In Search of Lost Data in the Era of Data Driven Planning

A while ago a very interesting Global Data Barometer was published, and it was great to see that Estonia ranked 2nd in the world. GDB is an expert survey conducted in 2019- 2021 observing data governance, capability, availability, and use and impact of data for public good. That is indeed excellent news in digitalizing world where we are kept told that *data* is everything.

However, from the urban scientist's perspective the situation is a bit more complicated and perhaps not as optimistic as the above survey hints. Here I discuss about some implications of datafication along with current situation of data use and availability from the viewpoint of scientific research and science-based urban planning I would like to elaborate here.

Data Driven Urban Planning

In marketing speech, data is today often sheeted with overwords: data is a merchandise but also a snake oil: it provides solutions to current problems from climate change and traffic chaos to terrorism and obesity. Data as such, however, just refers to any unorganized information; essential is that it must be organized to provide any useful knowledge for the issues to be solved, or for example those emerging in cities and urban planning. Those are particularly crucial for the future of humanity due to the ongoing planetary urbanization.

The highest hype naturally concerns so-called big data, which consist of the vast chaotic sea of information including anything we do or have ever done with our digital gizmo, gathered and stored by AI. The amount of this data is enormous. Globally we produce data 2.5 *quintillion* bytes per day: this means 2,5 with 18 decimals. Big data hence underlines the necessity of data organization: it is evident that, for example, the data our smart phones stores make no sense without sensible question. The question depends on the reason for which we need the particular knowledge. Films we have watched on Netflix provide knowledge of which movies to recommend to us, while our doctor's appointments offer other kind of knowledge about us. The appropriate classification of data – i.e. miscellaneous information - is the key for its further use.

Traditionally in urban planning data could be fetched from (physical or electronic) archives, regarding variety of maps, census data, traffic, and alike. For urban planning, both big and traditional data are important. Particularly the combination of two can, at least in theory, revolutionize planning. It could become possible to analyze and understand the whole spectrum of our lives, patterns in mobility and consuming, and choreographies of daily life; changes in these and the relationship between our behavior and environment, economics and society, considering this knowledge in urban planning for better quality of life. How did different groups change their behavior during covid pandemics certain times? Does the economic diversity of everyday environment affect consuming behavior and how? Are the educated people moving to certain places more than before? What kind of mobility patterns can be found among car owners, and do the new bus/bike options change these? The list goes on.

The order of day is hence *data driven planning* – planning that is based on data, or more likely, knowledge gathered by analyzing the data. Starting from 1990s, we have made huge leaps in fields of technology for storing data from digital twins to Building Information Models (BIM), not to mention advances in algorithmic architecture tools capable of dynamically produce imaginary built forms adjusting to environmental factors. Furthermore, computational simulation models provide us dynamic methods for realistic what-if scenarios of future cities based on data, helping us to sketch future directions for urbanity. All these tools are relying on myriads of different types of (real-time) data.

Evidently, a specific, meaningful question again must precede data analyses: which phenomenon are we scrutinizing? For example, research of thermal comfort concerting wind in urban environment we need data of wind directions and strengths and existing buildings masses. We can then experiment with new structures steering wind and improving comfort on the streets and squares. Similarly, such a model can concern traffic flows, daily rhythms of users, solar angles or other natural circumstances, or visibility axis. The needs create requirements for the data.

ABOUT OPENNESS AND 'OPENNESS' OF DATA

Availability of appropriate data is therefore essential, and the concept of *Open Data* has become prevalent in digital governance globally. Open data refers to freely available data online for anyone to download, modify, and distribute without any legal or financial restrictions. Open data is by definition available and free; accessible and modifiable; legally reusable and redistributable; and unrestricted regarding further use, modifications, and sharing regardless use (including commercial use).

