



BIROn - Birkbeck Institutional Research Online

Tarantino, M. and Lokot, T. and Moore, S. and Rodgers, Scott and Tosoni, S. (2019) Urban data cultures in post-socialist countries: challenges for evidence-based policy towards housing sustainability. University of Geneva, Geneva, Switzerland. (Unpublished)

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/31175/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

BIROn - Birkbeck Institutional Research Online

Tarantino, M. and Lokot, T. and Moore, S. and Rodgers, Scott and Tosoni, S. and Hedjazi, A.B. (2019) Urban data cultures in post-socialist countries: challenges for evidence-based policy towards housing sustainability. University of Geneva, Geneva, Switzerland.

Downloaded from: <http://eprints.bbk.ac.uk/31978/>

Usage Guidelines:

Please refer to usage guidelines at <http://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Urban Data Cultures in Post-Socialist Countries: Challenges for Evidence-Based Policy towards Housing Sustainability

Research Report

Matteo Tarantino,^a Tetyana Lokot,^b Scott Rodgers^d, Simone Tosoni^e, Susan Moore,^c Alexandre Babak Hedjazi^f

^a Department of Communications, Università Cattolica del Sacro Cuore di Milano, Italy; Institute for Environmental Sciences, University of Geneva, Switzerland. matteo.tarantino@unige.ch

^b School of Communications, Dublin City University, Dublin, Ireland. tanya.lokot@dcu.ie

^c Department of Film, Media and Cultural Studies, Birkbeck, University of London, London, UK. s.rodgers@bbk.ac.uk

^d Department of Communications, Università Cattolica del Sacro Cuore, Milano, Italy. simone.tosoni@unicatt.it

^e The Bartlett School of Planning, University College London, London, UK. susan.moore@ucl.ac.uk

^f Institute for Environmental Sciences, University of Geneva, Switzerland. Alexandrebabak.hedjazi@unige.ch

INTRODUCTION.....	2
<i>Localization and Data Cultures</i>	3
<i>Elements of Urban Data Cultures</i>	4
<i>Why Data Cultures Matter</i>	5
CASE STUDIES.....	5
<i>Rationale and Methodology</i>	5
<i>Case 1: Albania</i>	6
<i>Case 2: Georgia</i>	8
<i>Case 3: Kyrgyzstan</i>	10
<i>Case 4: Ukraine</i>	13
DISCUSSION.....	15
<i>Marketization and Decentralization</i>	16
<i>Dynamics of Knowledge and Expertise</i>	18
<i>Challenges to Building Data Infrastructures</i>	19
CONCLUSION AND RECOMMENDATIONS.....	20
<i>Dimensions of Data Cultures</i>	21
<i>Recommendations</i>	23

Introduction

The use of indicators, based on innovative or previously untapped data sources, has been identified as one of the most promising ways to monitor progress towards the Sustainable Development Goals (SDGs), agreed under the UN’s 2030 Agenda for Sustainable Development (SDSN, 2015). At the same time, cities and urban local bodies are arguably the most immediate and direct locus for integrating evidence-based policies towards achieving the SDGs (UN-HABITAT, 2018). Thus, on the one hand, there is a need to develop standardised sources of data, to allow for national and international benchmarking. Yet on the other hand, the idiosyncratic nature of cities, and the vast number of actors involved, means that data management often happens in ways that are highly localised. There are, in other words, geographically-varied ‘data cultures’. In different countries, regions, cities and localities data is produced and used according to variegated representations, values, norms, epistemologies, practices, infrastructures, modes of operation, standards, communication styles, artifacts, economies and power structures.

This Report summarises the insights of a preliminary study into the urban data cultures of four post-socialist UNECE member states. Our focus is on evidence-based policymaking, specifically in relation to SDG 11.1 (ensuring “access for all to adequate, safe and affordable housing and basic services and upgrade slums”). By ‘urban data’ we mean various sources of qualitative and quantitative information about urbanization processes, which can to varying extents also be utilized for monitoring trends toward housing sustainability, be they at local, regional, national or global scales. We are interested in urban data cultures related to

housing specifically as it is an area less studied in the literature, compared with other urban data such as environment, transportation or security. Housing is a key resource for the welfare of populations, and housing information is one of the oldest and most important forms of urban data. Moreover, as a policy domain, housing is difficult to isolate from other aspects of urban welfare such as planning, infrastructure, education and health. As such, it is an appropriate lens through which we can observe historical transformations in urban data cultures more generally.

Localization and Data Cultures

One of the most significant transformations affecting the kinds of urban data available to policymakers today is the growing ‘digitization of location’ (Wilson, 2012: 1266). While in the recent past, digital information was often seen to constitute a ‘virtual’ space, separate from physical location, today spatial location and digital information are increasingly joined together. Graham et al. (2013: 465–466) suggest this is the result of three successive developments. First, the rise of mobile internet connectivity, through smartphones and other devices. Not only does this mean that users can access online information on the go, but also, that geospatial data related to the real-time location of such mobile devices can itself become a form of urban data (Ratti et al., 2006). Such data about mobile users, for example, could be used as a proxy measure of housing density. Second, there has been an expansion of user-generated contributions, and specifically volunteered geographic information. Through a popular platform such as Google Maps, citizens are increasingly acting as digital sensors, volunteering their own spatial information to larger platforms (Goodchild, 2007). Finally, there has been a rise of what is sometimes called the ‘geoweb’, referring to how web content is increasingly being coded, often automatically, to specific geographic locations.

These developments point to the ways in which urban data is becoming more and more ‘local’ from a technical point of view. However, a growing number of scholars in fields such as science and technology studies (STS) and human-computer interaction (HCI) are making an additional, stronger claim related to the localization of data. This is that *all* data are local, and not just data coded structurally to geospatial locations. There is increasing recognition that all forms of data “are cultural artifacts created by people, and their dutiful machines, at a time, in a place, and with the instruments at hand for audiences that are conditioned to receive them” (Loukissas, 2019: 1-2). This includes geospatial and locative data, which are themselves produced in the specific contexts of local data cultures (Smith, 2019; Wilken, 2019). These deeper claims, emerging out of STS and HCI, have been very important for the approach we are taking in this study. While we remain interested in the geospatial locality of urban data, what is most important to our understanding of data cultures is investigating more generally the local patterns through which data are crafted, with what kinds of technologies, towards what audiences – and in turn those audiences’ capacities or readiness to use or understand such data.

Thinking about the locality of all data draws more attention to the fine-grained geographical specificities of data cultures. These ‘cultures’ of data (Bowker 2000: 635) are not to be confused with national or ethnic cultures. Rather, they refer to how local, regional and national norms, perceptions, practices, institutions and infrastructures both explicitly and implicitly frame how operators and policymakers make sense of data. Data cultures in other words are the ‘operating system’ of localized data-related practices, including those relating to our area of interest, housing data.

Elements of Urban Data Cultures

Based on the above conceptual approach, we outline four main elements of urban data cultures that have informed our inquiries in this pilot study:

1. Epistemologies and Representations

We are interested in how local actors know or understand ‘data’, in particular in an age when there are so many potential sources of data than can be tapped. Datasets were once a much rarer resource, a specialized type of information created by, for example, government departments and statistical agencies (Kitchin, 2014). Today, due to a series of technological changes, data is being collected from and generated through a much wider range of sources. This potentially means that the term ‘data’ is getting more and more amorphous, or being deployed with varied meaning. Communicating and representing such diverse data also poses a challenge to governments. Recent years have seen the introduction of well-designed and aesthetically attractive government data portals. But while these clearly make data more accessible, they can further downplay the distinct sources of such data, and also exaggerate their authoritativeness (Ratner and Ruppert, 2019). In exploring how actors on the ground understand data, we have avoided using the much-hyped term ‘Big Data’. This term is misleading in that what has changed in recent years is not so much the size of datasets, but the ease of searching, aggregating and cross-referencing a growing range of data, often with comparatively rudimentary computer hardware and software (boyd and Crawford, 2012). Our focus then is on the localized, contingent perceptions and representations of data as sources of information and measurement in housing policy and related fields.

