and create cohesion requires businesses, municipal services, and communities to come together and share the data each has collected. At present, this model of cooperation is difficult to attain as there is no set process for it, as Stanton demonstrated. This example shows that even in the most technically advanced cities in what's widely believed to be the world's leading open data country, obtaining, using, and collaborating with open data are difficult and slow.

5.3 Summary

Despite the slowness in the example above, there's immense potential in purpose gathered data, with the data gatherers directly benefiting from the system that will emerge from that (application of) open data. In order to create meaningful interactions and to make sense of open data for a local community and/or a particular local, government, companies and citizens must work together. As shown in the paper, it is the nature of open datasets that cannot be commercially free even if they are 'open' as they present creative commercial solutions and require venture capital investment for building applications and systems. Equally, open data cannot solely belong to the government domain as it lacks specialist knowledge and the copyright issue also makes it difficult to keep them open or free. And as discussed, Hackathons provide some solutions but are reserved for the techno elite. Beyond the current challenges of open data, these emerging new models of dynamic open data systems where citizens are gathering their own data, as these are developing in London, suggest new ways to make open data meaningful and accessible to people.

5.4 Open data and the street

In this context, and similar to Staton's talk, Cybersalon's research strand on civic data challenges, led by Eva Pascoe, explores collaborative models of sensor-generated open data and maps of municipal data gathering in order to improve the way these are processed. Pascoe studies how these findings are communicated to and used by the people who need

them. Pascoe's work on urban and retail regeneration using London as a case study has found three impediments to providing comprehensive city-wide open data: patchy and irregular data gathering; incomplete datasets; and the lack of user-friendly analysis tools for non-expert users such as small businesses, citizens, and decision- makers. In order to combat these problems identified by Pascoe's work is to shift from limiting data gathering to city or council functionaries, who have decreasing resources and insufficient knowledge of IT, to collaborative efforts by all parties (citizens, government bodies, companies) in order to create complete datasets.

In this model, collaborative practice would have every citizen and business contributing to local data gathering using easy smartphone-based tools. The data thus collected could then be collated and published as open data by specialists such as the London-based Local Data Company or the Open Data Institute to enable city management and regeneration projects. Wikidata, OpenStreetMap, and Guy Lansley's London Twitter footfall analysis projects are all examples of delivering collaborative, high-quality results at low cost (Black, 2012). In this model, easy-to-use interactive tools would be made available on a city-wide data portal that enables all who need to make better decisions to query, visualise, and perform simple trend analysis. With those capabilities in place to accept the right inputs and produce the right outputs, city services will be able to manage and regenerate with confidence as decision-making will be based on real-time, accurate, and complete information. Cooperation among civil services, companies, and citizens gathering their own data can bring real benefits to city services.

The key to all this is purposefully gathered Open data – but that by itself won't be enough. Completing datasets will require safe storage for data that has been anonymised to protect privacy, and some closed data will be needed to support the knowledge derived from open data. Closed data will need cities' institutional support. In the UK, this is emerging in the form of the ESRC-funded, multi-university Consumer Data Research Centre, which offers cities, retailers, and transport businesses a data platform for secure, anonymised storage for shared research projects.

6 Conclusion

The paper reviewed the current challenges of existing datasets : digital literacy, critiqued the idea of openness and discussed how mundane governmental data can be combated by a move to making open data meaningful and tangible to people – by allowing citizen to gather open data and interact with installations that make open data tangible and accessible to people. Also, this paper highlighted the challenges that cause delays in collaborations between interested parties, and the reasons why companies are disproportionately profiting by open data.

Citizen-gathered open data can be made more meaningful and applicable to everyday life as the Air Breather and Air Public demonstrate. The developing trend in collaboration between municipalities, businesses and individuals such as the 'Programmable City' and 'Citizen Sense' showcase the possibilities of those agents to come together and create applications and systems that are tailored for the needs of a particular area and a particular group of people.

The paper reviewed the current landscaped of open data based on Cybersalon's experience to assert that getting data to the people (or at least to the API) is difficult, circuitous, and slow, requiring an intricate process of leadership, public relations, and perseverance. Despite myriad tools and initiatives, there is no one solution for the actual transfer of that data. The route raw data takes from the company that owns the sensors and collects the data they gather to the company that will develop the API or the creative practitioner who will create a user application or visualisation is difficult and inefficient.

Despite regulations published in July 2015 that made government data open by law, data is not so open in practice. As this paper shows, among the many London initiatives attempting to bridge this gap and eliminate the bottlenecks, the ones that are effective are those that concentrate on local solutions. Open data practices are still being shaped; this is the beginning of an era of new modes of distribution and representation of, and interaction with, both real-time, sensor-generated data and government data.

Lastly, the paper showcased the future potentials of open data sensors. For the potential to be fully realised, this paper argued that more attention needs to be paid to making open data more accessible to citizens by simplifying formats, providing APIs that everyone can use, and designing meaningful interactions with open datasets such as the Breather project. As open data becomes more established, it will create more opportunities for collaboration and innovation. Throughout the HyperHabitat series, Cybersalon members and contributors highlighted the need for bottom-up approaches. Their practice demonstrated that engagement with the local community is a powerful force for real social change.
More broadly informed and engaged citizens are the key requirement for the future of open data. What's more, governments must change from being solely data gatherers to also being data publishers. The change will present a challenge, given their ever- decreasing resources and general lack of digital literacy, as well as broader issues around interoperability and data publishing methodologies. If these can be solved, purpose- gathered and citizen-gathered open data presents a real opportunity for a future in which new models of collaboration among governments, industry, and communities produce social change and innovation that equally serve all parties.

