EkoMiasto#Zarządzanie
Zrównoważony, inteligentny i partycypacyjny rozwój miasta

pod redakcją
Zbigniewa Przygodzkiego
Marcin Feltynowski*

CITIES IN THE AGE OF THE BIG DATA

* Ph.D., University of Lodz, Faculty of Economics and Sociology, Department of Regional Economy and Environment, e-mail: marcin.feltynowski@uni.lodz.pl
9.1. Big data basics and challenges

In the era of knowledge based economy an indispensable element of its creation is the access to information sources, resulting from the process of data analysis. Alongside development of information society, the need for access to data is growing and, alike, increases the capacity to generate new data, originating from qualitative and quantitative analyses, following the activities of various local actors. The development of information and communication technologies and accompanying expansion of access to personal computers and the Internet lead to a constant increase of computer memory capacity with simultaneous reduction of their physical size. The constant progress observed in computer memory capacity reduces the need to limit the size of data resources or the disk memory consumed by software. It is assumed that data stored in the digital universe will reach the level of 8 zettabytes by 2015 (1 zettabyte = 1000$^3$ terabyte = 1000$^4$ gigabyte) [Gantz, Reinsel, 2011, p. 3], and to 50 zettabytes by 2020 [Van Rijmenam, 2014, p. 9].

It should be stressed that access to data allows people to resolve properly a number of complex problems and to manage different areas of human activity. In the mid 19th century, owing to data collection on cholera incidence, John Snow managed to trace the source of a cholera outbreak in London Soho district. Nowadays, one can use various secondary data, encountered in daily life, as well as, the data collected by specialised entities and officials in their professional practice. More often, it is the public officers themselves, who notice the need to collect data, which is supposed to improve the quality of their work and can be used to make proper decisions concerning the management of their city. Thus, in the area of city management alike, we come across the notion of big data, which has not been equipped with a clear-cut definition yet. The literature of the subject gives a number of definitions, depending on the disciplines the authors specialise in.

We observe the most dynamic development of big data issues in the 21st century, but the origins of this phenomenon can be traced back to the 1940s, when the Oxford English Dictionary used the term ‘information explosion’ [Press, 2013]. The very term big data was used for the first time in the 1990s. According to the authors, the big data term referred to the data used for visualisation, which was large and thus required the user to possess appropriate data carriers [Cox, Ellsworth, 1997, p. 235]. The research into data collection suggests that beginning with the digital age in 2002 storing of large amounts of data on digital carriers has been more and more popular. Whereas, in the year 2000 the data storage ratio was 25% to 75% in favour of the analogue data carriers. The situation changed promptly and in 2007 as much as 94% of data was stored on digital carriers with only 6% stored on analogous ones. This was also related to the efficiency of storage on considerably lower number of physical media [Hilbert, Lopez, 2001, pp. 61–62]. Such a situation is reflected in today’s reality, when the tendency to store increasing amount of data translates into growing capacity of hard disks and portable storage media, with the price of 1 gigabyte falling from one year to another according to ceteris paribus rule.

The ‘big data’ term relates to the access to vast amounts of data, acquired with the use of e.g. mobile telephony networks, credit cards, administration systems or social networks[Offenhuber, Ratti, 2014, p. 7]. Every person is involved in that kind of activity and the age of those providing data for the system lowers, as, younger and younger people are using devices such as e.g. mobile phones. To put it in plain language, we can refer to Michael Batty’s definition, which implies that big data means the data stock that cannot be held in an Excel sheet [Batty, 2013, p. 274].
Big data can also be explained with the use of the three Vs: velocity, variety and volume, which are supplemented by the other four: veracity, variability, visualisation and value. This means that in the big data era, the data is processed in real time, regardless of their size (velocity and volume), whereas their variety lies, among others, in the lack of coherent structure and format [Van Rijmenam, 2014, pp. 10–13].

The above features are complemented by veracity, indispensable due to the need to use reliable data in the decision-making processes, in particular when these processes are automated and take place without human intervention. The velocity of data collection, also in city systems, entails the variability of data, which, in turn, implies verification of large number of data, which can give different results over time. Better understanding of large data sets requires their visualisation, facilitating data absorption by end users. This feature describing big data may be exemplified by terrain information systems, used increasingly often by Polish cities (more on the subject of utilisation of spatial information systems, see: Marcin Feltynowski in EkoMiasto#Środowisko). Additionally, the data presentation system should be based on infographics, providing an attractive way of presentation of complex contents. The last feature refers to the value of data market, estimated – in European public administration – at 250 billion € per year. It is important, however, that it is not the value of the data itself but rather the value of the data processed into information and often also into knowledge and wisdom, supporting the decision-making processes of local actors, with impact on the financial sphere as well [Van Rijmenam, 2014, pp. 13–14].