2. Practices and Norms

Another key interest in our inquiries has been exploring what local actors do with data, what they might take for granted about how data is handled, and who does that handling. This first of all means gaining an understanding of the everyday routines and understandings of actors working with different kinds of data, within specific organizational or bureaucratic environments. Often, such routines and understandings can be partly explained by understanding organizational, local and national histories. Beyond everyday routines and understandings of data, it is also important to investigate the broader norms governing how the generation of data is understood and valued. In many bureaucracies, practical norms can amount to a kind of informal regulation, which structures behaviour and decision making (Olivier de Sardan, 2017). For the post-socialist contexts in which we are studying, one key interest we have is the changing norms around who can and should produce data related to housing – for instance state agencies, private or market actors, civil society, etc – and the degree to which norms around the ownership of data and who might rightly access or use it go unquestioned.

3. Formats and Infrastructures

It is also important to study how data is formatted and stored locally, and shared and circulated across broader geographical scales. Data which is useful and valued in one place may not be compatible with

national data, or data from other localities. This may be because it is stored on paper hardcopies, or in non-machine-readable digital formats such as PDF files. There may also be barriers to sharing and circulating data due to the nature and extent of related infrastructures. Data infrastructures are “the institutional, physical and digital means for storing, sharing and consuming data across networked technologies” (Kitchin, 2014: 32). The design, availability or existence of, for example, dedicated data transfer systems, repositories and portals can, given the right conditions, act to bolster, sustain or transform a data culture. Yet broader digital infrastructures need to be considered as well, such as the geographical distribution and accessibility of high-speed internet connectivity or data processing centers.

4. Knowledge and Expertise

A final element of urban data cultures is the extent and nature of data-related knowledge and expertise. A longstanding problem associated with the development and expansion of information and communication technologies is information overload, and the need to develop new kinds of digital literacies to deal with this problem (Edmunds and Morris, 2000). If the growing availability of diverse data is going to empower actors besides large data analytics firms and private companies, there is therefore also a need to develop more ‘grassroots’ forms of ‘data literacy’ (D’Ignazio and Bhargava, 2018). It is crucial therefore to understand the existing capacities of local data-related knowledge and experience, and the opportunities as well as barriers to its development.

Why Data Cultures Matter

These four broad elements of urban data cultures are a heuristic lens guiding our research inquiries. Later in the Report, we build on these elements and detail some more specific observations of the urban data cultures we have observed in the case studies outlined below. Understanding data cultures is of considerable importance for designing guidelines for ‘good urban data’, especially those oriented towards cross-national sharing of data. Efforts to change or develop data practices will likely struggle unless overarching data cultures are addressed. Cross-national sharing is particularly sensitive for transitioning countries (as well as cities and regions) requiring the assistance of transnational bodies, which require some degree of alignment with international frameworks such as those offered by the UN’s SDGs. As yet, however, those countries are amongst the least studied in terms of data cultures: sociological approaches to data have so far focused on the more developed nations.

Case Studies

Rationale and Methodology

For this pilot project, the UNECE member countries of Albania, Georgia, Kyrgyzstan and Ukraine have been selected. The countries were considered fertile grounds for our research: they are all post-socialist nation states (Albania is a former independent socialist republic, and the other three are former republics of the Soviet Union), showing clear but also differentiated patterns of

urbanization and housing-related vulnerability, and presenting possible leapfrogging opportunities in the area of urban data management. From a practical point of view, the countries are also all involved in a pre-existing UNECE initiative about housing data. This included four National Conferences in 2019 on evidence-based policies for sustainable housing and urban development, organized by the UNECE Housing and Land Management Unit, in collaboration with UN-Habitat and local partners in each country.

As our approach in this pilot research is to understand how urban data cultures related to housing are localized (Loukissas, 2019), we have taken a qualitative approach that utilized the four National Conferences as sites of field research. That is, in addition to contributing our expertise to these conferences, we have used the opportunity to access national and local public officials on the ground, and more generally develop a preliminary appreciation for the current challenges facing housing and related professionals in each country.

Our main method was interviews with officials engaged in housing policy and management at the central, regional and local levels, undertaken in the field, and in some cases remotely, of which we conducted 24 in total (4 in Albania, 5 in Georgia, 12 in Kyrgyzstan, 3 in Ukraine). We also collected notes from a series of informal conversations, as well as observations made at the workshops and visits to state departments, balancing the overall quantity of material for each country. Finally, we have consulted a range of background policy reports, journalism and academic literature related to the case study countries.

Case 1: Albania

In 2018, Albania's population – 2.86 million people on a landmass of 28,748km² – has been declining for the third consecutive year (-0.3% between January 2019 and 2018), due to the falling birth rate and emigration. Yet the country is on a trend towards greater urbanization, especially due to internal migration to the main Albanian cities. The capital Tirana, in particular, has grown from 250,000 inhabitants in 1990 to between 820,000 (according to official registers) and 1 million (estimated actual population) today. Internal migration is tied to two main phenomena. First, it is tied to a widespread diffusion of informal settlements at the borders of urban territories, which started in 1992 with the fall of the communist regime and the subsequent liberalization of internal mobility, peaking around 2005-2006. The housing quality of these informal settlements varies, ranging from simple tents (especially for vulnerable populations like Roma) to conventional house structures. Second, internal migration is tied to a lack of monitoring procedures on the first phases of the process of migration, and to a persistent difficulty in keeping track of internal mobility. This is due to the existing local registration law lacking efficient enforcement measures, and to several factors discouraging registration in the destination municipality, for example more favourable welfare conditions in the origin municipalities.

The fall of the state socialist regime has created a rift in Albanian data culture, with fine-grained state control of internal mobility giving way to a generalized lack of monitoring, both on people's mobility in the territory and on the small enterprises that have responded to emerging housing needs. Furtherly deepening this rift, data collected before 1992 are no longer publicly or institutionally available, reportedly due to the destruction or loss of state archives, including the Real Estate Registry created under state socialism. In the context of this data vacuum, in 2012 the Albanian government restituted houses expropriated under state socialism to whomever could prove their property rights. As a result over 5000 family units, mainly elderly people, were forced to relocate. According to our respondents, in this data vacuum, several disputable records were accepted as original.

With increasing attention to a series of vulnerable populations (Roma, Egyptians, the elderly, disabled people, victims of human trafficking, unemployed people and young couples) the urgent need for more systematic gathering and processing of housing data has gained a high profile at the local and national levels of government in Albania. These needs relate in particular to: strategic planning for sustainable development; the planning, deployment and day-by-day operation of welfare housing initiatives at the municipal level; and the problem of informal settlements, requiring case by case assessments of both housing quality and position, to decide whether built structures should be legalized or demolished. These challenges have only been taken on in recent years by different levels of the Albanian bureaucracy, and respondents described their endeavours to us as tentative. In the present phase, actors involved in the process have gained a clear understanding of the different layers of complexity hindering a systematic and integrated program of housing data collection, gathering and processing. They are still in the process, however, of identifying viable long term strategies for its deployment.

There is a plurality of actors currently involved in collecting data, directly and indirectly (from other institutions), on housing in Albania. At the local level, the key actors involved in direct data collection are municipalities, the private sector, and the third sector. Albania's 61 municipalities not only differ in shape, resources, personnel and know-how, but also in their data needs, depending on their territorial specificities and the different forms of welfare housing programs they have in place. In bigger municipalities like Tirana, data collection tasks are distributed between departments with different functions: public buildings maintenance is for example independent from the Housing Policies department. While most housing data is collected in the routine operation of local departments (e.g. issuing building permits, monitoring welfare initiatives), larger municipalities have their own statistic departments which focus on socio-demographic data. Independent private sector actors (with no form of centralization) such as real estate agencies directly collect data on local housing market trends, and the quantity and quality of available houses. Third sector organizations and activists collect some of the most detailed data available related to specific vulnerable populations. Roma associations, for example, figure among the more proactive: they play a key role not only in the census of the Roma population, but also in

the discussion of national budgets, the localization of national indicators for access to welfare housing initiatives, and the density of informal settlements.

