Analytical Report n14

Analytical Report 14: Enabling Smart Rural: The Open Data Gap

The European Data Portal Consortium prepared this study as part of the European Data Portal project. The European Data Portal is an initiative of the European Commission, implemented with the support of a consortium led by Capgemini Invent, including Intrasoft International, Fraunhofer Fokus, con.terra, Sogeti, 52North, Time.Lex, the Lisbon Council, and the University of Southampton. The Publications Office of the European Union is responsible for contract management of the European Data Portal.

For more information about this paper, please contact:

European Commission
Directorate General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Daniele Rizzi – Policy Officer
Email: daniele.rizzi@ec.europa.eu

European Data Portal
Gianfranco Cecconi, European Data Portal Lead
Email: gianfranco.cecconi@capgemini.com

Written and reviewed by:
Johanna Walker
Email: j.c.Walker@soton.ac.uk
Gefion Thuermer
Email: g.thuermer@soton.ac.uk
Elena Simperl
Email: e.simperl@soton.ac.uk
Esther Huyer
Email: esther.huyer@capgemini.com

Last update: 15.07.2020
www: https://europeandataportal.eu/
@: info@europeandataportal.eu

DISCLAIMER

Abstract

This analytical report examines collection, availability and use of open data in rural areas. It does this through the lens of the 'smart city' and asks what the prioritisation of the development of services and products for urban populations and industries means for rural populations and rural industries and their specific requirements. It finds that these are not currently being met, and are insufficiently examined or supported by open data, with the exception of agricultural, fisheries and forestry data in more urban Member States, which is thriving. Further, it suggests that 'smart city' approaches should not be simply transposed onto the rural environment, but that there is value in 'smart region' approaches that recognise the relationship between cities and their surrounding countryside. It recommends the following activities directed at Member States with larger rural populations: the institutionalisation of open data skills and knowledge; increasing awareness of High Value Datasets; engaging with complementary data owners such as non-governmental organisations and researchers; developing skills links with urban areas and the creation of hackathons or challenges that specifically target rural issues.

Content

1. Introduction

Cities and urban areas have become avid users and sources of data, largely under the catch-all and loosely defined term 'smart cities'. Open data is considered a "defining element" of smart cities[1] . So vital is data for cities and citizens that the Urban Data Platform project[2] 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". This focus on cities makes great sense: they have dense populations which can be reached comparatively easily; infrastructural changes in smaller geographical areas can have substantial impact; and in many countries, such as Spain, Cyprus, Malta and the United Kingdom, reaching the urban population effectively means reaching 50% of the entire population.[3]

But what do smart cities imply for the 27% of EU citizens who live in rural (non-urban) areas?[4] While many densely populated EU Member States have a very high proportion of citizens living in urban areas, the opposite is true especially in most Central and Eastern European countries. Over half of Lithuanians, nearly half of Slovenians, Hungarians and Croatians live in rural areas. But even more Western states, such as Ireland and France, over a third of citizens live in rural areas.

One approach to tackling the domination of the urban paradigm is to directly import the tools and techniques of smart cities and big data into the rural environment. The Stonehaven Rural Co-op project in Scotland claims that, "Rural - it's just a low-density city."[5] This implies that smart processes, technologies and activities that succeed in cities can be successfully implemented in the rural setting. Another Scottish initiative, Smart Fintry, is a recently completed project that aimed to do just that - bringing urban innovation to a village in Stirlingshire.[6] However, there are challenges in merely transposing urban practices to a rural setting. Some of the key areas of smart city investment, such as those to do with traffic and parking management, and multi-modal transport, are far less applicable in rural settings. In comparison to their urban or suburban counterparts, rural populations tend to experience (negative) gaps in education, income, device availability, and internet access.[7] Further, according to the Rural Open Data Project, "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."[8]

---

[1] Ojo, A, Curry, E and Zeleti, F (2015) A Tale of Open Data Innovation in Five Smart Cities
https://www.researchgate.net/publication/280625326_A_Tale_of_Open_Data_Innovations_in_Five_Smart_Cities
[2] https://eu-smartcities.eu/initiatives/68/description
[3] https://ec.europa.eu/eurostat/statistics-explained/images/4/49/Focus_on_rural_areas_RYB2017.xlsx
[4] https://ec.europa.eu/eurostat/statistics-explained/images/4/49/Focus_on_rural_areas_RYB2017.xlsx
[5] https://www.smartrural.coop
[6] http://smartfintry.org.uk
[7] https://arxiv.org/abs/1901.01920
[8] https://www.ruralopendata.ca

This analytical report will focus specifically on the availability and role of rural open data in Europe, leveraging insights from existing research and use cases. By seeking to understand and define 'rural' open data, the quality, quantity, uses and barriers to use can be examined.

2. Rural Open Data – What Is It?

While 'open data' is a concept that has been well-defined[9], 'rural' is not such an easy concept to pin down. The Merriam-Webster dictionary defines 'rural' deceptively simply, as 'pertaining to the countryside, those living in the countryside, or agriculture'[10]. However, 'rural' is a particularly challenging concept to define in terms of developing a working research terminology, with an extensive research corpus dedicated to its consensus definition.[11] The US government alone has 15 separate definitions used by different departments for various purposes.[12] Further, the working definition is often recursive, being understood as 'pertaining to areas that are not urban', [13] and there may be 'shades' of rural, such as 'mainly rural', 'largely rural' and 'urban with significant rural'.[14]

In the EU, a new 'degree of urbanisation (DEGURBA) classification was created in 2014[15]. This classification distinguishes three types of areas: densely, intermediate and thinly populated areas. This new approach was developed to harmonise several similar but not identical spatial concepts and is based on the population grid. However, population is only one way to define rural - others, as the case study below shows, include levels of built-on surfaces and intensity of light at night.

Case Study: Defining Rural[16]

Researchers at the University of Washington set out to explore what would happen if they used different datasets to define areas as rural or urban in their target countries. They did this to understand what impact the definition of 'rural' might have on policy.

---

[9] www.opendefinition.org

[10] https://www.merriam-webster.com/dictionary/rural

[11] https://search.proquest.com/openview/2f7779f7bf760ae235084e1bd167c618/1?pq-origsite=gscholar&cbl=1821520

[12] https://www.washingtonpost.com/politics/the-federal-definition-of-rural--times-15/2013/06/08/a39e46a8-cd4a-11e2-ac03-178510c9cc0a_story.html

[13] https://www.hrsa.gov/rural-health/about-us/definition/index.html

[14]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/597751/Defining_rural_areas__Mar_2017_.pdf

15 https://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf

16 Definitions of "Rural" and "Urban" and Understandings of Transformation in Sub-Saharan Africa
A. Wineman1*, D. Alia 1, C. L. Anderson

As well as looking at the administrative definition (or official country statistics) they considered the following definitions, using publicly available datasets:

Definition	Data Set	Data Description
Population (rural = less dense)	WorldPop (University of Southampton)	Open high resolution geospatial demographic data on population characteristics in low- and middle-income countries.
Impervious surface (rural = little to no impervious surface)	Global Manmade Impervious Surface (NASA)	Open high spatial resolution estimates of global man-made imperviousness for the target year 2010, derived from global 30m Landsat satellite data
Nightlight intensity (rural = low nightlight intensity)	DMSP-OLS Nighttime Lights Time Series (National Ocean and Atmospheric Administration)	Longitudinal satellite record of intensity of light at night (therefore manmade). Defense Meteorological Satellite Program datasets are publically available and downloadable from the NOAA site.
Administration + population + density of building (rural = fewer buildings, lower official population))	Africapolis (OECD)	Open comprehensive and homogenous dataset on urbanisation dynamics in West Africa since 1950. Its original methodology combines demographic sources, satellite and aerial imagery, to provide population estimates and geolocation at the level of individual agglomerations
Subjective assessment (rural = areas the researchers agree look rural, based on a subjective understanding)	Google Earth Images (Google)	Although Google Maps is not 'open data' per se, this method of looking at them and then making a decision is available to all.