References

Bates, J. (2012) 'This is what modern deregulation looks like: co-optation and contestation in the shaping of the UK's open government data initiative', Journal of Community Informatics, Vol. 8, No. 2, http://ci-journal.net/index.php/ciej/article/view/845/916 (Accessed 1 March, 2016).

Batty, M. (2017) 'Producing smart cities', in Kitchin R, Lauriault, T.P. and Wilson, M.W. (Eds.): Understanding Spatial Media, Sage, London, pp.204–215.

Benjamin, S., Bhuvaneswari, R., Rajan, P. and Manjunath, R. (2007) Bhoomi: 'E-Governance' Or an Anti-Politics Machine Necessary to Globalize Bangalore, https://casumm.files. wordpress.com/2008/./bhoomi-e-governance.pdf (Accessed 1 March, 2016).

Berners-lee, T. and O'Hara, K. (2013) The Read–Write Linked Data Web, DOI: 10.1098/rsta.2012.0513,

http://rsta.royalsocietypublishing.org/content/371/1987/20120513 (Accessed 1 March, 2016).

Birchall, C. (2016) 'Shareveillance: subjectivity between open and closed data', Big Data and Society, Vol. 112, No. 1, pp.1–12.

Black, A. (2012) 'Open data movement', P.M. Public Management, Vol. 94, No. 6, pp.22, 23.

Davies, T. (2010) Open Data, Democracy and Public Sector Reform: A Look at Open Government Data Use From data.gov.uk, Unpublished Master's thesis, University of Oxford, Oxford, UK, http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open- government-data-being-used-in-practice.pdf (Accessed 1 March, 2016).

Gabrys, J., Pritchard, H. and Barratt, B. (2016) 'Just good enough data: figuring data citizenships through air pollution sensing and data stories', Big Data and Society, Vol. 3, No. 2, pp.1–14.

Gurstein, M.B. (2011) 'Open data: empowering the empowered or effective data use for everyone?', First Monday Peer Reviewed Journal on the Internet, Vol. 16, No. 2, http://firstmonday.org/article/view/3316/2764 (Accessed 1 March, 2016).

Johnson, J.A. (2014) 'From open data to information justice', Ethics and Information Technology, Vol. 16, No. 4, pp.263–274.

Kitchin, R. (2014) The Data Revolution Big Data, Open Data, Data Infrastructures and Their Consequences, Sage, London.

Lauriault, T.P. (2017) 'Open spatial data', in Kitchin R, Lauriault, T.P. and Wilson, M.W. (Eds.): Understanding Spatial Media, Sage, London.

Lin, Y. (2015) 'Open data and co-production of public value of BBC backstage', International Journal of Digital Television, Vol. 6, No. 2, pp.145–162.

Littledale, A. (2010) Data Art on Backstage, http://www.bbc.co.uk/blogs/researchanddevelopment/ 2010/06/data-art-on-backstage.shtml (Accessed 1 March, 2016).

Okwuone, N. (2015) Open Data Leaders. Network ODI Interview, https://theodi.org/open-data- leaders-network (Accessed 1 March, 2016).

Ritzer, G. and Jurgenson, N. (2010) 'Production, consumption, Prosumption. The nature of capitalism in the age of the digital 'prosumer'', Journal of Consumer Culture, Vol. 10, No. 1, pp.13–36.

Scassa, T. (2014) 'Public transit data through an intellectual property lens: lessons about open data', Fordham Urban Law Journal, Vol. 41, No. 5, pp.1759–1810.

Scassa, T. and Campbell, L. (2009) 'Data protection, privacy and spatial data', in Devillers, R. and Goodchild, H. (Eds.): Proceedings of the 6th International Symposium on Spatial Data Quality, Taylor & Francis, St. John's, Canada.

Shelton, T. (2015) 'Research resource review', Progress in Physical Geography, Vol. 39, No. 3, pp.412–413.

Slee, T. (2014) What's Yours is Mine: Against the Sharing Economy, OR Books, New York, London.

Susha, I., Zuiderwijk, A., Janssen, M. and Gronlund, A. (2014) 'Benchmarks for evaluating the progress of open data adoption: usage, limitations, and lessons learned', Social Science Computer Review, Vol. 33, No. 5, pp.613–630.

Tananbaum, G. (2008) 'Adventures in open data', Learned Publishing, Vol. 21, No. 2, pp.154–156.

Verrilli, A. (2013) 'Air quality egg shows power of open development', Environmental Monitor, http://www.fondriest.com/news/air-quality-egg-shows-power-of-open-development.htm (Accessed 27 April, 2017).

Von Hippel, E. (2005) Democratising Innovation, MIT Press, Cambridge, Massachusetts, London.

Wang, X. (2016) 'Unequal publics: open geographic data, development and community access in ulaanbaatar, Mongolia', Paper Presented at the 62th Annual Meeting Association of American Geohraphers, 28 March–4 April 2016.San Francisco, California.