At the national level, there are three key public bodies directly and indirectly collecting housing data. First, the new State Agency of Cadaster, which since March 2019 has taken over the responsibilities of different agencies to produce coherent and complete data about immobile properties across different areas. A new law also prescribes the online registration of immobile properties, promoting the digitalization of data, even if a complete and updated picture of housing is still to be achieved. Second, INSTAT collects data through a network of local offices, and through a decennial census (the next taking place in 2020-2021) for the production of more fine-grained datasets. Third, there are other government departments and units with specific responsibilities for housing, such as the Housing Unit within the Ministry of Finance or the National Planning Department within the Ministry of Infrastructures, that indirectly collect data, mainly from municipalities. While all these institutions disclose some data to the public (in different formats including maps, publications, or websites), only INSTAT has a designated responsibility for direct data collection.

There is considerable heterogeneity in the housing and related data collected by the various actors outlined above. Data varies in their subject (depending on specific needs and aims); procedures and methodologies of collection; format (with many documents still produced and archived in paper); and ontologies (with different databases rarely interoperating). This heterogeneity is especially seen at the local level, where procedures vary considerably, depending on resources, commitment, and know-how of personnel. In addition, there are difficulties in data exchange among institutions. Informants both at the local and the national level reported difficulties in accessing data gathered by the private sector in particular, notwithstanding their key relevance for budgeting and managing housing welfare initiatives (e.g. the average building cost or rent cost per square meter is often informally deduced through the public websites of real estate agencies).

Notwithstanding the difficulties outlined above, housing data gathering and monitoring is at present regarded as a strategic priority by stakeholders at all levels, and there are strong signs that housing data cultures are changing for the better in Albania.

Case 2: Georgia

Georgia, while a relatively small state with a population of 3.7 million on a landmass of 69,700km², has a fairly young population and is seeing growing urbanisation, with an urban population at just under 60% in 2019, and projected to grow to 72% or more by 2050 (UN DESA, 2018b). Up to 97% of the housing stock in Georgia is privately owned and home ownership is exceptionally high, with the rental market mostly reserved for tourism. At the same time, because of privatisation and

deregulation, there is little data on the conditions of most older housing stock constructed in Soviet times, or on construction standards of new urban housing.

On the state level, housing policy, and housing data collection and management, is dispersed across multiple agencies. Key actors mentioned in our interviews include 1) GeoStat, the National Statistics Agency, responsible for collecting data on housing and households; 2) The Ministry of Regional Development and Infrastructure, involved in policymaking, development and planning as well as coordinating local data collection; 3) The Ministry of Labour, Health and Social Affairs, tasked among other things with collecting data on homeless persons and internally displaced populations, which feeds into public housing and affordable housing provision; and 4) The National Agency of Public Registry (part of the Ministry of Justice), responsible for maintaining a publicly available registry of immovable property ownership in the country that relies on data from the cadastre and from private entities. In addition, there are also local actors collecting housing data in the 68 municipalities across Georgia, as well as civil society organizations and business entities which have accumulated various relevant datasets.

The key challenges and issues identified by our respondents with regard to housing data reflect the transitional nature of housing policy and urban data culture in Georgia, turning it into, in the words of one respondent, “a living lab of [urban] policy development”. The post-Soviet legacy of ageing housing stock and loss of many housing records, coupled with rapid privatisation of urban housing, has made it difficult to preserve and maintain a comprehensive set of historical housing data in the country. There is increasing variation in populations who qualify for affordable housing or are otherwise in need of housing support, such as internally displaced persons in some of the areas bordering occupied territories, who often live in informal housing settlements.

The distributed nature of housing governance among different state agencies and further decentralisation of data collection also present a challenge in terms of cross-agency communication and national-local data exchange. Respondents noted that there is no comprehensive framework for consolidating and communicating various elements of housing-related information and statistics, and that an inter-departmental agency dealing specifically with housing issues and monitoring the evidence would be useful. There is currently an interdepartmental working group within the Open Government Partnership framework which is working on an action plan, which involves among other things considerations of how housing data is collected, managed and made available.

Housing data collection by municipal authorities is currently highly heterogeneous: data is collected in multiple formats, to multiple standards and has proven difficult to analyze. Another issue is that definitions of key housing data indicators also vary wildly between municipalities. For example, there is no definition of ‘homelessness’ that is shared across municipalities, so different locations compile their data on homeless populations according to their own context and

understanding of the concept. This makes it difficult for national agencies to analyze the collated datasets, and to advise on the different housing provision initiatives and policies in the municipalities. Respondents indicated that there were attempts to standardise data collection practices and to empower municipalities to collect quality data, but that these required resources such as standards of data processing, robust indicators and definitions, and improved communications between national and local levels.

In terms of technology and digitization, Georgia has made significant progress in digitizing available housing and land data. Initiatives include: a number of urban datasets being made available on the state open data portal; GIS mapping projects in collaboration with the Norwegian Mapping Agency; and a publicly available online registry of immovable property maintained by the National Agency of Public Registry. Larger local municipalities, such as the Tbilisi Municipality, have excelled at creating municipal open data portals that pool available urban data from various sources. However, this openness is uneven, as not all municipalities have the capacity and expertise to do the same. Even national agencies, such as the National Statistics Agency, have difficulty attracting and retaining data scientists and other relevant experts to work on data management and analysis, as commercial jobs in this area offer more competitive remuneration. Despite this, respondents were generally positive about the prospects of digital technology in fostering more homogenous housing data practices, and discussed adapting approaches or platforms developed in one municipality to other locations, in order to standardise data collection and to make the resulting datasets more accessible to everyone.

Despite the challenges and issues discussed, respondents in our research remained hopeful about the development of sustainable housing policy and fostering a shared urban data culture in Georgia built on openness, interdepartmental collaboration and shared standards of data collection. Such a data culture would be informed by local contexts and a common understanding of the meanings of key data signals. Supported by technical innovation, it would in turn inform international SDG indicators and create more opportunities for delivering sustainable housing to those who are most in need of it.

Case 3: Kyrgyzstan

Kyrgyzstan's population today (about 6.2 million on a land mass of 199,990km²) is predominantly rural, though UN projections forecast a majority urban country by 2050 (UN DESA, 2018). With this trend a number of housing-related issues are coming into view, perhaps most visibly in informal settlements (*novostroiki*) around the urban centers of Bishkek and Osh, resulting from internal rural-urban migration. Other prominent issues include: ageing 1950s/60s Soviet era multistorey apartments, and the challenge of professionally managing and improving such housing stock, particularly towards better energy efficiency; affordable housing needs being met through a privatised housing sector; a lack of affordable mortgage lending options towards homeownership;

and high levels of external labour migration (especially to Russia and Kazakhstan), and the resulting impact of remittances for housing rents and prices, particularly in Bishkek and Osh.