The researchers discovered that the different definitions of rural and urban considerably affected certain indicators, such as the rural poverty rate or rural electrification rate. For both the countries they studied, definitions not used by the government decreased the level of urbanisation. In conclusion, using an appropriate definition of rural had the ability to greatly impact policy choices.

However, it is necessary to develop a working understanding of the term in order to direct our research. A common method is to look at population densities, and to so define the 'degree of urbanisation'.

Eurostat[17] uses this method, which directs our research towards investigating open data pertaining to, in Merriam-Webster's terms, 'those living in the countryside.'

However, as Merriam-Webster indicate, 'rural' is also about certain industries. Eurostat Rural Development associates fisheries, forestry and agriculture as rurally thematic industries. This directs our research towards the availability and use of data that impacts rural industries: primarily agriculture, forestry and fisheries.

Thus, this report focuses on two types of rural open data - those about people living in rural areas ('rural populations') and those about traditionally rural industries ('rural sectors').

2.1 Rural population data

By rural population data, we mean data about what it is like to live and work in rural areas. This might include data on housing stock, condition and affordability; broadband and mobile connectivity, speed and cost; levels of material deprivation; access to transport (owned, shared and public); demographic profile (while the proportion of older people living in rural areas is higher than in urban areas, does it become more difficult for people to reside rurally when they are no longer able to drive / need greater access to health and care services?); types of jobs, sectors, pay, employment levels and access to public services (doctors, hospitals, social care, ambulance, the different levels of education, postal services, welfare services and so on).

2.2 Challenges faced by rural populations that open data can help address

Solutions developed for more densely populated areas – smart city solutions – may be technically inappropriate in a rural context. The Eurostat Rural Development Statistics[18] also shed some light on how different the underlying challenges in rural and urban areas are. Whereas housing stock and availability is a major challenge for cities, 80% of the rural EU population lives in a house (as opposed to apartments/flats). The 'overburdening' of housing costs is also less prevalent in rural areas. However, in some south-eastern member states, especially those with few non-subsistence farming activities, housing cost overburdening is greater. This reflects a wider general truth, which is that in western and northern member states, poorer citizens tend to be concentrated in urban areas, whereas the reverse is true for eastern, southern and Baltic member states. Solutions developed for poverty alleviation must be aware of and address this difference.

---

[17] https://ec.europa.eu/eurostat/web/rural-development/methodology
[18] https://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_rural_areas_in_the_EU

Health challenges also differ for rural versus urban populations. Urban populations frequently suffer from extensive waiting times, resulting in a focus on speeding up diagnostic processes. Rural populations are often deterred from seeking treatments due to the distances and resulting necessity to travel involved, which may require altogether different solutions. Overall, rural populations, especially those in 2004 accession states, have more unmet healthcare needs than urban populations.

Across the EU, the level of those not in education, employment or training (NEETS) is higher in rural areas. In addition, In the rural labour market the gender gap is wider.

2.3 Rural economy data

This data is pertaining to the (predominantly) rural industries of agriculture, fisheries and forestry. Agriculture covers 47 % of the EU territory and represents around 40% of the EU budget. Forestry accounts for 1% of total EU GDP, and is also important culturally.[19] Lastly, fishery is one of the key investment areas for developing sustainability and growth across the EU, with €6400M allocated to this from 2014-2020. Such data might be about all aspects of growing conditions, including soil, irrigation, weather, climate and other environmental issues, as well as all aspects of the plants or animals that are grown, including genetics, husbandry and best fit. Food provenance and supply chains are closely related. As well as supporting those who work in these industries, regulators of these industries also require this data. At an EU level, the Agricultural Trade Statistics, the Agricultural and Rural Development Budget and the Farm Accountancy Data Network data have all been identified as Datasets of high value[20][21], and recent regulations have changed the way that agriculture is reported and accounted for, through the new integrated farm statistics and statistics on agricultural input and output.[22]

2.4 Challenges that can be addressed by rural economy data

Some issues are naturally similar to those found in more urban settings. Agrifood businesses, use data extensively, and monitoring of certain environmental issues such as water run-off, C02 levels and soil erosion, reflect urban concerns[23]. However, there are specific challenges related to rural economy data that differ significantly in focus from urban challenges, for example, accurate and verifiable forestry

---

[19] http://www.europarl.europa.eu/factsheets/en/sheet/105/the-european-union-and-forests
[20] This is based on the most recent published report, which is dated 2014, however, there is ongoing work to establish more up to date High Value Datasets.
[21] https://ec.europa.eu/isa2/sites/isa/files/publications/report-on-high-value-datasets-from-eu-institutions_en.pdf
22 http://www.europarl.europa.eu/legislative-train/theme-deeper-and-fairer-internal-market-with-a-strengthened-industrial-base-services-including-transport/file-integrated-farm-statistics
23 A list of agrifood accelerators can be found on p29 of https://enrd.ec.europa.eu/sites/enrd/files/publi-enrd-rr-24-2017-en.pdf

mapping and monitoring can be helped by transparency initiatives and linking to existing documentation. A key area is livestock monitoring, necessary for a wide variety of purposes, including preservation of animal welfare standards, prevention of disease and supply chain monitoring. A generally more transparent supply chain – whether regulated by law or willingly supported by enlightened market forces – may substantially reduce the opportunities for malicious market operators to cheat, situations like that in 2013, when several European countries discovered that processed meat that was sold in supermarkets and other locations as beef, included in fact traces of horsemeat. Provenance of other products can assist with protected designation of origin (PDO)monitoring. As well as decisions at the individual producer level, state level production policy decisions and planning for farming and use of land can all be addressed.

In our research, (the methodology of which is explained at the beginning of each relevant chapter), we encountered rural economy-specific challenges around the best locations, timings and methods for successful, sustainable crop growth; protection of marine reserves through fisheries transparency; and improved access to existing agricultural support. We interviewed 6 respondents, 3 from national portals in rural countries (Ireland, Romania and Slovenia) and three experts in agriculture, fisheries and forestry. We then used these insights to guide desk research from the literature to understand more about the context and generalisability.

3. Availability of rural open data

We interviewed three experts for insights into the availability of rural population data, Larisa Panait, Co-ordinator of the Romanian National Open Data Portal; Niall Ó Brolcháin, of ODI Galway and Ales Versic, of the Slovenian National Open Data Portal. These three interviewees were selected for their extensive experience with open data portals in countries with large rural populations across the Open Data Maturity scale. We developed the insights below based on these interviews, using existing research to establish their generalisation.