Figure 9.1. Popularity of ‘big data’ entry in the Google browser

The figures represent the ratio of total number of queries for the ‘big data’ entry to the total number of Google queries at the same time. The line with downward trend implies downfall in relative popularity of the entry. This, however, does not necessarily mean that the number of searches for this entry has fallen down. It illustrates solely its decreasing popularity vis-à-vis other queries.

Source: Own elaboration on the basis of Google Trends www.google.com/trends, [dostęp 10.10.2015].
For the sake of simplicity, we can say that thanks to big data itself it is possible to conclude that the market of data bearing those features shows an upward trend. The search query into the Google resources relating to the global number of queries for the 'big data' entry suggests that beginning with 2004 till the end of 2011, the term was not very popular and was rarely searched for. Conversely, from the end of 2011 till July 2015, a change in trend was observed with the systematic growth of searches for 'big data' entry. Beginning with September 2012, we can observe a similar trend in Poland too (Figure 9.1). Thanks to Google’s acquisition of data bearing big data features, every user, equipped with Internet access may compile him/herself a number of similar collations for endless variety of entries registered by search engines. Additional conclusions from the analysis allow us to name the voivodships (regions), where 'big data' terms is searched for most often. In Poland these are: mazowieckie, małopolskie, dolnośląskie, pomorskie, wielkopolskie and śląskie.

The growing popularity of the 'big data' term is also linked to the challenges facing the actors utilising the data and those providing them. It should be observed that these challenges in a way result from the very definition of 'big data' and their descriptors. Large data sets require application of new methods of analysis, enabling collation of elements with different structures [Batty et al., 2012]. This, in turn, implies the need to adjust the tools and enabling their efficient cooperation with diverse data sets. Working on large data bases encourages the development of appropriate equipment, enabling drawing up the required conclusions in fast and efficient way. It is important, however, that the data cannot be accessible only for selected users, but for the widest possible group of recipients, who could use the data to generate information.

Data analysis constitutes the final element of big data utilisation, however, one should bear in mind that this area is still developing. Whereas the elaboration of new and more efficient methods for data acquisition and the development of already existing ones poses a considerable challenge for data management professionals. The problem of data storage implies the need to expand data carriers. The assumption behind this trend consists in the conviction that it is better to store rather than delete the data, for the sake of their future, cost-free utilisation. A major problem, however, is the process of data retrieval from large data bases, which, as mentioned before, often have different structures. The efficient search engines, enabling prompt retrieval of required data, are being developed in order to provide access to data for end users. The Internet constitutes the easiest way to provide access to data, however, fast data transfer is required for data deployment to the computers of end users. This forces continuous development of broadband networks and a need for bigger accessibility for local communities. According to Eurostat data for 2014, 71% of Polish households had broadband access to the Internet. The methods for data visualisation pose another challenge, especially that they should not be based exclusively on specialist software but also supported by graphic modules of internet sites exploiting big data [see more: Kitching, 2013; Moorthy et al., 2015].

Big data is becoming an indispensable element of the decision-making process in all spheres of human activity. However, similarly to the knowledge gap diagnosed in relation to the use of terrain information system by public administration, a knowledge gap will also apply to the use of big data. Better developed data mining techniques in correlation with the appropriate skills in hardware and software use should result in positive impact on decision-making processes in public administration. Nonetheless, public administration will have to tackle the problem of necessary investment in new hardware and software, which are indispensable for building competitiveness in the global world.
9.2. Big data in the city – sphere of application

Big data can be acquired from various sources, including Internet, social media, mobile applications, resources of different levels of public administration, commercial data bases, spatial data, results of scientific research or scanned documents, processed by OCR systems. To these we can also add the GPS location data, wi-fi networks mapping or mobile phone information through the use of localisation of the GSM transceiver stations [Van Rijmenam, 2014, p. 6; Mellody, 2014, pp. 4–6].

Thanks to the above sources of data acquisition and the data sets created under this process, local authorities and the public administration officers enjoy better possibilities of big data utilisation in their daily professional practice. This, however, involves the need for constant upgrading of their skills and cooperation with higher education institutions, which is becoming the basis for a complex approach to solving numerous problems facing municipal authorities.