Through our research we found many different understandings of what counts as housing ‘data’, ‘information’ or ‘records’ in Kyrgyzstan, and who the main players are in collecting, requesting and maintaining such housing data. Respondents at the local level (e.g. municipalities, districts) described a range of bespoke forms of localized data collection related to housing, usually aligning with their specific needs, be they for example new land for construction, access to clean water or better flood or rockslide protection. Three dominant actors were most frequently mentioned at the national level, however: (1) the National Statistical Committee, which collects housing and related statistical data through a network of district offices, publishing some of its datasets and outputs through a modern website; (2) the State Agency for Architecture, Construction, Housing and Communal Services, which monitors and collects data related to urban development and housing construction; and (3) the State Registration Service, which monitors and collects data relating to land boundaries and ownership, as well as other spatial attributes including housing construction, civil engineering and communications infrastructures. Regional and District branches of national departments (e.g. Urban Architecture, State Registration Service) are seen to be an interface for two-way information sharing between local and national state agencies. It should be noted that – as Kyrgyzstan’s civil society sector is regarded as one of the strongest in Central Asia (ICNL, 2019) – it is likely that many third sector organizations also generate housing datasets; this has so far however been outside the scope of our research.

Despite some gaps in our research corpus to date, there is nevertheless evidence of local/central tensions in the availability and compatibility of housing data. There are evidently many forms of housing data generated by local municipalities and private developers, much of which is acknowledged to be either incompatible with national or international data standards, inaccessible and/or proprietary. Meanwhile, national state agencies generate and maintain more standardized housing data. While in principle centralized control of such datasets and infrastructures is accepted by most local actors with whom we spoke, it was also clear that local access to some data is often highly restricted. Accessing land data from the State Registration Service (SRS) was frequently mentioned as an area of difficulty. While the SRS maintains a unified, near-comprehensive land registry information system, local municipal representatives described their access to this data as largely on land parcel-by-parcel basis, for a fee. Other actors, including state departments and private companies, appear by contrast to have negotiated access to larger, aggregated land datasets from the SRS (in some but not all cases for a fee).

One barrier to more open data sharing is the market-based model of many government services and programmes in Kyrgyzstan. The SRS has, for instance, been mandated to be financially self-sufficient, and this has clearly influenced its approach to both restricting access to, and charging for, land data. Moreover, virtually all affordable housing programmes in Kyrgyzstan today entail only indirect state involvement (e.g. providing free land plots, mortgage lending programmes),

with private sector property developers, managers and construction companies more directly involved in the actual provision of housing stock. As a result, respondents we spoke with tended to take-for-granted that a significant amount of housing-related data is inaccessible, since it will be produced and owned by private sector bodies. This even appears to be the case with the State Mortgage Company (SMC), which (in addition to providing loans) aims in the next 5 years to construct 130,000m² of new housing. As part of this process, the SMC has undertaken regional visits across the country, assessing local housing needs in depth; it does not however make available or even necessarily retain this rich information and data, in effect treating it as entrepreneurial market research.

Another strong perception we observed related to the promise of digitization initiatives addressing many if not most of the inefficiencies and inconsistencies in the interactions between state agencies, local municipalities and private actors, including in the area of housing. Kyrgyzstan President Sooronbai Jeenbekov has declared 2019 the “Year of Regional Development and Digitalization” (Levina, 2019), and within this rubric a number of digital technology initiatives are underway. Some of the state and local respondents we spoke with mentioned in particular the Tunduk system, which will create a secure process for certifying interdepartmental electronic interactions, and in principle would eliminate paper-based reporting and document transfers. These and other initiatives (such as improved regional Internet access and data center facilities) seem to hold promise for much improved data-sharing and data infrastructures, though they will not be realised without considerable challenges.

For example, the Norwegian Mapping Agency has been supporting the State Registration Service (SRS) to build and consolidate geospatial data in Kyrgyzstan, towards the development of a National Spatial Data Infrastructure (NSDI). The overarching intention of an NSDI is to provide open-access spatial data to state agencies, local bodies and the private sector. However, even in our small-scale research, we found evidence that there is at present low levels of awareness about, and expertise in, making effective use of geospatial information. It will be a challenge for state and local actors in the housing area to develop or attract that expertise. Although there were clear signs of existing, professionalized information technology expertise, notably within the National Statistical Committee and SRS, state (not to mention local) agencies may struggle to compete with the higher salaries on offer in the private sector, both in Kyrgyzstan and abroad. There were also indications that the SRS might not intend (or have the financial capacity) to make some of its own datasets as open and accessible as the ideals of an NSDI would suggest. Finally, another challenge to surmount will be the diversity of extant datasets, particularly at the local level, which may deploy varying data definitions (e.g. of informal housing, structural conditions, affordability), or exist in different formats (e.g. non-machine-readable data, paper-based information).

Despite these and other challenges, the overall direction of travel in Kyrgyzstan appears positive. With an implementation programme involving all stakeholders, there are strong prospects for

much-improved capacities to measure progress towards national and local housing objectives, including those reflected in SDG 11.

Case 4: Ukraine

Ukraine is a comparatively large country, both in terms of land mass (603,628km²), and a significant, though diminishing population, projected by UN DESA to reach 43.74 million by 2020. Ukraine has been a majority urban country for some time, with 70% of the population living in urban areas, and projected to reach 80% by 2050 (UN DESA, 2018a). Unlike most countries in this region, Ukraine's urban population is concentrated in either small cities (50,000 inhabitants or less) or large cities with more than a million inhabitants. The current housing market is characterised by an over-saturation of newly built apartments, but an absence of affordable mortgage lending and effective state support programs. In recent years, Ukraine has seen significant rural-urban migration, but also internal displacement of hundreds of thousands of people, caused mainly by the ongoing conflict in the East and South of the country and the occupation of parts of those regions.

Over 94% of Ukrainian housing stock is private, as a result of post-Soviet privatisation. However, there is little reliable data on house ownership or rental activity, as the rental market is unregulated and informal, and there is next to no recently updated information on housing conditions or needs. Ongoing housing issues include: dealing with ageing Soviet-era housing stock (mainly apartment blocks) and outdated records; delegating housing management from district administrations to community organizations in a mostly privatised housing market; investing in energy-efficient housing solutions; and regulating the private construction market, which has been reinvigorated since the economic crisis of 2008-2009, but where housing quality and timely delivery have consistently been impeded by corruption and distrust towards developers.

The key actors in the housing data arena that were mentioned during our research include: 1) The State Statistics Service of Ukraine, responsible for collecting data on housing and households through national census, national surveys and other instruments with the help of regional offices, and then making aggregated data available to state bodies and the public; 2) The Ministry for Regional Development, Construction, Public Housing and Utilities, responsible for public housing infrastructure development, and in part for affordable housing provision, as well as for designing and implementing state programmes in the area of housing; 3) The Ministry of Social Policy, also involved in affordable housing provision; 4) state agencies such as The State Fund for Support of Youth Housing Construction, tasked with developing and implementing affordable housing programmes for younger Ukrainians; and 5) The State Inspection for Architecture and Construction, responsible for control, registration, issuing permits, and monitoring of construction projects in Ukraine, including housing. Housing data is also collected by municipal

administrations, and there are multi-way data sharing flows between national-, regional- and municipal-level entities.

There are a number of challenges with regards to collecting, analysing and sharing urban housing data in Ukraine. In regulatory terms, Ukraine finds itself very much in transition from more centralised Soviet and early post-Soviet housing management and data collection practices to significantly decentralised housing data management and collection infrastructure today. The Housing Code currently used was inherited from the Soviet period, and though a new one is under development, many of the provisions are outdated and miss new forms of ownership or new issues. A number of new laws have been adopted in the area of housing management, but bear little connection to the Housing Code. Respondents stress that developing strategies for housing policy, including managing housing data, should come before the redesign of the Housing Code and inform how it is reformed.

The executive bodies and agencies involved in housing data collection have been reorganized several times since Ukrainian independence in 1991. A key network of state agencies responsible for collecting data on housing stock and ownership rights, The Bureaus of Technical Inventory, were abolished in 2012, with their responsibilities outsourced to a chaotic collection of communal, municipal and private agents. This significantly impeded the centralised and standardised collection of housing data, generating internal variation and challenging the data flows between the local and national levels.