3.1 Key factors of providing open data on rural populations

While member states gather statistical data on their populations across rural and urban areas, there are challenges in both creating (that is, collecting and collating) and publishing useful open data from this. These challenges can be broadly categorised as those of digitalisation, awareness, IT capacity and capabilities, usefulness and standardisation.

3.1.1 Level of digitalisation in rural areas

Even in highly urban states, smaller or more rural authorities will produce some data in less useable formats such as PDF. In large, rural states like Romania, most rural administrations are still using analogue tools, working with paper records with few or no databases. Central government aims to assist

these administrations to increase their digitalisation, beginning with budgets and expenses, which most should have in Excel tables. As with all changes, overcoming resistance is part of the process just as developing the necessary skill set.

3.1.2 Awareness for open data in rural areas

The Romanian national data portal has been active since 2013. However, our research suggests that open data barely exists in rural areas of Romania, both in terms of being published and also conceptually. Most public servants and decision makers in these administrations have rarely heard of open data before. Even potential users of rural open data – those who could benefit from it – are not familiar with it. The approach of the Romanian open data portal is therefore to help rural stakeholders understand that open data follows on from the Public Sector Information directive, and that compliance is required. To address this, they have recently been reaching out to rural public administrations to assist with facilitating this.

Conversely, in more urban Member States such as Belgium and France, there is a higher level of awareness of the importance of rural open data – especially that which promotes transparency of efforts to bridge the rural divide such as broadband connectivity and access.[24]

To achieve improved awareness, local champions are key. Last year the Romanian national data portal developed a new methodology for publishing at the national level. They attempted to disseminate this to the local level, but were successful only to county level. However, they have recently experienced increased success, with one of the less developed regions starting to publish on the national data portal.

3.1.3 IT capacity in rural areas

Ireland is one of the most developed open data countries in Europe and highly digitised. However, despite this, like Romania, some local rural authorities have no IT department. In Romania, open data is 'piggy-backing' on advanced communes who are implementing digital EU projects in non-data areas. These tend to be close to cities and municipalities, and therefore the influence of urban areas and their IT capacity and capability is reaching them – once they have data it can potentially be open.

Rural County Roscommon in Ireland (with 64.5k residents) is extremely unusual and is something of a beacon project in terms of the sheer volume of data it has published as a rural area - it has the second largest open data output in Ireland[25]. Roscommon published 124 datasets on the European Data Portal, covering areas as the graveyard survey, council planning applications, and the thatched building survey. This latter dataset in particular illustrates Dix' (2014) point about the size of datasets – it is very small, with only 37 line items. It is available in 6 formats. It is unlikely that a complete set of thatched building

---

[24] www.opendatasoft.com
[25] http://data-roscoco.opendata.arcgis.com

surveys from all Irish councils can be compiled. However, as a record on its own merits it has value. For instance, citizens may use it on which to base their own informal survey, for hyper-local purposes. Roscommon has used their own open data to provide services for residents and tourists. The winter service gritting plan has received 2,514 views to date.[26]

However, the case of Roscommon also illustrates something larger: virtually all our interviewees noted that open data rests on practices, processes and skills, and in almost all cases, these are not institutionalised. In the case of Roscommon, their exceptional output was driven by one specific employee. As seen in Romania, local champions are important for initiating open data; but as Roscommon illustrates, practices and ways of thinking about open data must be institutionalised in order to survive the loss of individual champions. The Roscommon employee has subsequently moved to another area of Ireland. It remains to be seen whether they successfully institutionalised this commitment to the publication of open data.

3.1.4 Usefulness of open data for rural areas

One reason that there is not a greater use of rural data from the open data point of view that was suggested by interviewees is because data publication is often prioritised on an 'ease of opening' basis. Given some of the collection, IT and format challenges noted above, rural data is not often the low hanging fruit that is opened first.

The nature of rural in comparison to urban local authorities is partly to do with space and culture. As described before, the geographical urban area is small and has strong coherent strands. It is therefore relatively easy to reach and serve diverse populations. In the rural Irish region of Connemara, for example, the population is thinly spread over a large area. Within this space, Irish speakers and English speakers in different cultures may have conflicting needs for resources and influence. This is further complicated by external assumptions of monoculturalism. Consequently, there may be multiple layers of data needed in the same area, pertaining to each culture.

Small groups of data are often amalgamated into larger units where individual needs and circumstances are lost[27]. For example, the island of Tiree's electric meter usage data is amalgamated with neighbouring island Coll and the much larger area of Mull, so that patterns specific to Tiree are lost.[28]

26http://roscoco.maps.arcgis.com/apps/PublicGallery/map.html?appid=ad0e44aec4ea471b905c1f20f2697b85&group=8e883303099547bdae6ed7ef920007eb&webmap=01082f0f5a52403c888e4dff115d5709 (views correct to August 2019)

[27] Dix, A (2014) Open Data Islands and Communities, https://tireetechwave.org/wp-content/uploads/2014/07/Open-Data-Islands-and-Communities-v3b.pdf

[28] https://tireetechwave.org/wp-content/uploads/2014/07/Open-Data-Islands-and-Communities-v3b.pdf

3.1.5 Standardisation

Large, decentralised countries such as Romania struggle with standardisation in many areas. The central administration receives some data from almost 3,000 local authority units which they can potentially publish. However, the format and content of this data is hugely varied, and requires substantial work to be made into a coherent dataset.

We hypothesised that, as our interviews suggested that rural data could be more challenging to collect and use, the Member States with more rural populations are in earlier stages of Open Data Maturity. The figure below shows open data maturity (in green) plotted against rural population (in purple).
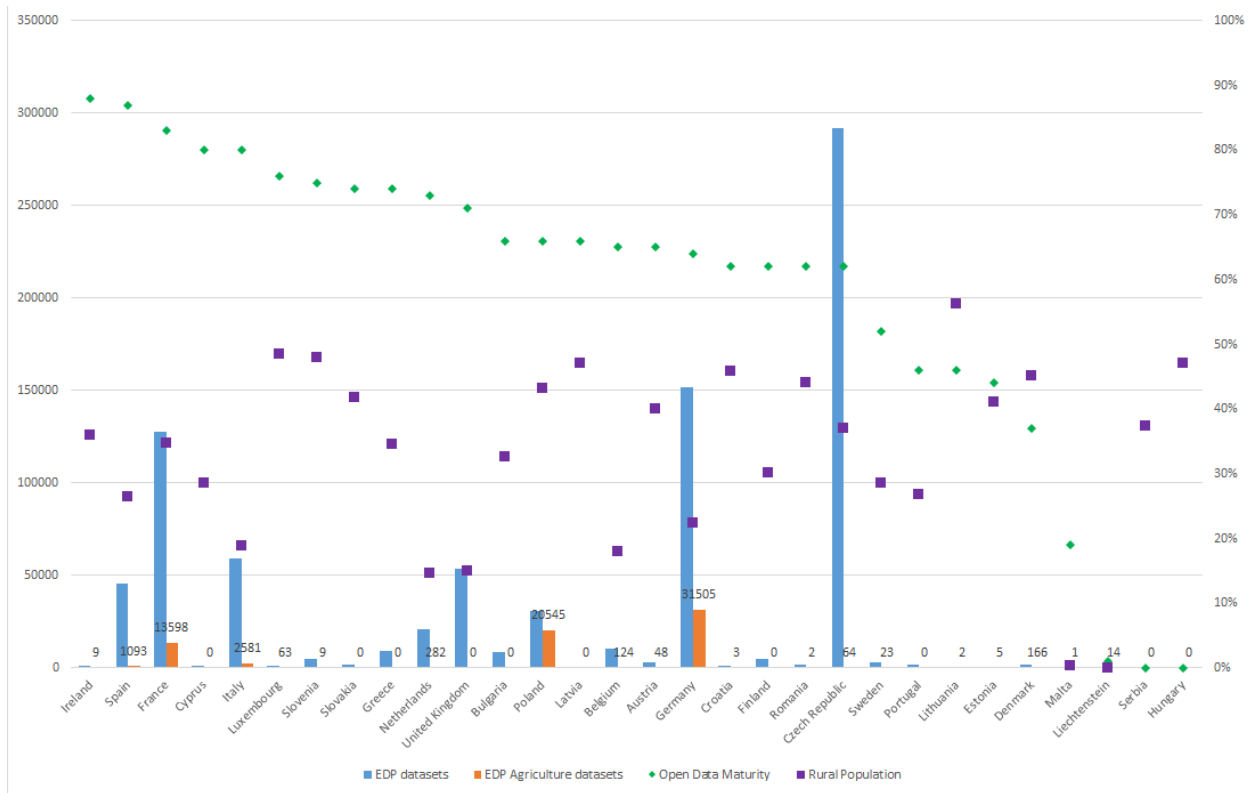
Comparison of EDP dataset representation, Open Data Maturity, and rural population, by country

