Smart Rural: The Open Data Gap

Johanna Walker[1]*, Gefion Thuermer[2], Elena Simperl[2] and Les Carr[1]
[1]University of Southampton, [2]King's College London
* Johanna.Walker@soton.ac.uk

Abstract

The smart city paradigm has underpinned a great deal of the use and production of open data for the benefit of policy makers and citizens. This paper posits that this further enhances the existing urban rural divide. It investigates the availability and use of rural open data along two parameters: pertaining to rural populations, and to key parts of the rural economy (agriculture, fisheries and forestry). It explores the relationship between key statistics of national / rural economies and rural open data; and the use and users of rural open data where it is available. It finds that although countries with more rural populations are not necessarily earlier in their Open Data Maturity journey, there is still a lack of institutionalisation of open data in rural areas; that there is an apparent gap between the importance of agriculture to a country's GDP and the amount of agricultural data published openly; and lastly, that the smart city paradigm cannot simply be transferred to the rural setting. It suggests instead the adoption of the emerging 'smart region' paradigm as that most likely to support the specific data needs of rural areas.

Keywords **–** Open data; agriculture; European Data Portal; rural open data;

1 Introduction

The urban-rural divide is one of the most well-known foci of digital inequality (Vicente Cuervo & López Menéndez, 2006). In comparison to their urban or suburban counterparts, rural populations tend to experience gaps in education, income, device availability, and mobile and internet access. The Eurostat Rural Development Statistics[1] illuminate points of difference in underlying challenges in rural and urban areas across issues such as health, housing and education, which mean that the digital demands in each of these domains in rural areas differ from those experienced in the urban setting.

Twenty seven percent of EU citizens live in rural areas[2]. The more densely populated EU member states have a very high proportion of urban population, but the opposite is true for many countries, especially in Eastern Europe. Over half of Lithuanians and nearly half of Slovenians, Hungarians and Croatians live in rural localities. However, even in some Western member states, such as Ireland and France over a third of the population are classified as living rurally. Although in Western and Northern member states rural living offers comparatively better housing and relatively available employment, in Eastern, Southern and Baltic member states, rural citizens tend to be among the poorest. Being on the wrong side of the digital divide therefore is exacerbated by rural poverty, and, we argue, potentially being excluded from the beneficial possibilities of open data.

Cities and urban areas across both the Global North and South have become avid users and sources of data in a variety of ways, but generally under the umbrella term of 'smart cities'. Conversely, the Rural Open Data Project[3] states that, "few if any rural local governments provide open data, and little is known about how open data affects rural communities. If there is a benefit to communities from open data policies, it is likely that rural communities are benefiting less than urban ones, if at all." Yet data is perceived to be so vital for cities and citizens that the Urban Data Platform[4] project has the aim of speeding up the adoption of common open urban data platforms and ensuring that "300 million European citizens are served by cities with competent urban data platforms, by 2025". Open data is at the very heart of the smart cities concept (Ojo et al., 2015).

In this short paper we attempt to define what rural data might consist of and identify some dimensions of its availability and use. We also engage with the smart city paradigm to explore its use for invigorating 'smart rural' open data.

2 Defining Rural

How to define 'rural' is virtually a research area of its own. As Wineman et al. (n.d. forthcoming) show, there are a variety of ways to define rural, and the different definitions affect certain indicators. Our approach is based on the quotidien dictionary definition of 'pertaining to the countryside, those living in the countryside, or agriculture' (Merriam-Webster, 2019). This therefore led us to explore both population (those living in the countryside as defined

[1] https://ec.europa.eu/eurostat/web/rural-development/data/database
[2] https://ec.europa.eu/eurostat/statistics-explained/images/4/49/Focus_on_rural_areas_RYB2017.xlsx
[3] https://www.ruralopendata.ca
[4] https://eu-smartcities.eu/initiatives/68/description

by the EU DEGURBA population statistics[5]), and what we term 'rural economy' – specifically those economic activities defined by EuroStat Rural Development as 'rurally thematic industries', primarily agriculture, forestry and fisheries.

2.1 Rural Population Challenges

As noted above, there are key differences in the underlying challenges for rural populations. This means simply transferring solutions created with or on data from urban areas is unlikely to be of use. Whereas housing stock and availability is a major challenge for cities, 80% of the rural EU population lives in a house rather than apartment blocks. Health challenges also differ. Rural populations are often deterred from seeking medical treatments due to the distances involved, unlike urban populations, where the focus is on speeding up the diagnostic process to reduce waiting times. Across the EU, the level of those not in education, employment or training is higher in rural areas. In addition, the gender gap is wider in the rural labour market. Further, solutions developed for more densely populated areas – smart city solutions – may be technically inappropriate in a rural context.