Housing affordability emerged as a key issue from our research with regards to urban data collection. On the one hand, respondents indicated that the decentralization processes have made it harder for national agencies to maintain comprehensive datasets on regional and municipal affordable housing needs and practices, as such data is often collected on an ad-hoc basis, and tends to be quite heterogeneous. On the other hand, some respondents stressed that there is a lack of a standardized definition of “affordability” and of parameters which would qualify certain populations as eligible for affordable housing or affordable mortgages. Adding to this, there are emerging categories of individuals such as military veterans or internally displaced persons that also qualify for affordable housing, but data collection about these populations is not standardized. Instead it is managed in an ad-hoc way, both locally and centrally, resulting in a lack of reliable indicators and datasets that would inform the provision of housing rights for citizens.

Dispersal of housing management and data collection to regional and local levels has resulted in better provision for local needs and local contexts, but has created a dearth of reliable, centrally collected and managed data, according to respondents. This impedes the development of evidence-based housing policies and the implementation of international indicators for evaluation of sustainable housing in the country. There is also uneven collaboration among state agencies, with some exchanging data quite successfully, and others severely limiting their data sharing. This is

further complicated by the fact that housing data-related state agencies have various kinds of powers: e.g., the State Statistics Service cannot generate its own policy solutions but can only implement policy generated by relevant Ministries.

As evident from our interviews, technological solutions, such as designing and implementing online registration for waiting lists for affordable housing or digitizing connections between central and regional statistics offices, have made it easier to collect certain kinds of data. Despite this, some types of data require a national effort supported by both technological solutions and political will. Ukraine last held a national census in 2001, and though the State Statistics Service has been designing a new census survey, with a significantly expanded housing data section, the implementation of the census is taking time, with a pilot in Kyiv and Kyiv region set to take place in late 2019, and a national census to follow later.

Technological and data expertise is highly valued by our respondents, but they indicate that it is hard for state institutions to attract employees with requisite skills, as they face stiff competition from the private and business sector. Improving data and statistical literacy for state agencies, media, business and civil society, including in the area of urban data, is seen as crucial to developing an urban and housing data culture in the country, as are innovative data gathering solutions such as sensor-based technologies. Ukraine has also been active in promoting an open data and open government culture, and is a key participant in the Open Government Partnership framework. However, efforts to make housing and other urban data open and available to the public have been uneven and sporadic, with larger cities seeing more success from collaborating with the private sector and non-profit groups to design and implement open data repositories. Moreover, only a segment of open datasets are made available in machine-reading formats, and a significant proportion are only available as PDFs or aggregate reports.

Overall, our respondents in Ukraine see considerable potential for developing an urban data culture that would support citizens' housing rights, enable provision of affordable housing and inform SDG indicators to further help the country adapt to international best practice standards for urban data management and evidence-based sustainable housing policy.

Discussion

Having briefly overviewed the data cultures emerging from the four contexts included in our sample, we now turn to trace some common threads. Three main threads will be discussed here: the transition towards marketization and decentralization; the question of knowledge and expertise; and the evolution of data sharing infrastructures.

Marketization and Decentralization

The four countries in our sample have all been undergoing the transition from planned to market economy. This transition has been coupled with one from a centralized governance model to a decentralized one, providing more autonomy to localities, be they municipalities, districts, cities, regions or provinces. Compared with other contexts undergoing similar decentralization processes (for example the Mediterranean countries of Europe), post-communist countries have experienced far more rapid transitions since 1989. These dramatic changes have left an imprint on the representations and practices of our observed settings.

Across our sample, marketization is never discussed critically in itself. It is either consistently described as a positive and desirable transformation, or is taken for granted as natural and unchangeable. In turn, institutional decentralization is perceived as a necessary consequence of marketization. An important implication, which directly impacts data cultures, is that fragmentation and communication problems are largely assumed as necessary and natural corollaries of institutional decentralization. This naturalization of dysfunctional aspects may contribute to consolidating, to a certain extent, a self-centered attitude by each institution, problematizing forms of cooperation which might be necessary for evidence-based policies towards sustainable housing.

Aside from how they are interpreted by our informants, marketization and decentralization appear to exercise very concrete impacts on multiple dimensions of data cultures: most significantly: (1) data definitions (how different kinds of data are defined and validated); (2) data access (how datasets are accessed and by whom); and (3) data conservation (how datasets are stored and preserved).

Housing *data definitions* refer to the system of formal qualities and labels upon which the specific knowledge domain of housing policies rests. What counts as a ‘dwelling’ for example? What expenses are to be counted as ‘housing expenses’? If concepts are not defined in the same way across the different institutional levels, measurements will tend to be misaligned, with cascading impacts on policy design, implementation and especially monitoring. A typical example of this is ‘informal housing’, which tends to be defined differently across geographic scales. The local level tends to privilege the material quality of the housing (whether or not the housing matches basic habitability criteria), especially in the case of long-term informal housing settlements. The central level tends to give more prominence to juridical status (including building permits, official land deeds, etc.). Furthermore, the content of those forking priorities can be, in some instances, misaligned: for instance, the parameters to be included in definitions of housing quality can vary significantly, especially when datasets about housing materials and techniques (for example proceeding from building licenses) are not readily available. Clearly, data definitions have an important political dimension, and reflect institutional and cultural priorities. In some contexts, we

observed civil society actors participating directly in the definition of indicators. While this involvement further complicates the definition (and propagation) of housing data, it can play an important role in ensuring fair and equitable housing policies. Civil society organizations help in negotiating particular definitions of housing data so that they include the needs of the disadvantaged (an example of this would be the involvement of minorities associations in Albania).

The second dimension impacted by decentralization and marketization is *access to housing data*. Here we focus on the issue of data access across institutional levels. The question of public access to housing data was not widely discussed or practiced in our case study sample (but we return to this issue under Conclusions and Recommendations). Across all contexts, we encountered ‘broken data pipelines’. Data is routinely transmitted from one subject to the other through pipelines, or pathways. When these pipelines are ‘broken’, one subject is not getting the expected or requested data from another subject. For instance, in some contexts localities reportedly fail to receive national income statistics needed for calculating access thresholds to housing aids. In others, central institutions reportedly fail to receive from localities disaggregated information on building quality. Marketization meanwhile confers new opportunities and responsibilities to private sector actors, along with significant rights over the information they generate and produce. For example, detailed data on building construction and materials is often inaccessible to local or central public bodies, and this complicates monitoring the status of housing infrastructure at larger scales be they urban, regional or national. Access to such private housing market data is described by both central and local actors as either inconvenient or problematic, or is simply assumed to be out of reach. This forces public agencies to reconstruct approximations of housing market dynamics from publicly available sources, such as websites and realtors.

Across our case studies, broken pipelines often lead to the development of idiosyncratic methods of data collection or monitoring. For instance, in one of our country studies, data about family income is used to help determine whether or not the family qualifies for public housing. When this data is not provided by central tax agencies (because it is unavailable or not shared), in one municipality, it is instead estimated using on-site inspections, and arbitrary parameters such as the number of television screens in the household. This ‘semi-official’ data is saved in institution-specific databases (typically spreadsheets), which become the basis for other practices and estimations. In general, the longer data pipelines remain broken, the stronger reliance on this institution-specific data tends to become. Data pipelines can become broken in this way because data is ‘siloe’d’ by one institution or department. In other words, access is restricted or made exclusive to one public body, whether formally (e.g. explicitly placing limitations on access) or informally (e.g. by discouraging access through various forms of inconvenience). Siloeing can lead to the same or very similar information being recorded in different and often incompatible ways by multiple agencies, which not only hinders cooperation but is economically inefficient. On the other hand, data siloeing was clearly seen by some of our respondents as desirable, since it ensures

maximum efficiency of a dataset within the organization, with data collection tailored to their specific needs.¹

A third dimension impacted by marketization and decentralization is *housing data conservation*, that is, how housing-related data is stored and preserved over time. All four countries experienced a disruption in the continuity of housing data when transitioning from state-planned to market economies. Our respondents reported significant issues in maintaining standards of data collection and monitoring after the end of state socialist bureaucracies, and in some cases the material loss of entire housing data archives. This relative chaos associated with rapid marketization led to significant inconsistencies and arbitrariness in data collection, and even basic record-keeping, across these countries' changing housing sectors. In turn, this led to episodes of speculation, corruption, proliferation of illegal housing and private seizing of public assets, in many cases originating or intensifying the problems our informants were facing daily. There does however appear to be a growing understanding in these contexts, almost three decades after state socialism, of the impacts rapid marketization has had on effective housing data management practices and infrastructure, and the need to address them.