This hypothesis was incorrect. We found a very weak correlation the other way, suggesting that countries with a higher percentage of rural population in fact correlated with a higher ODM. The effect is miniscule: 0.017 per percent of rural, and 0.014 per percent of urban population, which in practice is equivalent to no relationship.

3.2 Availability of open data on the rural economy

Again, we interviewed three experts in the fields of agriculture, fisheries and forestry, whose organisations either consume or produce open data[29]. These insights inform the following sections. We further analysed the European Data Portal catalogue for insights into the availability of agriculture, fisheries and forestry datasets ('agriculture' for convenience).

3.2.1 Open data from agriculture

Our first step was to compare the availability of all datasets per country against the agriculture datasets.

Share of datasets in EDP by country (in %)

---
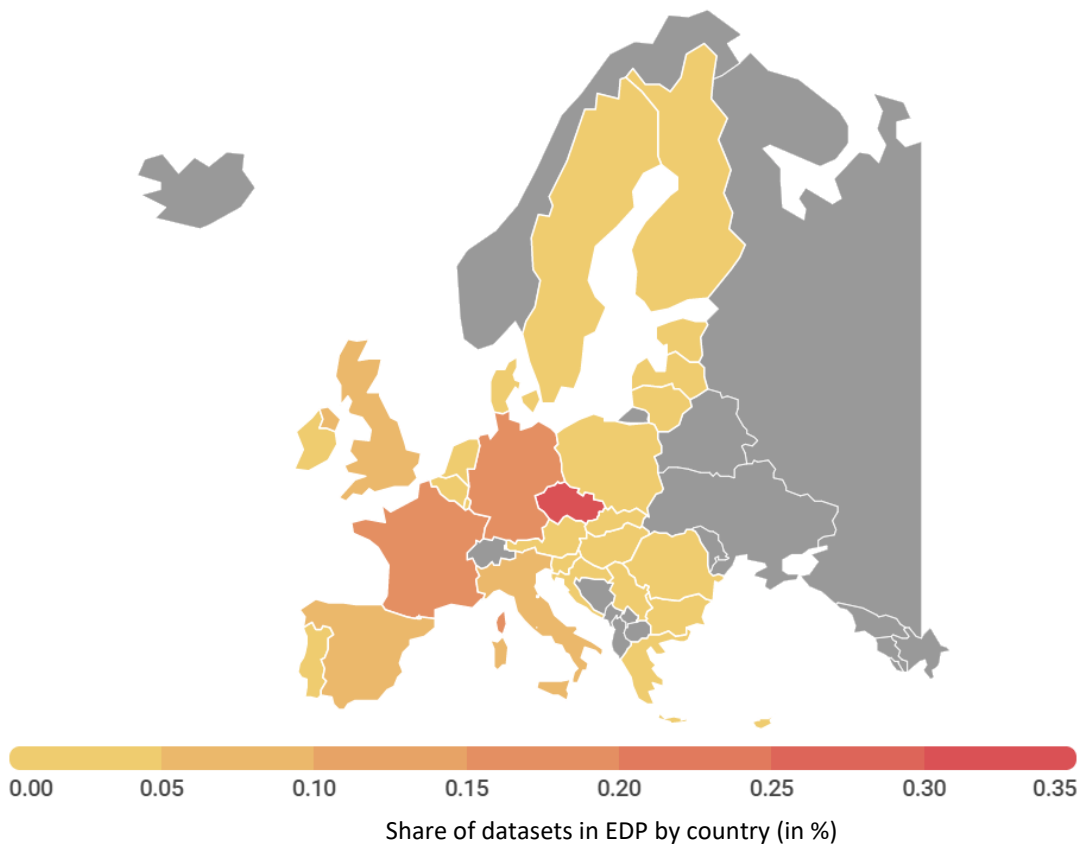
[29] Tyler Clavelle, Global Fishing Watch; Jos Berkvens, Regio Foodvalley, NL; Jonas Fridman, Head of the Swedish National Forest Inventory (by email).