The portal entitled ‘Na co idą moje pieniądze’ (Where does my money go), set up by The Story company with the support of PricewaterhouseCoopers (PwC) and the European Place Marketing Institute, is an interesting example of application of data bases is the public sector. The portal is based on data from communes (municipalities) budgetary resolutions. From the portal, the citizens can learn what proportion of their income tax (communes receive 37,53% of PIT) goes to public roads and transport, education, housing, green area maintenance or public administration. At the moment only 9 communes joined the initiative, yet it is still worthwhile to present the potential which the big data sources carry with the view of providing information for the public on the use of financial resources, derived directly from the local communities.

Appropriate visualisation with the use of infographics and cartograms, also suggest considerable marketing value of big data. Presentation of complex data in a user-friendly form, provides a useful tool, by which public administration offices and local authorities can explain and communicate to the local communities the phenomena and processes characteristic for municipalities. The city of Poznań, used the data and resources of ‘Na co idą moje pieniądze’ portal to promote the city as an attractive tax residence by presenting PIT resources distribution. What is interesting, portals of that kind are the sources of big data themselves. Thanks to the data uploaded onto the portal, their administrators can create a data base devoted to the analysis of gross incomes earned.
Big data acquired in real time may serve the purpose of improving the quality of life in a city. This, however, requires the cooperation of many different public services obtaining the data: e.g. offices responsible for security, transport, technical infrastructures or crisis management. All data obtained from their equipment, including city monitoring systems, facilitate urban management, in particular in emergencies resulting from system breakdowns, accidents or natural phenomena. Big data sets support the decision-making with respect to necessary changes in vehicle and pedestrian traffic required in emergency situations.

Urban management should translate into making the real-time information accessible not only to local authorities but also to other local actors in the city.

The example of London and seven other big cities of Great Britain (Birmingham, Brighton, Cardiff, Edinburgh, Glasgow, Leeds and Manchester) demonstrates that data concerning the weather conditions, air pollution, accessibility of city bicycles or river water levels are made accessible in real time to the local community via an internet portal created for this purpose. In this case the local community is seen as the end user, who can visualise the data accessible in the form of an on-line map.

Data collections characterised by highest diurnal variability is those concerning road traffic. The data is acquired by the use of open data collected from transport companies, providers of navigation systems (such as e.g. TomTom), smartphones, smartcards, used in underground and other forms of public transport, information from public transport providers and taxi companies, traffic census or spatial statistics and information services. All these elements may form a huge data base on transport-related phenomena and their trends, both in diurnal cycle and a longer-term perspective. Such a city transport management platform proves to be an easy to use, productive and efficient solution, and results in increased security of the city traffic. High costs and the need to coordinate the activities of numerous city actors are, without doubt, the factors that affect the feasibility of application of such solutions. Still, cooperation of various actors may contribute to provision of valuable sources of information, which – owing to the analysis of big data – can serve not only the purposes of local governments, but also those of the business sector, through the analysis of traffic flows within a city [see: Morioka et al., 2015].

Big data resources include the maps supplied by commercial providers and the open source map resources, e.g. maps by Google, Bing or the Open Street Map (OSM). In the case of the OSM project, the users often generate data in the form

---

of GPS mapping and subsequently import it into the portal, which enables supplementing or updating already collected data. The data concerning the building development and the land use is accessible on the OSM portal. Thanks to this initiative, created by the Internet community, the local authorities, in particular those of smaller administrative units, can enjoy access to a sui generis geoportal and a collection of spatial data. In the case of commercial portals, additional value lies in satellite images, which can be analysed on the timeline. Thanks to that kind of analyses, the public officers may undertake basic monitoring of the changes in the city space with the use of open data sources.