For data open to this extreme extent, privacy becomes an issue: open data does not naturally include personal data about individuals. When making data openly available one must consider the safeguarding of information, obtaining consent and the secondary use of identifiable information, and how identifiable information is handled during data linkage. In EU accessibility to open data is promoted strongly. Set for this purpose, the *Data Governance Act* will increase trust in data sharing, strengthen mechanisms to increase data availability and overcome technical obstacles to the reuse of data. DGA should steer public and other data holders in the frame of local laws. Underlying idea is that open data benefits society and economy in many ways. Particularly it is suggested to promotes collaborative research and increase its reproducibility and integrity.

However, requirements of maximal openness while maintaining individual privacy, and requirements for innovative research producing new knowledge often collides in regard to what data is needed for answering a unconventional research question. Even in data drive planning, research cannot start from 'what is openly available' but from the well-defined lack of (scientific) knowledge about a certain phenomenon. From the urban science point of view, the question is how can we get the data we *need*, not just the data that happens to be online? Essential is that not all data can be open, but it should be still available for research. This is not always the case.

Some time ago I talked with a big data owner wondering whether it would be possible to get for a very restricted research purpose cell phone data about anonymous users with for example age or gender. The response was terrified – of course not since it would enable *profiling* of users. I was dumbfounded. Profiling – i.e. classifying data according to features – is naturally the key to make big data usable in both urban research and planning. It would be impossible to study, for example, the safety of the children's routes to school, or moving patterns of retired people if our data covers everyone from babies to bedbound patients and marathoners. The dichotomy becomes palpable: sometimes researchers seem to have either (very general and aggregated) open data, or no data.

CONTRADICTION IN HOW WE THINK ABOUT BIG DATA

Regarding big data we are – typically to our species – blind to the *general*, and considerate with the *particular*. We are emotionally very concerned about our privacy – and for a good reason. Who would want to give away our personal history we constantly collect with internet-connected mini-computers attached to us to potentially hostile parties!

Hence we jump to a conclusion and also strongly deny the restricted, monitored use of our anonymous tracking data by a certain university researcher for scientific purposes, regardless the fact that this data is

very coarse due to the behavior of the signal jumping between masts, and researcher being responsible and signing all necessary confidentiality agreements including sanctions. Simultaneously we happily ignore the underlying double standard and grant permissions to myriads of foreign applications that collect every millisecond much more detailed GPS data and combine it with our and our contacts' data, our photos, browser and purchase history, microphone and camera – and what is more, permission to sell this all to unknown third parties. (And yes, here the *no-data* becomes available for some).

Consequently, by doing so we lose the opportunity to get practically free scientific knowledge for building better cities *along* with losing our privacy – although in exchange we get of course access to funny applications making it possible to attach dog ears to our profile picture. Who would refuse from such a deal!

TRADITIONAL DATA AND DATA TRADITIONS

Public actors also provide 'open data' collected using taxpayers money. This data has been the primary source of data for researchers for over a decade now. One could assume this type is the most available kind, and once paid, also free. However, many public actors provide their (very general) open data online, and make a nice profit with selling the more detailed, applicable data to researchers. Such general open data online is very useful in open data -seminars in which it can be used to show that it exists. For serious research it needs more imagination to find meaningful applications for e.g. average income by district, and local professional planners already know such things by heart.

To an extend the data classification and aggregation is just coarse but sometimes it astonishes researcher: what would be the intended purpose of a data set, in which residential and public building are grouped as one class. I however accept the fact that some questions just retain unanswered.

With the right mind set genuine Data Driven Planning could enable much more with just a bit more detailed data. It could be possible to discover groups of urban blocks with high level of socio-economic or functional diversity increasing resilience and vitality, or those in which small maneuvers would help them to flourish, or areas where progress has just started to go wrong, and it is still possible to fix it. The possibilities are endless if we start from the problem, not from the data.

I must stress that raw data violating any privacy is very rarely necessary in research, and should not be openly available – but the reasonable level of aggregation and easy availability of data is something to pay more attention to.