Share of agriculture datasets in EDP by country (in %)[30]
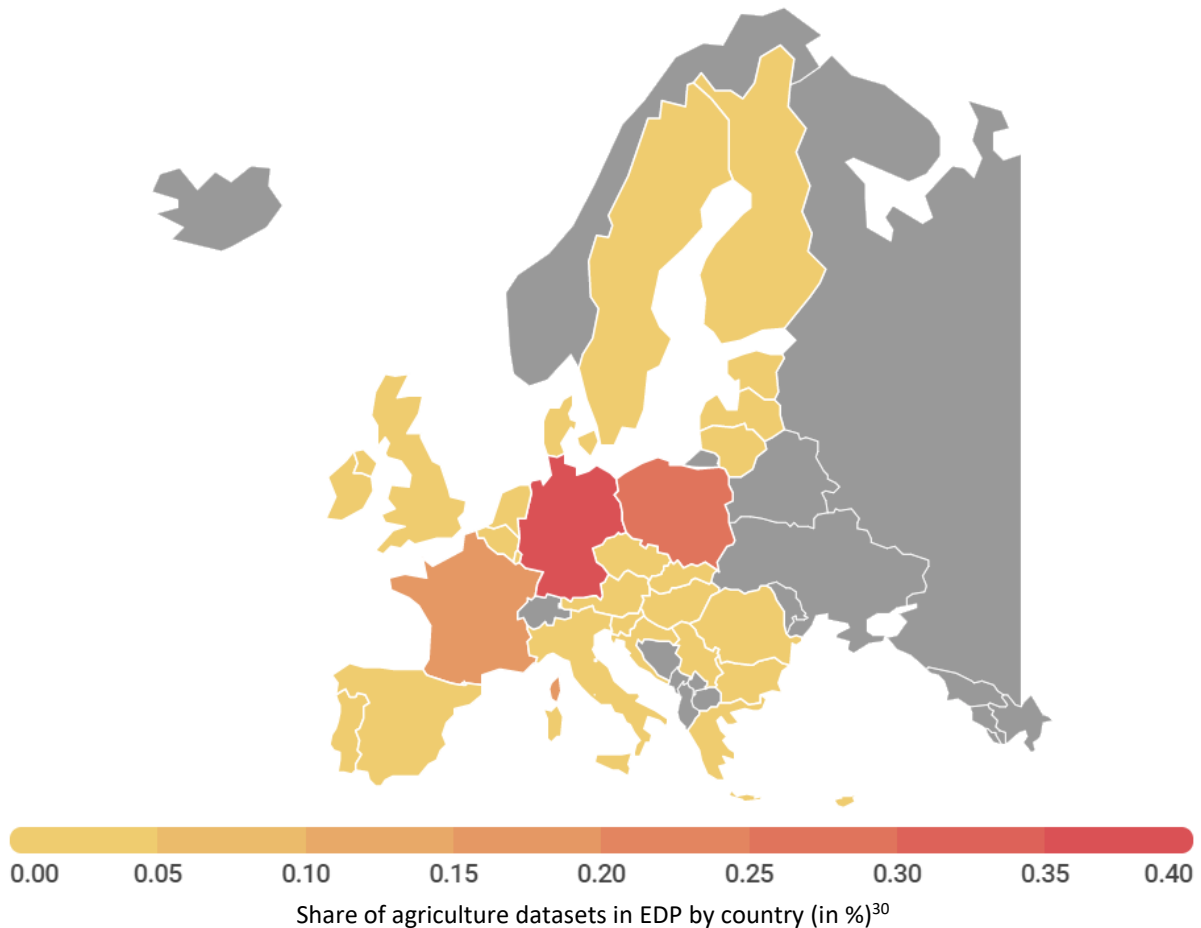
Nearly 14% (13.75%) of all datasets on the EDP are agricultural data, making it the largest category on the portal. However, it is highly concentrated – the three main contributors to agricultural datasets in EDP are Germany (agriculture 0.7% of GDP), Poland (agriculture 2.4% of GDP) and France (agriculture 1.6% of GDP), who between them share 93% of all datasets. In comparison to the proportion of GDP that agriculture accounts for, this is surprising: Poland has the 11th largest share in the EU, Germany ranges near the bottom as 26th. However, Germany in particular shares 18% of the total datasets, but 45% of the total agricultural datasets. Poland shares 4% of all datasets, but 19% of agricultural data[31]. This may be because the data is added to the national portals (from where the EDP harvests the metadata) simply because it is available, rather than following an assessment of its potential value.

---

[30] https://infogram.com/app/#/edit/c5dc7586-b641-4c13-bc57-9540f65cebb4

[31] Poland has more than doubled their data sets on EDP since the writing of this report. In September 2019 they added 37,000 data sets, of which 30,000 are agricultural, making them the new largest contributor. ahead of Germany. At the same time Germany has reduced the count for agriculture data sets. It is not possible to tell whether this is because data sets have been moved to other categories, or if they were removed from the portal. The overall findings still stand, so we have left them with the original counts.

Other countries with a high proportion of agriculture in their GDP, such as Serbia (6.2%), Romania (4.3%), Greece (3.7%) or Bulgaria (3.7%) make little to no agricultural datasets (or datasets in general) available. Most remarkably, the Czech Republic (agriculture 2% of GDP) shares 34% of all the datasets in the EDP, but less than 1% of agricultural data. It is difficult to draw conclusions from these basic comparisons, as the numbers alone do not say much about the context. It may be that these countries do not collect data in a systematic fashion, and consequently it is not harvested by the EDP; it is also possible that these countries, despite being agriculture intensive, lack awareness of the value of data about this. The key finding is the inconsistency between the relative value of agricultural data to these countries and the availability of datasets, and also the inconsistency across the Member States in general. We discuss some qualitative case studies below to add more context to these considerations.

Although there is no discernible pattern in the relationship between proportion of GDP and agricultural data shared, this does raise a number of crucial questions. Why are so comparatively few member states sharing 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? How valuable are the agricultural datasets, or do they have a very long tail?
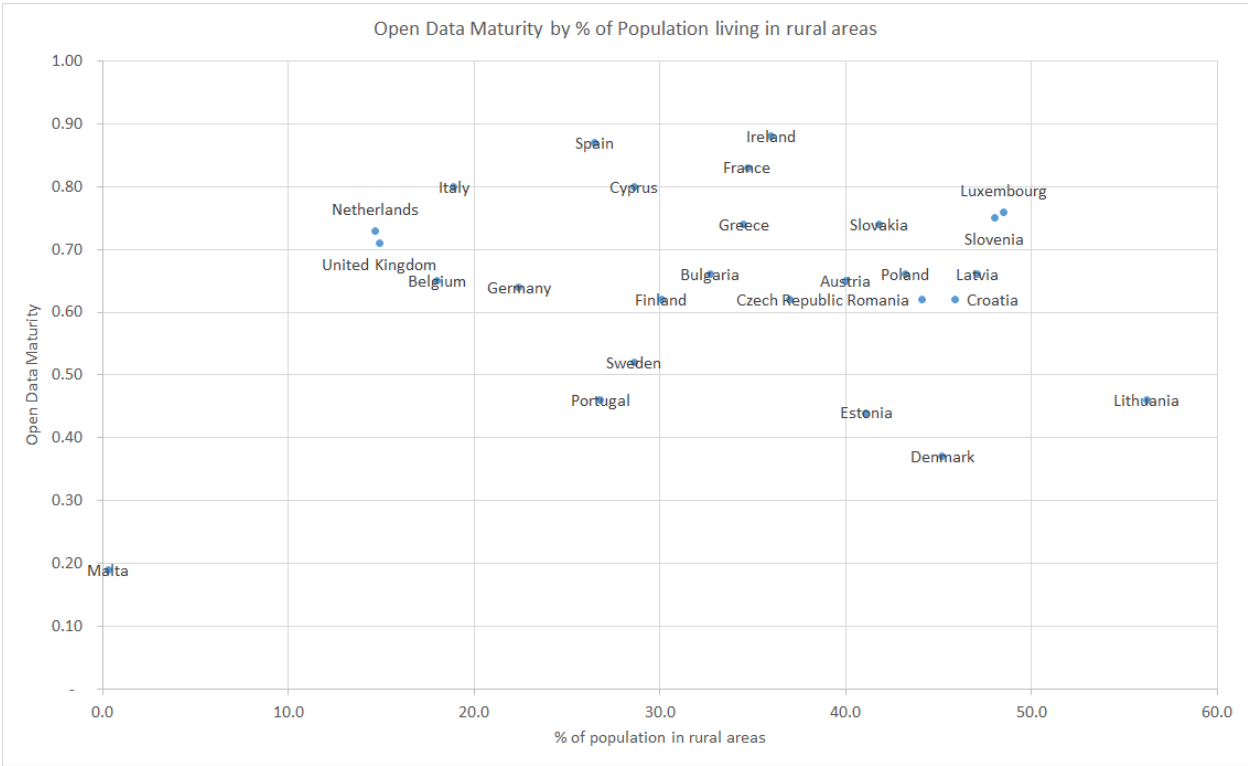
There is also no significant correlation between countries' ODM and their representation in the EDP with datasets in general, or with agricultural datasets specifically.