Social initiatives, undertaken under the SoftGIS methodology are similar in their nature are, and involve simultaneous analysis of ‘hard’ spatial data with ‘soft’ data retrieved from geosurveys. Solutions of that kind enable research of spatial planning, green areas and environment, as well as, other spheres of the city functioning in connection with spatial information. Use of the SoftGIS enables the collection of vast data sets originating from the users of urban space and their subsequent application in undertaking strategic decisions. [Kahila, Kyttä, 2009, pp. 389–391; Feltynowski, 2015, p. 73]. Scientists and NGOs promote SoftGIS-based practices, for example a project by the Sendzimir Foundation ‘Licz na zieleń’ (Count on green). SoftGIS methodology also provides the basis for city portals, enabling reporting problems identified in the city space, e.g. in Gdańsk. Another example of an NGO project is the portal entitled ‘NaprawmyTo!’ (LetsFixIt!), initiated by the Stefan Batory Foundation and involving 13 other organisations. Cooperation with city inhabitants based on the concept of collecting spatial data, allows to build up considerable data bases, and additionally makes it possible for relevant services to monitor the problems in the longer perspective. This can be done with the use of www sites.

Spatial planning is yet another area for big data utilisation in cities. It can be supported by SoftGIS methods in order to implement social consultations. On the other hand, spatial planning can use the sources of spatial data collected in order to establish an infrastructure for spatial information, in accordance with the provisions of the INSPIRE Directive (Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in
the European Community (INSPIRE)). Currently, local self-government can use the big data sources in the form of point clouds, created with the use of laser scanning. LiDAR (Light Detection And Ranging) in photogrametry provides the data enabling the creation Digital Surface Models and Digital Terrain Models. This technology is applied in the planning of routes of linear objects, creation of flood maps, detection of relics of archaeological objects or in the management of territorial units. Similarly, large data sets are produced with the use of orthophotomaps, whereas satellite photographs, having undergone appropriate transformations, may also provide a spatial data set.

The areas of big data utilisation in city management presented above do not exhaust all potential spheres of their use by communes. It should be noted that analytical capacities, supporting the city activities are often located outside municipal offices, namely they are found in local communities. Therefore, an important means of support for big data utilisation is the creation of portals containing open data that can be used by local actors in order to conduct analyses related to the functioning of cities. In such a case, the implementation of open government formula is indispensable, which in the case of city management takes the form of open local government. This idea also translates into openness of data related to self-government activities, which should share the data and information collected in the process of public resources spending. Is should be noted that making the data available by local authorities, constitutes a vital factor supporting operationalisation of the good governance concept via implementation of transparency and accountability rules.

9.3. Evidence based urban planning with big data

The area of spatial planning, through the linkage between the data and local space, is also involved in the concept of evidence based policy, inseparably connected with data, and the big data even more so. The concept of evidence based policy is derived from medicine, whereby the exploitation of newest data and research results in making the decisions concerning patients health was already proposed and practised in the 1990s [Li-Wan-Po, 1998, pp. 57–58]. Beginning with the mid 1990s the concept of evidence based policy has been implemented on governmental level in the United Kingdom. The success of the evidence based approach translated into the possibility of its application on the local level too. It is precisely the domain of spatial planning that enables application of such an approach in connection with big data sets. Spatial policy is that sphere of activity of self-government authorities, which should combine all elements related to the functioning of the city. It is spatial planning and appropriately arranged instruments that make it possible to influence the local actors and to take care of the spatial order and sustainable development. Therefore, we should underline the need to develop evidence based urban planning, as integral part of evidence based local policy.

Evidence based urban planning consists in the use of both the scientific research and suggestions by local residents on the changes required in the city space when drawing up the land-use plans and studies of conditions and directions of land use. It should be underlined also that new scientific methodologies should be developed, the ones that enable clear-cut assessment of land-use plans, as well as the studies of conditions and direction of land use in terms of their feasibility vis-à-vis the potential financial burden to be put on municipalities, investment capacities in the discussed areas and the demographic changes observed in communes. The literature of the subject stipulates that evidence based spatial planning requires the development of tools that allow an involvement of larger groups of residents in
decision-making processes and the use of the data thus collected in the process of land-use planning. Such a method promotes evidence based urban planning based on quantitative data and at the same time encourages a bottom-up urban development approach [Khan et al., 2014, p. 224]. Such a stance results from the fact that, due to its complex nature, spatial policy needs to use various tools in the process of participatory planning.

In the case of spatial policy both the local authorities and the local community should expect of each other to justify the solutions proposed for inclusion in the land-use plans and the studies of conditions and directions of land use. Justifications related to spatial policy should be based on reliable source material, which in turn should be supported by research results and analyses made by experts and researchers knowledgeable in relevant areas. This is all the more important since the results, research and analyses undertaken in the domain of spatial planning, may contribute to solving a number of problems of vital importance for local communities.