Fragmented and redundant approaches to data collection, as well as data siloing, are certainly not exclusive to these four contexts. These are global issues affecting all complex data-management endeavors. Yet, in these four contexts, the pervasive assumption of the inevitability of those issues may often be fostering the development of balkanized data cultures, rather than cultures cultivated around data sharing and coordination. The complexity and multidimensionality of housing issues would appear to benefit from such sharing and coordination, which could lead to the creation of 'data lakes' (large shared datasets), ensuring a more accurate and comparable monitoring of housing policies.

Dynamics of Knowledge and Expertise

Data-related knowledge and expertise – i.e. how 'data' is conceptualized, and the knowledge required or desired to make sense of data – was clearly an issue across the four countries in our sample. For the most part, 'data' is equated with quantitative information, with demographic statistics perhaps being the most long standing method, and geospatial coding of data being a more recent development. Qualitative information was seldom discussed as 'data' by our informants, nor were qualitative methods of information gathering. This representation underpins another one we consistently found in our sample: 'data' is framed as a specialized subject for specialized people, rather than as a mundane commodity flowing through the everyday processes of organizations or even citizen's daily activities. As a result, day-to-day responsibilities that may involve data collection or generation are often distinguished from that involving those regarded as data professionals. At the same time, the talent pool for those seen as identifiable data professions

¹ Furthermore, the development of agency-specific methods due to access constraints can, conceptually at least, lead to experimentation and innovation. However, we couldn't observe significant instances of this phenomenon in our sample.

(e.g. statisticians, GIS and database specialists) is described as being drained by the private sector. Some informants, working in agencies they described as overworked and understaffed, had a negative perception of data processing, management and sharing. Not only did they believe they lacked the expertise, but also that the workload entailed in, for example, preparing or cleaning data so it can be shared, added work that was largely unproductive for the agency in meeting its own objectives. Some agencies, especially at the local level, could only support the development of data-related competencies subject to resources, weighed against what they perceive to be their more 'vital' responsibilities.

Up-to-date knowledge around housing stock maintenance and improvements was also noted as an area lacking in expertise, with important impacts on housing data. Ageing Kyrgyz multistory housing blocks in Bishkek and Osk, mostly built during the socialist period, are for example often managed by non-professional building supervisors. After the fall of state socialism, the management of these housing blocks often fell to elderly tenants, who may be effective representatives of the other occupants, but are not professional housing managers. Such non-professional supervisors have often proven unable to identify issues with, or convey consistent information about, the status of the housing infrastructure they manage. In Kyrgyzstan, this has notably affected the attainment of policies seeking to make such multistory apartments more energy efficient. Problems with the professionalism of housing maintenance on the ground, across all four case studies, appears to be forcing state bodies to compensate through their own monitoring efforts (i.e. inspections), with the resulting issues of timing, costs and comprehensiveness. Technologically-advanced solutions, such as automatic sensors transmitting information about building systems, are likely too costly in the contexts we studied, and also structurally difficult given the private ownership of the housing. Professionalizing such building managers will likely have the greatest impact in this area.

One final area noted in some of the case studies was a perceived or real lack of capacity among the general public to make use of openly available data. Media personnel, in particular, were described as lacking proper understanding of hard data and related methods, and accused in some instances distorting the meaning an agency intended to convey through public data sharing. This perceived distortion can result in agencies limiting future data disclosures. In general, across all four contexts, a general lack of confidence in the ability of the public (mediated or otherwise) to interpret data, and in some instances in the quality of the data itself, sometimes complicates the adoption of open data solutions.

Challenges to Building Data Infrastructures

The most common recourse to building better data cultures across our four cases is to build better data infrastructures ("the institutional, physical and digital means for storing, sharing and consuming data across networked technologies" – see Kitchin, 2014: 32). Fragmented, localized

infrastructures still predominate, with different agencies operating siloed, ‘walled gardens’ of databases, and most sharing being done on an ad-hoc basis. However, significant investments are being made towards unified platforms. Centralized, cross-agency data sharing infrastructures are being developed in three of the four countries (Albania, Kyrgyzstan and Georgia), and three countries (Albania, Kyrgyzstan, Ukraine) are replacing their cadasters with digital ones. These leapfrogging projects, mostly driven by central or high-level institutions, and often financed through international grants, denote an awareness of the limitations of the existing solutions, and seem to possess strong formal support.

There are likely to be significant challenges however to ensuring such unified data infrastructures are useful to all stakeholders. At present, the use of inter-agency, persistent databases, accessed remotely, appear limited. In Kyrgyzstan and Georgia, housing and land registration agencies have begun for example to develop unified geospatial data infrastructures, in some cases with the direct aid of foreign bodies such as the Norwegian Mapping Agency. Geospatial data meanwhile is more commonplace in Albania and Ukraine, however it appears to be mostly confined to specialized subjects or bodies (such as national mapping agencies). The ability of actors across agencies and in particular locally to make use of geospatial data seems as yet limited. Some informants also expect that these and other new data infrastructures could encounter resistance with some bodies, due to their idiosyncratic data cultures, as well as a lack of incentives for them to use such emergent data infrastructures. In addition, the varied format of local data could pose problems for integration with larger unified data systems. While spreadsheets appear to be the core instrument of housing policymaking, implementation and monitoring, much local data continues to be stored on paper or in non-machine-readable formats such as PDF.

Prediction-oriented methods – such as machine learning methods – appear to be outside the scope of the examined agencies. Methods involving citizens-as-sensors, delegating to citizens the job of collecting data via the Web or mobile apps (e.g. to report malfunctioning infrastructure) are largely absent in the sample.² Finally, methods based on remote sensing, for example monitoring structural integrity of buildings, also appear rare.

Conclusion and Recommendations

In the introduction we put forward a set of preliminary elements for data cultures which we used as a heuristic tool in our research. In this conclusion, we build on these elements and put forward five more specific dimensions which can be used to investigate and describe data cultures. These

² While the city of Tirana deployed in 2016 a mobile application to record citizen complaints about issues in urban infrastructure, this application does not include housing infrastructure.

dimensions represent the axes upon which organizations or countries can differ. They emerge from our exploratory research, are grounded in our fieldwork and are not meant to be definitive. However, they appear to be sufficiently generalizable outside of our sample, and complexify our preliminary elements of data cultures into a multi-dimensional construct. This construct can be used by policymakers and researchers alike, in the fields of housing and beyond.

Dimensions of Data Cultures

The first dimension we have identified is the *relationships between data and decision-making*: that is, how the quantifiable information embodied in datasets factors into decision-making processes. These processes can be considered both within particular public organizations, as well as more abstractly, across a society at large. Data can be positioned before (*control*) or after (*monitoring*) decision-making, or not really impact it (*irrelevance*). For instance, all four countries share a similar historical trajectory in their housing data culture, in which the three stages correspond to three moments. The state socialist period was marked by the ‘control’ mode, in which systematic data collection was instrumental to long-term planning: most housing stock was traced and accounted for, and was allocated through strict policymaking. The period of transition to market economy appeared dominated by the ‘irrelevance’ mode. The relevance for policymaking of systematic housing data declined: an important role was played by resource scarcity, but also by the fact that housing data collection could not match the speed of the phenomena it was observing (in other words, housing developed in so rapidly that official data could only return very partial image of the situation). Finally, in all four countries, the current period appears dominated by the mode of ‘monitoring’. In this mode, data is produced through the autonomous agency of social actors (citizens, migrants, realtors, developers, other agencies and so on); data track their actions, and it is used to drive those actors towards desirable outcomes (fairer allocation, lesser environmental cost etc.) by means of rewards or penalizations. While striving to achieve a robust monitoring culture, all four countries are also attempting to promote evidence-based decision-making. The challenge appears to be balancing the free agency of social actors with more consistent and comparable data monitoring.