Therefore, we need to look elsewhere for drivers of opening national agricultural data. Some interviewees suggested that this continues to be subject to a kind of 'availability bias': countries publishing a great deal of agricultural data are doing so because it happens to be available, rather than because of any awareness of its value or necessity. However, if data is being opened because it is 'low hanging fruit' then it is less likely to create impactful re-use opportunities. Other challenges to availability we encountered included agricultural data that was only available to citizens or organisations that applied for it on a request basis, and often shared only as a web page or PDF rather than in easily re-usable formats; and datasets being published on the Ministry of Agriculture site, but not on the relevant national portal.

The availability of open data, the availability of useful open data and the awareness of the availability of open data are entwined. Concerning availability, there is still data that is, in principle, open, but is not publicly accessible. For example, Slovenian soil and forestry datasets are available on request from the relevant ministry, but not on the open data portal. Regarding the availability of useful open data, we found in our interviews that many institutions and provinces are still operating at an *ad hoc* level of data publishing. Performing meaningful, timely research under these circumstances is difficult. All our interviewees from national data portals described intense resource challenges in trying to develop a more consistent level of publishing, frequently with two or three employees attempting to engage with dozens of ministries and institutions. Finally, regarding the awareness of the availability of open data, potential users are still unaware of the existence of much open data. Interviewees noted that the potential use was likely to be driven by urban-based initiatives. As one respondent said: "There are few fishermen sitting in a boat thinking of a way to use open data." Thus, there is a gap between the rural industries and audience that might benefit from the use of open data, and those who are working with open data. While open data about fisheries might be of much value to individuals in the fishing industry, they are unlikely to have the skills or disposition to realise this value.

However, an audience that is located rurally and that both creates and consumes a vast amount of data is farmers. In many cases, patterns that might affect livestock or crops - particularly to do with disease in both – cannot be seen in the data of one farm but can be easily spotted by combining data from multiple farms. In this case, it may make sense for farmers to crowdsource the data themselves. While there may be some commercial sensitivity in the data, there are also strong arguments for publishing openly – for instance, it may help government departments in identifying and combating the spread of disease and planning prevention.

Case Study: Creating open data for successful agriculture[32]

Soil moisture levels are key for successful agriculture. It is currently measured through satellite imagery with a very small number of ground measurements. GROW Observatory, a Horizon 2020 project, collected data on soil moisture from thousands of participants across Europe, via sensors. The collection of ground truth data enables the calibration of remote-sensing data and supports the interpretation and

---

[32] www.growobservatory.org

analysis of what is being recorded. Small scale farmers and growers in Poland, Luxembourg, the Canaries, Greece, Croatia, Italy, the UK, Austria, Hungary and more supply soil moistures measurements that are published as open data. In return, GROW provides online training in topics such as how to run growing experiments and citizen science. The measurements also have significant impacts on scientific understanding of the role of soil moisture on several key environmental variables such as soil erosion, flood risk, or fire risk.

3.2.2 Open data from Fisheries

The next example illustrates a case of the non-profit organisation Global Fishing Watch (GFW) that opens fishing data with a very specific purpose, protecting fish reserves and marine life. Fishing activity and stocks have long been a fraught issue for some member states from both an environmental and an industrial point of view. In our interview with Tyler Clavelle from Global Fishing Watch it got evident that many of the old barriers to opening data – data hoarding, lack of incentives, unclear licensing – still prevail. The example is illustrated in the case study below.

Case Study: Making open data available to protect fish reserves and marine life

Global Fishing Watch (GFW) produces maps of fishing effort and presence at very high resolution and publishes them as open data.[33] Europe is where most fishing effort is, and has the highest concentration of vessels, behind China. GFW collects 30 million data points per day, and distils that into downloadable datasets. Currently, the challenges of their raw data is such that it takes them a full year of processing to make the data understandable and useable. Their goal is to automate this, to make data accessible faster, while still allowing for quality control.

Uses
GFW's data has two main uses. The first is to facilitate a better understanding of where fishing occurs. The UN Food & Agriculture Organisation (FAO) oversees global fishing statistics. Countries self-report to FAO what was fished, how much, and where, largely based on estimates. GFW's data is much more fine-grained, and is therefore better scientific data for studies of marine ecosystems – a "game changer for marine research" on species distribution, fishing, bycatch, and seabirds, according to GFW.

The data is also used for monitoring, helping countries understand what happens in their waters. Previously countries could only track their own vessels, if that. GFWs data helps countries prioritise and make policy decisions about fisheries and the environment, and GFW engage often with governments who use their data for that purpose.

---

[33] https://globalfishingwatch.org/datasets-and-code/

Users fill in a short survey to gain access, which helps GFW collect other use cases. Examples include the positioning of transatlantic telecommunications wires.

Challenges
Even if distilled, GFW works with an immense amount of data. While they want to make as much of it available as possible, most potential users struggle to use that amount of data. It requires extensive computing resources, and they don't want to overwhelm their users, nor do they have the capacity to meet the differing needs of every potential user organisation. Therefore, they carefully balance what they publish.

As a small organisation, GFW want to engage with and encourage their users, but delivering support and training is not always practical. They mainly work with academics, governments and NGOs, not all of whom are necessarily well funded enough and / or sufficiently skilled to use the data.

Further, while their underlying automatic identification system (AIS) data is open or purchasable, it can also be politically sensitive. Although GFW's objective is not to judge the legality of fishing, users of the data may do so, for example, by combining it with local knowledge. GFW cannot therefore control the narrative of how their data gets used.

Finally, most of the tracking data for mammals or sea birds belongs to individual researchers and, despite open science initiatives, remains closed. While GFW are trying to facilitate collaboration and convene researchers around these themes, legal access and licences are further complicating this.

3.2.3 Open data from Forestry

The third key rural industry sector is forestry. Reliant as we are on trees for wood, paper, and of course, oxygen, there are numerous examples of national forest organisations making forestry data available, especially in the Nordic member states. An example is illustrated in the case study below.

Case Study: Swedish National Forest Inventory

The Swedish National Forest Inventory has the task of describing the state and changes in Sweden's forests. The information collected is used, for example, as a basis for forestry, energy and environmental policy in Sweden. Data from the Swedish National Forest Inventory is part of Sweden's official statistics.

Forest information is needed for many different purposes, including supply studies for the industry, e.g. long-term scenario analysis, operations planning, biodiversity studies and carbon sequestration for climate studies.

They opened up their data as part of an objective to, "connect with more users of forest data as well as with users in need of forest decision support. We hope that companies will develop new products based on the data and the knowledge that we can provide to them."

Using LIDAR data from the Swedish national LIDAR scanning, and ground truth from Swedish NFI plots, forest companies have developed tools where each pixel representing 10x10m has been assigned forest attributes (height, density, volume etc.). This makes management planning in the forest so accessible that it can virtually be done on a smartphone.