2.2 Rural Economy Challenges

Issues around food supply and environmental concerns such as CO2 levels and water run-off are common to both urban and rural settings. However, significantly different challenges include accurate and verifiable forestry mapping and monitoring, livestock monitoring for animal welfare standards, prevention of disease, and supply chain monitoring. More transparent supply chains are needed to reduce the likelihood of events such as that which arose in 2013, when several European countries discovered that processed meat sold in a variety of locations as beef, in fact included traces of horsemeat.

Data is required not only for decisions at the producer level (for instance, geomatic and meteorological data is useful for agriculture practices), but also for state level production policy decisions and planning for farming and use of land.

3 Methodology

To analyse the relationship between open data and rural economies and populations, we pursue a mixed methods approach, combining qualitative interviews with statistical analysis.

For our qualitative analysis, we conducted interviews with open data experts from EU member states with high levels of rural population, and experts from the rural data economy. We asked them about their view on the current status of rural open data, the challenges and potential use they saw, and their own and observed use of rural open data. The interviews were then thematically coded, both deductively, using the categories from the interviews questions as a starting point, and inductively, with themes emerging from the conversations and observations of the interviewees themselves. Existing research was used to develop generalisations from the interviews.

For our statistical analysis, we collated data from several sources:

- The European Data Portal and its meta-statistics (August 2019 and July 2020)
- The EDP Open Data Maturity Reports 2018 and 2019 (European Commission, 2019, 2020)
- Eurostat data sets on GDP, rural areas, and agriculture

We tested several combinations of indicators for correlations, to compare the rurality of countries with their open data practices pertaining to agriculture, focusing on the Open Data Maturity, population in rural and urban areas, and number and proportion of published data sets.

4 Results

4.1 Rural population data - availability

These challenges of collation and publication can be classified as those of digitalisation, awareness, IT capacity and capabilities, usefulness, and standardisation. Digitalisation, even in highly urban states, can often take the form of publishing PDF documents. In more rural areas of countries such as Romania, administrations still use analogue tools, although most budgets and expenses should be digitalised. Awareness of open data in rural areas may be non-existent, even amongst public servants and decision makers. To achieve improved awareness, local champions are key. The lack of IT capacity in many rural administrations is a key challenge, not least because responsibility for open data initiatives, even in urban governments, often lies with the IT department. Ireland, a highly digital nation, has no IT capacity in many rural areas. The Roscommon County Council Open Data site[6] is a well-known example of open data being published by a rural (and very small) council. However, this was largely driven by the enthusiasm and skill of one employee, who has since left. Lastly, large, decentralised countries like Romania struggle with standardisation. Almost 3,000 local authorities submit data for potential publication, but the format and content of this data is hugely varied, and requires substantial work to be made cohesive.

We hypothesised that, as our interviews suggested that rural data could be more challenging to collect and use, the EU

[5] https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/degurba

[6] http://www.roscommoncoco.ie/en/

Member States with more rural populations are in earlier stages of Open Data Maturity (EDP, 2018).

Table 1: Correlation between proportion of rural and urban, and total population size (2017) and Open Data Maturity in 2018 and 2019

	ODM 2018		ODM 2019	
	effect	p	effect	p
rural	0.017	0.000	0.016	0.000
urban	0.013	0.000	0.014	0.000
total	0.015	0.000	0.015	0.000

When we measure the rurality of an EU country by where their population resides, there is a very weak correlation with its Open Data Maturity (ODM) score, as shown in Table 1. A higher percentage of both rural and urban population correlated with a higher ODM. As the effect is tiny and occurs for both population groups, this does not provide further insight.

4.2 Rural population data - use

Our interviewees agreed that the greatest potential of rural open data was to impact rural areas rather than have rural users, because the skills, knowledge and connections to ideas and innovation were most often found in urban areas with exposure to the quadruple helix of business, government, academia and citizenry. This mirrors other research on 'smart and competitive rural areas', such as that done by the European Network for Rural Development (2016).

4.3 Rural economy data - availability

In terms of data availability for the rural economy we looked once again at the correlation between a countries' Open Data Maturity (2018) and this time, the relevance of agriculture for their economy, approximated through the proportion of agriculture to their GDP. We find a significant correlation (p=0.000, st. effect=0.871 – see Figure 1), indicating that higher ODM is linked to a higher relevance of agriculture. Assuming that the less mature countries would have more agriculture is incorrect – indeed, the opposite is the case. We therefore reject the idea that rural open data challenges are simply caused by a lack of maturity.

Next, we looked at the correlation between a countries' Open Data Maturity (2019) and the number of datasets they publish in general, as summarized in Table 2. We found a significant, albeit small, correlation. This is not surprising as datasets published is one of the measures on which the ODM score is based. This correlation disappears, however, when we look at the number of agricultural datasets published, which does not have a statistically significant correlation with the ODM. The maturity is there, but it is not leveraged for rural economy open data.