Four years ago colleagues reported about the local challenges of the DDP the following:

"Lack of political will, fragmentation of data ownership, difficulties in machine readability, limited access to data, lack of standards and low data quality." Kalvo et al 2018

Despite the nice Global Data Barometer rank, to my personal experience, these unfortunately appear quite similar.

WATCH YOUR ATTITUDE!

The big challenge in making data not just 'open' but truly available for innovative research is in attitudes. For data providers, openness should be a common mindset, not just a directive and a gold standard that looks good on paper. We seem to not yet have a profound collective understanding of where and why data could be applied in a way that benefits everyone in the society – not just state's sometimes appearing as

data museums, commercial actors buying and selling it behind the closed doors, or dilettante researchers scrambling 'open data' online for fun.

Particularly we often lack understanding of how data could be used in urban planning, and research preceding it carried out by researchers and professional planners. Contemporary cities *cannot* be planned unaware of what is going on right now, and how processes and their spatial configurations have changed in time. How can we guide its evolution blindfolded without knowing where different people live and work, where different types and sized firms and services are located, and about their transition. Cities cannot be planned for common good if the urban dynamics and mechanisms are obscure.

It is good to mote that promoting 'common good' does not imply adopting a socialist stance. Market liberal city nevertheless requires certain level of steering, of which the private actors are well aware. It has been known for over a 100 years that the location decisions of firms are never based on complete knowledge, and individual decisions cumulate in a surprising ways in a spatial city, creating for example clusters of certain industries. Companies have no intention or resources to follow up such overall regional location dynamics – nor is it their task to do. Overall, externalities, i.e. unintended consequences of an activity to other parties, may induce nasty side effects for firms from non-steady land prizes or accessibility issues to lack of educated work force or societal turbulence. Understanding of how urban complex systems operate is necessary to avoid these. Already Friedrich von Hayek, the father of neoliberalism, acknowledged this: urban planning is necessary for stabilizing chaotic forces operating in urban dynamics. For companies, loose guidance would be the most preferable, enabling the gravitation to the best location while ensuring continuous urban processes. To build such understanding, a lot of urban research is needed, which today requires availability of volume of different data – both big and traditional.

Afterthoughts

To conclude, on the one hand we have all the data we can imagine, policies encouraging its sharing and tools for planning and design. On the other, our mind set is still to an extend tuned with the 'old world' implying archives, specialist gate keeping, and data requests following chain of command. However I am confident on that once the future generations genuinely internalize the profound significance of variety of data and its potential in urban research and planning, possibilities would open to imaginary futures and dynamic design tools to build fascinating, viable and sustainable cities yet unseen. Data, which is actually just the representation of all environmental factors, will then be able to fully play a role in the analyses, digital models and planning, and steering of dynamic urban systems for better cities of the future.

***

Some literature

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in human geography*, *3*(3), 274-279.

Fujita, M. (1989). Urban economic theory. *Cambridge Books*.

Kahnemann, D. (2017). Thinking, fast and slow.

Ogulenko, A., Benenson, I., Toger, M., Östh, J., & Siretskiy, A. (2022). The fallacy of the closest antenna: Towards an adequate view of device location in the mobile network. *Computers, Environment and Urban Systems*, *95*, 101826.Kalvo et al

Renee Puusepp and Raul Kalvo (2018): Data-driven spatial planning. https://inimareng.ee/en/data-driven-spatial-planning.html. Accessed on 15.2.2022

Internet links

Global Data Barometer: https://globaldatabarometer.org/

EU Data Governance Act: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN

Infographic: how much data is produced every day? https://cloudtweaks.com/2015/03/how-much-data-is-produced-every-day/

Future of Architecture? Minimaforms: Stephen Spyropoulos & Theodore Spyropoulos - All is Architecture https://www.youtube.com/watch?v=FbUU8SsrdU8