As observed by researches, evidence based spatial planning, like other policies applying the evidence based approach, face certain barriers. They note that continually promoted cooperation between business and science communities does not always take place. This is reflected in the divergence of opinions between science communities, policy-makers and practitioners as regards priorities, procedures and work cycles of these parties. This, in turn, leads to the scattering of data and information, which – despite well developed ICT – is difficult to access. Therefore, it is proposed that the so-called knowledge brokers, responsible for building up of cooperation networks between the research communities and practitioners, should operate in order to guarantee the best possible exploitation of information and knowledge [Meyer, 2010, p. 118; Olejniczak, Kupiec, Raimondo, 2014, p. 69].

Numerous attempts at exploitation of big data in spatial planning are observed in urban planning departments of large cities (e.g. the city of Łódź), which can afford to undertake on their own the laser scanning procedures, the results of which are subsequently used in city planning. 3D images, presenting city areas have been popularized by such applications as: Google Earth or Bing Maps. Owing to the capacities of those portals, the planners recognised the value of laser remote sensing. Analysing the data with the view to establish land use plans allows the generation of vector and raster maps from point clouds. Indisputable benefit of laser scanning is that it provides the possibility to verify urban layouts or the visibility of single buildings or developments in the city space. This approach also enables the presentation of variant solutions for land use plans, of particular importance for the process of consultation with local actors so that the land use plans are drawn up in accordance with the requirements of the local community. The changes proposed to the Spatial Planning and Land Use Act and several related acts (amendments to the act currently under legislative procedures, as of 29 July 2015) also follow this direction: it is stipulated that the residents should be consulted at the stage of planning, prior to elaboration of a draft land use plan. Regardless the legislative provisions, it should be noted that the use of laser scanning data and visualisation of new elements in space, allows to maintain a dialogue between the local community, authorities and the authors of the land use plan [Tress, Tress, 2002, p. 74]. It is worth noting that exploitation of big data and projects’ reliance on hard evidence related to space visualisation should allow an easier implementation of solutions proposed in land-use plans in future.

A number of Polish cities are faced with the problem of inadequate land management, especially of areas intended for urbanisation. Many companies implemented projects meant to support the management of city space in Poland. The
projects attempts to designate potential areas for urbanisation on the basis of spatial data and analyses with the use of GIS applications. The workflow consists in the use of spatial data in a few consecutive phases, resulting in the designation of areas intended for: restricted expansion of development, non-restricted expansion of development and those areas, where building development is inadvisable. Such an effect can be obtained by superimposition and analysis of data related, among others, to: existing building developments, land use, road network, water supply and sewage collection networks, together with the indicators representing the development percentage and the intensity of development. Existing conditions related to anthropogenic transformations are enriched with data sets, making it possible to define the environmental threats caused the planned development and vice versa: the hazards on the side of the environment that the prospective developments may be exposed to. The users of this method use, among others, the information on open terrains, environmental limitations, cultural heritage, areas exposed to flooding, landslides, mining areas, protection of agricultural land and forests, conservations areas or the terrains, where founding of buildings is difficult. All mentioned elements come from data bases, collected by the sectoral entities of public administration. The data obtained during studies provide evidence for the local authorities and the land owners, suggesting the desirable direction of city development taking into account the current environmental conditions and land use. Analyses allow a clear cut justification for inclusion or excusing of given areas from building development.

<table>
<thead>
<tr>
<th>Use of big data for 2D and 3D imaging in the domain of spatial planning</th>
</tr>
</thead>
</table>

![Part of Jersey City visualised thanks to vector and raster laser scanning (A) and in the form of 3D point cloud (part of image A) (B).](image)

*Source: own work based on LiDAR data from U. S. Geological Survey resources visualised with the used of QGIS software (image A) and the Furgo Viewer (image B).*

Exploitation of big data in spatial planning is more and more common. We should develop methodologies, which allow the local actors to understand which solutions are the best from the point of view of spatial order, sustainable development, as well
as, socio-economic development of the city. All this should be backed up by scientific research. Its findings should be applied in various aspect of functioning of territorial units. It is the local authorities which should invite experts to provide scientific support and build proper relations with scientific centres. The finding of scientific research will provide well documented bases of decisions undertaken by the local authorities. This will encourage the concept of evidence based practice, which is to enable a better understanding of processes taking place in the cities.

**Bibliography**


EkoMiasto#Zarządzanie
Zrównoważony, inteligentny i partycypacyjny rozwój miasta

pod redakcją
Zbigniewa Przygodzkiego