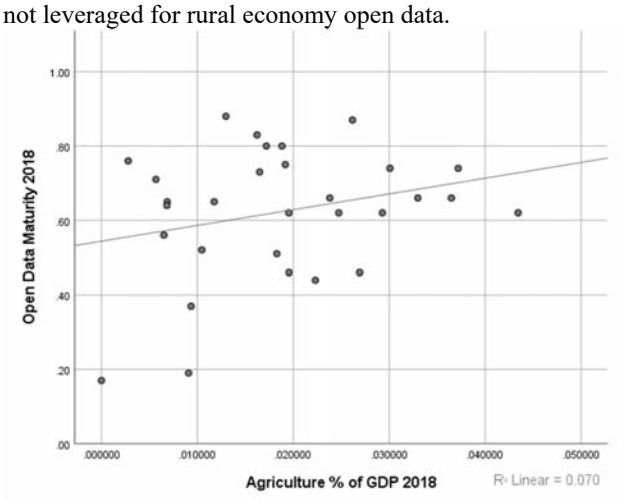

Figure 1: Open Data Maturity 2018 by Agriculture % of GDP 2018, with fitted line

Thus, a third key finding from our statistical analysis is the lack of relationship between the importance of agriculture to a country's GDP, and the number of open data sets published in the category of 'agriculture' on the EDP (which includes fisheries, forestry and food). Agriculture covers 47% of the EU territory and represents around 40% of the EU budget; forestry accounts for another 1% of total EU GDP (Nègre, 2020), and is also important culturally. Fishery is a key investment area for developing sustainability and growth across the EU.

Table 2: Correlation between total and agriculture datasets and Open Data Maturity (2019)

	effect	p
datasets 2019	0.000004	0.021
agriculture data 2019	-	0.075
datasets 2020	0.000004	0.014
agriculture data 2020	-	0.074

Agriculture is the largest category on the EDP, but highly concentrated: Germany, Poland, and France provide over 94% of agriculture datasets. If we use the % of agriculture of GDP in 2018 as a proxy for the importance of the rural economy sector, and correlate this with the total number or proportion of agricultural datasets they published, as shown in Table 3, the results remain negative: There is no significant correlation between a countries' economic reliance on agriculture and the total number of agricultural datasets, or the proportion of agricultural of their overall published datasets, with neither the dataset count from 2019 nor 2020.

Table 3: Correlations (p-values) between total and proportional agriculture datasets and countries % of GDP derived from agriculture (2018)

	2019	2020
total	0.308	0.253
proportion	0.166	0.293

The key finding here is the inconsistency between the relative value of agricultural data to these countries, and the availability of datasets; and also the inconsistency of data publication across the member states in general.

Although there is no discernible pattern in the relationship between proportion of GDP and published agricultural data, this raises crucial questions. Why are so comparatively few member states publishing agricultural data? Are Germany, France and Poland simply opening agricultural datasets because they are easily accessible, or should other member states be opening agricultural data at the same levels?

Some of the reports from our respondents shed some light on these issues. They informed us about agricultural data that was only available to citizens or organisations on a request basis, and often shared only as a web page or PDF, and datasets published on the Ministry of Agriculture site, but not on the relevant national portal, all of which would prevent inclusion in the EDP catalogue.

4.4 Rural economy data - use

We interviewed several open data portal representatives from countries with large rural populations about the level of agricultural datasets requested. These were rarely amongst the most requested datasets, which in Slovenia, for example, concern vehicle licensing, public sector salaries, and building permits, with agricultural statistics on crops and the number of livestock further down the list. There are a number of plausible reasons for this, one of which might be fewer exemplar uses available. We found relatively few use cases (28 out of 548) concerning agriculture-adjacent topics on the European Data Portal.

As with population data, interviewees noted that in many cases these use cases were likely to be driven by urban-based initiatives or agribusiness, creating a gap between the potential audience and the potential creators. This gave rise to the fisherman in the boat image: while open data about fisheries might be of much value to individuals in the fishing industry, the majority are unlikely to have the skills or resource to realise this value.

5 Discussion

Much of what we found regarding the lack of institutionalisation, in infrastructure, skills and champions, reflected earlier challenges of open data in urban settings. So it is unsurprising that there have been several attempts to import the tools and techniques of smart cities rurally. Two such projects have happened in Scotland, based on the idea that, "Rural [is] just a low density city" (The Stonehaven Rural Co-op[7]). Fintry[8] is a village with an open data portal and interactive energy dashboard, hailed as the United Kingdom's first "smart" village. However, as noted above, such transfers may fail to take into account the specific needs and requirements of rural areas.