The second dimension of data culture are the *relationships between the data of particular public organizations and civil society*. That is, the degree and forms of attention that civil society agents (chiefly media and third-sector organizations that monitor the activity of state agencies) pay to the information collected, processed and/or made available by public sector organizations. We identified three main statuses for this dimension: indifference, content focus and epistemological focus. Civil society can exemplify indifference when it does not exhibit significant interest in the data of public sector organizations. This can be seen as safer by public organizations, since civil society scrutiny tends to heighten political attention. Yet, it can also foster self-centered data cultures (with the issues outlined in the previous section), and limit the perceived usefulness of

particular public organizations, even putting them at risk during bureaucratic reshuffles. Civil society agents have a content focus when they pay attention to the information or concerns of public sector data – for example the number populations made vulnerable because of housing shortages. Finally, civil society agents can adopt an epistemological focus when they pay attention to how the content of public data is produced. This focus can encourage or even force public organizations to reveal for example their methods, definitions and labels used in data collection and gathering. For example, Roma associations in Albania pay close attention to how public agencies identify and record informal housing, and try to influence it.

A third dimension we can use to describe data culture regards the *systematicity of data collection*. Data collection practices range from highly systematic (data collected at regular intervals) to ad-hoc (data collected only in response to specific events, for example policy changes or disasters). We would add, in the middle range, occasional data collection practices (data collected in a consistent way but at irregular intervals, e.g. when resources are available). In each country, the systematicity of data collection varies significantly across specific organizations and departments. While there appears to be a general consensus about moving towards more systematic data collection, which is generally considered good practice, it should be noted that there can often be valid policy justifications for ad-hoc or occasional forms of data collection.

A fourth dimension of data cultures are *attitudes towards data sharing*. Data sharing in our case examples ranged from openness to siloing. A commitment to more open sharing usually entails an investment of resources in maintaining data pipelines, even when such data are not immediately used. A commitment to open sharing may also involve the promotion of open data ideals, for example the establishment of data portals and data literacy programs designed to encourage the active use of data. And accordingly, organizations committed to data sharing will often also be preoccupied with the risk of usable data going to waste. For example, the Norwegian Mapping Agency, who are working with public agencies in some of our case countries, was actively promoting the use of its urban geospatial data, collected at significant cost, before changes in housing patterns make it obsolete. At the other end of the spectrum, some organizations tend towards siloing their data, either because data is sensitive, because there is a will to capitalize on the value of data (often to recoup the costs incurred in data collection), or because resources are not available to establish data pipelines. The most widespread example of this in our case study research were private real estate actors (developers, realtors, construction companies, etc.), who unsurprisingly consider their data to be proprietary. However our pilot research also found that many public services also siloed or restricted access to data.

A fifth and final dimension regards the *specialization of data-related functions* inside public organizations. Specialization can range from unspecialized to expert-oriented. Unspecialized does not necessarily refer to a lack of data competencies, but a low level of formalization. We observed in many public agencies, particularly at the local level, employees who either developed some

form of data skills on the job, perhaps out of necessity, or inherited such skills from personal interests, education or previous careers. In such organizations, data-related skills may not be a hiring criteria. Public organizations with an expert-orientation to data are more inclined to allocate resources towards staff with verifiable, data-related skills, and consider such skills as a hiring criteria. Moreover, such organizations might involve themselves in broader training and capacity-building efforts, for example with local universities, to train data-related specialists. This appears to be the case, in our sample, of national statistical agencies and mapping agencies.

Recommendations

We now outline some recommendations regarding housing data cultures, aimed at supporting evidence-based policymaking towards the Sustainable Development Goals related to housing (most notably SDG 11.1). Based on our preliminary research, a well-functioning data culture would support:

- a) Inclusive processes of data definition that involve all stakeholders, which can reduce conflict and increase literacy as well as trust.
- b) The collection of comprehensive, recent, high-quality and coherent data across agencies, encompassing the various dimensions of housing, including, but not limited to: housing occupancy, quality and expenditure; household income levels; housing market real prices; energy consumption; and land availability.
- c) Robust, established and replicable data aggregation and interpretation practices being routinely integrated into decision-making processes.
- d) Efficient inter-agency relationships, with shared and well-established data definitions, labels and collection methods, well-maintained data pipelines, and consequently reduced data redundancy (i.e. multiple agencies collecting the same information).
- e) Good outward information sharing mechanisms, notably through open data platforms, empowering civil society (especially media and other third sector monitoring organizations) to hold public organizations to account, and encouraging transparent, inclusive procedures for data definitions, labels and collection methods.
- f) Transparent, reliable data reporting mechanisms towards the international community, for example for the purpose of tracking progress towards SDGs.

These six points outline an ideal type of data culture, but one which we have observed all four countries to be tending toward. In Table 1 we summarize the key measures currently being implemented in our sample to enhance the effectiveness of the national systems. For each measure, we include evaluations of: the impact on public finances ('Economic Cost'); the impact on public accountability and society at large ('Social Impact'); the obstacles it might encounter ('Expected Resistance'); the impacts on 'Data Coherence' (how much the data collected from different

agencies is comparable); and the impacts on ‘Data Comprehensiveness’ (how accurately data describes the phenomenon). None of these solutions are free of negative implications. This is a testament to the complex and multi-dimensional nature of housing and housing data.

	Economic Cost	Social Impact	Expected Resistance	Data Coherence Impact	Data Comprehensiveness Impact
<i>Deploy centralized, cross-agency data sharing platforms</i>	Very costly	Positive	Very high at the agency level (agencies need to radically change MO)	Very positive	Very positive (differences are immediately visible)
<i>Hire data specialists inside housing-related institutions</i>	Very costly, especially at the local level	Neutral	Low at the agency level (Agencies relieved of load)	Neutral	Neutral
<i>Strengthen positive reinforcement for maintenance of data pipelines</i>	Costly	Positive	Low	Modestly positive (more agencies would be sharing data, possibly increasing heterogeneity)	Positive
<i>Strengthen negative reinforcement for the maintenance of data pipelines</i>	Neutral, but risk of indebting localities	Possibly negative (penalties can impact localities' budgets)	Very High	Modestly positive (more agencies would be sharing data, possibly increasing heterogeneity)	Positive
<i>Involve localities and civil society in data definitions</i>	Cost neutral	Very positive (inclusion and empowerment of civil society)	Possibly high at the central level.	Negative (complicates data definition, possibly increasing heterogeneity)	Depends on the local data collection schemes
<i>Centralize data definitions</i>	Low	Neutral	Very high at the agency level (agencies need to change MOs)	Very positive	Depends on the local data collection schemes
<i>Increase open-data disclosure of housing data</i>	Costly	Positive (empowerment of civil society)	Possibly significant at the agency level; none at the social level	Neutral	Positive (disparities are immediately visible)
<i>Train building managers with basic data skills</i>	Costly	Very positive	Possibly modest at the social level, due to change in MOs	Negative (multiple individual reporting practices)	Very positive
<i>Involve citizens as sensors</i>	Costly	Very positive	Very low at the social level, possible at the agency level	Negative (multiplication of data sources)	Positive

Table 1 Measures currently being implemented or discussed in the four countries.