4 Re-use and re-users of rural open data

4.1 Open rural population data

In the same way that scale affects the publication of data, scale affects its use. Cities are host to a greater number of data consumers (those who are actually or potentially using open data to create a product, service or insight). This has a number of effects. With fewer users actually based rurally, and city consumers based in the city, there is not enough awareness of what data there is available, and how it could be used.

Our interviewees agreed that the greatest potential of rural open data was to *impact* rural rather than have rural users. This is 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. Our research shows that in Romania, overall user numbers are low, and focused on the capital and large smart cities who are working on their own open data portals.
This mirrors other research on 'smart and competitive rural areas', such as that done by the European Network for Rural Development, which finds that the 'digital divide' is still a very real barrier to non-agrarian productivity rurally. [34]

This, then, challenges the idea that the smart cities concept can simply be transplanted to 'smart rural'. However, certain concepts of smart cities have extended to reflect the fact that urban areas are not completely independent of the rural areas that surround them. This approach is known as 'smart regions'. Smart regions, such as Cork Smart Gateway, are focused on the city as well as the rural regions surrounding it. The term 'smart region' has also been used to discuss city clusters, without acknowledgement of the rural/urban relationship. We follow the Helsinki Smart Region definition, which states 'smart city plus smart countryside = smart region'. That this extension from the city is natural is

---

[34] https://enrd.ec.europa.eu/enrd-thematic-work/smart-and-competitive-rural-areas/rural-businesses_en

exemplified by 'accidental' smart regions. For example, Smart Dublin is a smart city initiative, but covers some regional areas around Dublin as well.

Smart Atlantic Way[35] is a project assessing the feasibility of creating a smart region in the nine mainly rural and sparsely populated Irish counties of Cavan, Clare, Donegal, Galway, Leitrim, Mayo, Monaghan, Roscommon and Sligo. The investigators have identified 'smart community' as a key component of smart regions: allowing communities to identify their specific needs as the basis for developing initiatives in the areas of protection of vital local services (e.g. community policing), tourism, and cultural facilities, by exploiting digital and smart technologies. Giving communities the ability to publish their own data means that they are able to make their own decisions about what is said about them[36].

"Knowledge creation, dissemination, and use are essential for keeping regions smart."[37] The issue to highlight is that there is a culture around open data that is driven by knowledge institutes, universities and governments – which are largely based in cities.

Case Study: The role of Smart Villages in Smart Regions[38]

In 2018 the EU launched its action for Smart Villages. In the definition of smart villages, it states, "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."[39] Sometimes, the relationship between cities and regions has sometimes been seen as a zero-sum game - what the cities gained, rural areas were thought to lose. However, the OECD finds that in nearly all of its member countries, it is the rural areas close to or accessible from cities that are the fastest growing in terms of Gross Domestic Product (GDP), productivity and population For smart villages, to succeed, it is not just a case of overcoming the urban-rural divide, but of harnessing the unique potential of each for mutual benefit.[40] It is also important that local actors - the metaphorical fishers in the boat mentioned above - are enabled to unlock their potential.[41]

In seeking to achieve the most successful outcomes smart village projects are increasingly cooperating – both with other similar rural areas and with their associated small and large population centres – to develop mutually beneficial planned territorial solutions. For example, the French government has supported a series of so-called reciprocity contracts between cities and their surrounding countryside.

---

[35] N. Ó Brolcháin, A. Ojo, L. Porwol, D. Minton, C. Barry. 2018. Examining the feasibility of a Smart Region approach in the North West Atlantic and Borders Region of Ireland. In Proceedings of the 11th International Conference on eory and Practice of Electronic Governance, Galway, Ireland, April 2018 (ICEGOV'18), 7 pages. https://doi.org/10.1145/3209415.3209512

[36] Dix, A (2014) Open Data Islands and Communities

[37] https://timreview.ca/article/932

[38] https://enrd.ec.europa.eu/sites/enrd/files/enrd_publications/publi-enrd-rr-26-2018-en.pdf

[39] https://enrd.ec.europa.eu/sites/enrd/files/enrd_publications/publi-enrd-rr-26-2018-en.pdf

[40] https://enrd.ec.europa.eu/sites/enrd/files/enrd_publications/publi-enrd-rr-26-2018-en.pdf

[41] https://europa.eu/regions-and-cities/programme/sessions/627_en

in France, the City of Brest and the rural Centre Ouest Bretagne area signed a 'mutual-reciprocity contract' in 2016. In the energy sector, the Brest-Brittany contract precipitated a new wood energy cluster, which aggregated waste from rural timber to feed a biomass incinerator in the city, that fuelled public lighting. The reciprocity contract has also increased health care collaboration, addressing Center-West Brittany's risk of becoming a medical desert.

These existing infrastructures and initiatives provide a solid model for the development of collaborative data smart regions.

4.2 Open rural economy data

Rural economy open data use is affected by the same knowledge and skills issues as rural population data. During our interviews, we discovered national portal representatives had clear ideas for rural economy open data applications relevant to their situation. The use of agricultural data for confirming the provenance of fruit and vegetables was one such area with multiple applications.

In the case of Slovenia, the country is at a trading crossroads, neighboured by Austria, Croatia, Italy and Hungary. This makes it difficult to trace the provenance of fruit and vegetables in the market, and each year many more tons of produce are sold labelled as 'produce of Slovenia' than are actually grown[42]. Open data would allow interested parties to monitor which produce truly was Slovenian, and which that of their neighbours, as it the volume of produce sold in the markets could be cross-checked with the volume of produce that relevant acreage is able to produce, thus increasing the reliability of the information.

Slovenia is relatively early in its open data journey, but that does not stop it from having ambitious ideas for its use. Tracking the provenance of meat from farms to slaughterhouses to restaurants was another posited use for open data. Further lighthouse examples are illustrated in the case studies below.

Case Study: Foodvalley for healthy and sustainable food

Netherlands-based Foodvalley is the leading agro-food centre in Europe; the top region for knowledge and innovation in healthy and sustainable food.[43] Foodvalley Region is a framework of cooperation involving eight municipalities with altogether 350,000 residents, and many educational institutions and businesses.

The co-ordinating group is a small organisation. They use data of all types, including open data, as a basis for advocacy. They are currently looking at livestock demands and the implications on CO2 levels. This then allows them to draw conclusions about agricultural developments, the size and

---

[42] Interview with Alex Versic of the Slovenian National Open Data Portal, August 2019
[43] https://www.foodvalley.nl

number of farms, use of land, etc. They advise policy-makers based on the understanding derived from this data.

Again, while some individuals are highly skilled, as an institution and across municipalities, they report that using open rural data is not institutionalised. Across the area of the Netherlands dealt with by FoodValley, there is little consistency in the open data processes of the 8 municipalities involved.

As well as government institutions, Foodvalley uses open data of all kinds from Wageningen University. National statistical data on housing, agriculture and livestock is also easily obtainable. Much of the rural data is available on a national level, which means they find it easier to obtain than urban data, which is affected by the fact that availability differs between municipalities, with varying systems and set-ups. Occasionally provinces and municipalities will differ from official records. Bringing together data from all the municipalities to get an overview of the entire area is, therefore, extremely challenging, and sometimes involves a level of educated guessing.