A solution may be the emergence of 'smart regions'. Smart regions have extended the smart city to reflect the fact that urban areas are not completely independent of the rural areas that surround them. This creates the necessary links between urban and rural, while acknowledging the differences. Examples of these include Cork Smart Gateway[9]; Smart Atlantic Way (Brolcháin et al., 2018) and Helsinki Smart Region[10]. Understanding the key factors of smart regions allows for the identification of the key data that should be opened. The EU's 2018 Action for Smart Villages notes that, "Smart means thinking beyond the village itself. Some initiatives are taking place at village level, but many involve the surrounding countryside, groups of villages, small towns and links to cities" (Paneva, 2017). This kind of thinking can be used to help develop the smart region concept.

These links with urban areas are also established through universities with complementary interests. There are already examples of this in the agricultural sector, for instance, the Wageningen University Masterclass Accelerator. This approach fosters new agricultural business models based on rural-urban linkages. It is managed by a partnership of two organisations: Wageningen Economic Research (WecR, Wageningen University) and the Municipality of Rotterdam's 'Food Cluster' (*EU Rural Review*, 2017, p. 31). Foodvalley, the leading agro-food centre in Europe, brings together eight municipalities to create knowledge and innovation in healthy food, and also uses Wageningen's data. Such approaches could also help to strengthen the intentional selection of key agricultural, forestry and fishery datasets for publication by countries whose rural economy is important, to assist and empower local actors.

Identifying local actors is key. As our respondent said, "There are few fishermen sitting in a boat thinking of a way to use open data." However, farmers are rurally-located consumers and creators of a great deal of data. In many cases, the value of the data is created by combining with other farms, or other data types. While there may be some commercial sensitivity issues, there are also strong arguments for publishing openly, particularly around identifying and combating the spread of disease.

[7] www.smartrural.coop
[8] www.smartfintry.org.uk
[9] http://www.corksmartgateway.ie/
[10] helsinkismart.fi

6 Conclusion

During the execution of the original and further research for this paper, we examined many reports and policies on open data, and on rural populations and industries. We were struck by the limited overlap between the two areas. It is crucial that rural populations and industrial sectors are not (unintentionally) excluded from the benefits of the data economy. If the focus of data collection and its subsequent re-use in technological artefacts in urban areas is prioritised, those who are not part of the data become invisible, and are subsequently ignored.

Acknowledgements

The initial research for this paper was funded by the European Data Portal as part of their Analytical Reports series. Further research was supported by Data Market Services (H2020 #825014). We thank Gianfranco Cecconi, Esther Huyer, the EU Office of Data Publication, and our interviewees, for their generous insights.

References

Brolcháin, N. Ó., Ojo, A., Porwol, L., Minton, D., & Barry, C. (2018). Examining the feasibility of a Smart Region approach in the North West Atlantic and Borders Region of Ireland. *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, 568–574. https://doi.org/10.1145/3209415.3209512

EDP. (2018). *Open Data in Europe 2018 | European Data Portal*. https://www.europeandataportal.eu/en/dashboard/2018

ENRD. (2016). *Rural Businesses*. European Network for Rural Development. https://enrd.ec.europa.eu/enrd-thematic-work/smart-and-competitive-rural-areas/rural-businesses_en

European Commission. (2019). *Open data maturity report 2018*. https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

European Commission. (2020). *Open data maturity report 2019*. https://op.europa.eu/publication/manifestation_identifier/PUB_OABE19001ENN

Nègre, F. (2020). *Fact Sheets on the European Union: The European Union and forests*. European Parliament. https://www.europarl.europa.eu/factsheets/en/sheet/105/the-european-union-and-forests

Ojo, A., Curry, E., & Sanaz-Ahmadi, F. (2015). *A Tale of Open Data Innovations in Five Smart Cities*. *2015*, 2326–2335. https://doi.org/10.1109/HICSS.2015.280

Paneva, V. (2017). *EU action for Smart Villages* [Text]. https://enrd.ec.europa.eu/news-events/news/eu-action-smart-villages_en

Skakelja, N., & Thorpe, E. (Eds.). (2017). *EU Rural Review* (No. 24). European Network for Rural Development. https://enrd.ec.europa.eu/sites/enrd/files/publi-enrd-rr-24-2017-en.pdf

Vicente Cuervo, M. R., & López Menéndez, A. J. (2006). A multivariate framework for the analysis of the digital divide: Evidence for the European Union-15. *Information & Management*, *43*(6), 756–766. https://doi.org/10.1016/j.im.2006.05.001

Wineman, A., Alia, D., & Anderson, C. (n.d.). *Definitions of 'Rural' and 'Urban' and Understandings of Transformation in Sub-Saharan Africa*.