All of the measures outlined in Table 1, taken in themselves, possess recognizable positive impacts on the overall quality of sustainable housing policymaking, implementation and enforcement. However, all of them appear vulnerable to the various forms of resistance or criticism we outlined in this Report. For instance, even when centralized data sharing platforms are deployed, individual agencies might continue to collect data using idiosyncratic methods or formats, possibly periodically importing such data to the centralized sharing platform. When this happens, in the best case scenario the new platform adds a further burden on agencies which already perceive themselves as overworked, reducing their capacity, and likely impacting negatively on the quality

of the data. In the worst-case scenario, this could introduce incoherent or incompatible data to the shared platform (depending on the strictness of its data schemes), undermining the purpose of the platform itself.

Data cultures are a key backdrop for these kinds of failures. If data cultures are not understood holistically, and addressed in a nuanced way, (e.g. at the organizational level), transitions will be much more costly, and possibly less sustainable.

We put forward, therefore, five recommendations for possible interventions related to fostering better data cultures, in the cases studies we examined and potentially beyond:

- 1) **Deconstruct data pessimism.** We found in this research a pessimism related to the perceived inevitability of data fragmentation in decentralized contexts. This can heighten self-orientation and siloing in data management practices. It is true that decentralization and fragmentation tend to be correlated. But dysfunctional data pipelines and massive data heterogeneity are not inevitable in decentralized governance contexts. Fragmentation should be seen as one *possible* outcome of decentralization, which can be counterbalanced with a coordinated strategy and buy-in from all involved stakeholders.
- 2) **Promote data as lifeblood.** We found that in some cases data sharing requirements are perceived as a kind of institutional window-dressing, detached from the real priorities of housing related agencies, such as increasing affordable housing supply. Countering this perception means showing how housing data is a core resource - a 'lifeblood' - in delivering such priorities, both by showing the problems stemming from a lack of data visibility, and the benefits when it is made more visible (e.g. efficiency, fairness, accountability). Promoting data as a lifeblood can expand public agencies' horizons beyond data collection to include better sharing.
- 3) **Consolidate trust in data.** When data definitions (e.g. what is 'affordable' housing, what is an 'informal' settlement) are inconsistent, citizens and civil society can lose trust in data, and in turn, policymakers and public agencies. Data may again be perceived as window-dressing, unrelated to actual policy decisions. Trust in data can be promoted by: instituting open data as a public good, vital resource and even a citizenship right; proving this commitment by establishing open data platforms (not just for large urban centres); demonstrating through public channels how policy decisions are taken on the basis of evidence; and involving civil society in establishing data definitions, while publicizing that involvement.

- 4) **Gradually transition data practices.** Centrally imposing a radical overhaul of an organization's data culture may be a tempting proposition. But drastic transitions can be risky, and encourage the maintenance of existing (or even the development of new) parallel data cultures inside organizations. This can have the effect of duplicating data-related efforts, and further decreasing the efficiency and effectiveness of the organizations affected. Gradual transitions, progressively standardizing single practices, in which stakeholders are consulted and heard, are more likely to succeed and be sustainable.

- 5) **Promote data roadmaps.** The benefits of moving towards evidence-based or data-driven policymaking may not be immediately clear to the officers and employees of public organizations. A gradual transition in data practices is more likely to be understood when there is a convincing 'roadmap' of the changes ahead. A data roadmap will clearly specify steps, milestones, expected challenges and benefits of any transition project, and should always be visible to all affected stakeholders. If the involved parties have a clear vision of where the organization is heading, the impacts or costs incurred make more sense.

In conclusion, we have argued that holistically understanding urban data cultures is a key step in designing and implementing transitions towards more evidence-based policymaking in the housing sector, as well as other urban policy areas. Developing holistic national data cultures avoids the risk of a 'hollow' transition to more superficial harmonized data systems only used for central reporting or international benchmarking. At the same time, holistic national data cultures reduce the risk of localised, idiosyncratic systems proliferating underneath the surface of such harmonized data systems.

This preliminary study only investigated a small fraction of complex, national data cultures related to housing in Albania, Georgia, Kyrgyzstan and Ukraine. More research is needed to account for data-related practices in the housing field that were not investigated in this study, such as closer observations of daily routines inside specific public organizations, the practices of the private real estate sector, and the practices of different third sector organizations involved in housing provision and advocacy. Current research on these topics has largely focused on developed countries. A better understanding of the urban data cultures in developing or transitioning countries, and specifically those examined in this Report, is all the more urgent in building capacities towards evidence-based policy towards the SDGs.

References

- Bowker GC (2000) Biodiversity datadiversity. *Social Studies of Science* 30(5): 643–683.
- boyd d and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- D’Ignazio C and Bhargava R (2018) Approaches to Building Big Data Literacy. In: *Bloomberg Data for Good Exchange Conference*, New York, USA, 28 September 2018.
- Edmunds A and Morris A (2000) The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management* 20(1): 17–28.
- Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4): 211–221.
- Graham M, Zook M and Boulton A (2013) Augmented reality in urban places: contested content and the duplicity of code. *Transactions of the Institute of British Geographers* 38(3): 464–479.
- ICNL (2019) ‘Civic Freedom Monitor: Kyrgyz Republic’ The International Center for Not-for-Profit Law. Available at: <http://www.icnl.org/research/monitor/kyrgyz.html> [last accessed 20 September 2019].
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.
- Levina M (2019) Kyrgyzstan aims to be in the forefront of digitalization in Central Asia *The Times of Central Asia*. 27 April 2019. Available at: <https://www.timesca.com/index.php/news/26-opinion-head/21103-kyrgyzstan-aims-to-be-in-the-forefront-of-digitalization-in-central-asia> [last accessed 20 September 2019].
- Loukissas YA (2019) *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, Massachusetts: The MIT Press.
- Olivier de Sardan J-P (2017) Practical norms: Informal regulations within public bureaucracies (in Africa and beyond). In: de Herdt T and Olivier de Sardan J-P (eds) *Real Governance and Practical Norms in Sub-Saharan Africa*. London: Routledge, pp. 19–62.
- Ratner H and Ruppert E (2019) Producing and projecting data: Aesthetic practices of government data portals. *Big Data & Society* 6(2). DOI: 10.1177/2053951719853316.
- Ratti C, Frenchman D, Pulselli RM, et al. (2006) Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design* 33(5): 727–748.
- SDSN (2015) *Indicators and a Monitoring Framework for the Sustainable Development Goals: La5*.
- Smith H (2019) The locative imaginary: Classification, context and relevance in location analytics. *The Sociological Review*. DOI: 10.1177/0038026119878939.
- UN DESA (2018) ‘World Urbanization Prospects 2018: Kyrgyzstan Country Profile’ United Nations Department of Economic and Social Affairs, Population Division. Available at: <https://population.un.org/wup/Country-Profiles/> [last accessed 20 September 2019].

- UN DESA (2018a) 'World Urbanization Prospects 2018: Ukraine Country Profile' United Nations Department of Economic and Social Affairs, Population Division. Available at: <https://population.un.org/wup/Country-Profiles/> [last accessed 20 September 2019].
- UN DESA (2018b) 'World Urbanization Prospects 2018: Georgia Country Profile' United Nations Department of Economic and Social Affairs, Population Division. Available at: <https://population.un.org/wup/Country-Profiles/> [last accessed 20 September 2019].
- UN-Habitat (2018) *Sustainable Development Goals and Urban Local Bodies: The Future We Want*. Guidance Note Prepared by UN-HABITAT in India with Support of UN Agencies in India, May 2018.
- Wilken R (2019) *Cultural Economies of Locative Media*. New York, NY: Oxford University Press.
- Wilson MW (2012) Location-based services, conspicuous mobility, and the location-aware future. *Geoforum* 43(6): 1266–1275. *unching a Data Revolution for the SDGs*. A Report to the Secretary General of the United Nations by the Leadership Council of the Sustainable Development Solutions Network, 12 June 201