Case Study: A Slovenian open data success story

EVineyard is an app for vineyard management[44]. It assists vineyard owners in finding the best time to spray vines to ensure efficacy and reduce run-off, which in turn reduces costs. eVineyard is used vineyards in multiple countries around the world, from single hectares to the large producers. eVineyard mainly aggregates and enriches geospatial and soil data, but also uses open data on pesticides and weather to provide its service.

Case study: Countryside and Communities Research Institute[45]

The CCRI wished to assess the level of access to information and markets in rural Britain. As a proxy, they used the average speed of the broadband connection to farm properties within each unitary authority or county. To achieve this, the open broadband data from the telecommunications regulator was joined to the Ordnance Survey postcode boundaries, and then overlaid with a spatial dataset of built-up areas from the Office for National Statistics (to remove built up areas).

Using GIS software (QGIS), the average download speeds within each county/UA were then calculated. The results of this data analysis show that (as of 2015) rural broadband speeds in the UK were generally poor with only 7 out of 49 counties/UAs achieving the minimum average download speed.

.

---

[44] https://www.europeandataportal.eu/sites/default/files/slovenia_-_evineyard.pdf
45 http://www.ccri.ac.uk/recent-projects/mapping-rural-broadband-with-gis/

Case Study: Reducing Bureaucracy in Forestry Planning in Spain[46]

A pilot in the Smart Open Data project focused on forest management in Allariz and Mecida, in Galicia in North-west Spain. The stakeholders were forest owners, foresters, public administrations, seed-harvesting companies and plant and seed nurseries.

The main concern for owners and foresters was knowing which species would be most suitable for a land plot, and then obtaining that particular species of seed. Species selection before the project was mostly based on experience or intuition. Public seedbeds, known as Admission Units, are controlled by the administration and could not be visualized online, so a public officer had to be contacted in order to know which seeds are accessible and where they are located.

By making the seedbed location and content (and growing information) available online, the project intended to remove intensive overheads that included substantial paperwork and almost as much footwork.

However, in all the countries we interviewed, agricultural datasets were not highly requested. The most requested datasets in Slovenia, for example, regarded vehicle licensing, public sector salaries, and building permits, with agricultural statistics on crops and the number of livestock much further down the list. There may be a number of reasons for this – one of which could plausibly be that there are fewer use-cases available to demonstrate what might be done.
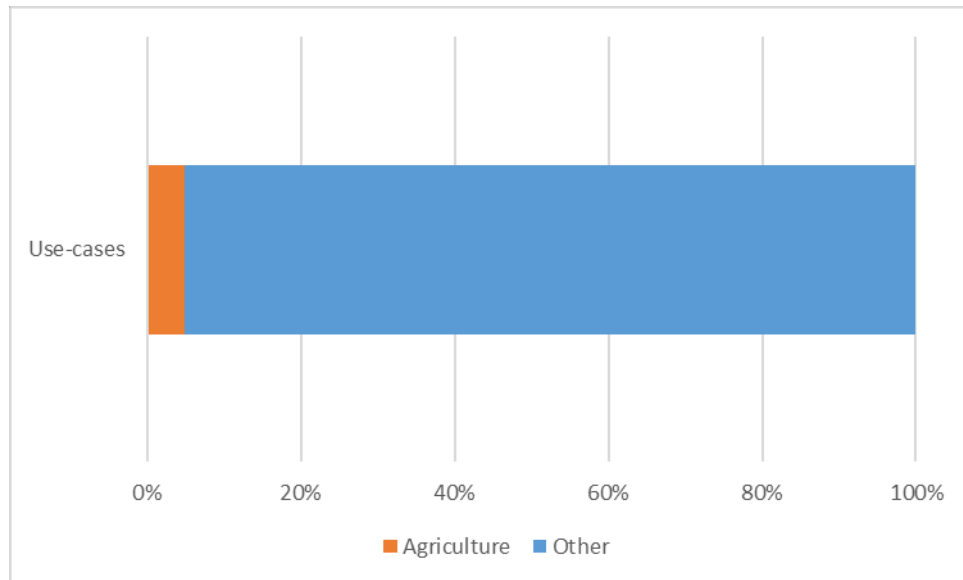
The European Data Portal, for example, contains use cases from both publishers and re-users of open data. These are intended both to showcase impact of open data, but also to inspire future use.

---

[46] Smart Open Data Project 2014
http://www.smartopendata.eu/sites/default/files/SmartOpenData_D5.1_Rationale%20of%20the%20Pilots.pdf

EDP use-cases comparison: Agriculture vs all other categories

Of the 548 use cases listed on the EDP only 28 are concerned with the agriculture sector, and only 21 are based in Europe.[47] This may be because Member States, while contributing data on the EDP, are not adding use-cases. For example, Ireland showcases two uses on their national portal, but none on the EDP. Romania or the Czech Republic do not publish any case studies on either national portals or the EDP, while Germany shares no case studies on their national portal, but one on the EDP. Most of the data open on the portal concerns agriculture, but this data is either not used extensively, or if it is used, best practice of this use is not shared in a proportionate, meaningful way.

---

[47] Based on a search of https://www.europeandataportal.eu/en/using-data/use-cases using the sector filter 'Agriculture, Forestry and Fisheries'.

5. Recommendations

Six key recommendations to policy makers who wish to encourage the use of data at various levels of government at the rural level emerged from our analysis of the rural open data situation:

Data Collection

1. Support local champions to establish and institutionalise the processes and practices of open data in rural administrations, with close relationships to urban open data groups (this may be the national portal, a nearby city portal or an activist group. This could potentially be achieved by the development of training materials and incentives for local champions to participate in this training. This will also assist with improved digitisation and standardisation both within the rural administration and across other, related regions. Bath:Hacked[48] is an example of a community group working with the local North-West Somerset council in this way.

2. Highlight the importance of the agricultural High Value Datasets to the Member States that have been identified as having a gap between their agricultural percentage of GDP and their open agricultural datasets. This could be delivered as a training workshop showing examples of the variety of agricultural data applications, and the development of a greater number of use cases following those already on the EDP.

Data Opening

3. Create collaborative links with complementary data owners, including Not for Profits, NGOs, researchers and private organisations, that allow data to be opened and made available together. This could also be done via alternative models by which data can be shared, when open licensing is not an option (for instance, as exemplified by the JoinData rural data sharing platform in the Netherlands[49]).

4. Invest in understanding the key factors of smart regions that mean they are more fit for purpose than simply being smart cities with sprawling reach, in order to identify key data to open. How does the ecosystem differ? What are the challenges? How is a 'smart community', with the bespoke approach for each community this implies, successfully integrated with the larger 'smart city' concept? Helsinki Smart Region and the work on the Turino-Milano smart region are the current models to follow.

---

[48] www.bathhacked.org
[49] https://eudatasharing.eu/examples/data-sharing-agricultural-sector

Data Reuse

5. Establish links with urban areas, via such instruments as hubs or 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'.[50]

6. Identify and support a wider range of rural-specific challenges, such as out-migration, health solutions that address distance, and food provenance. (These can/should still be run in urban settings.) Support might take the form of assisting with the opening of key datasets, or financial support. One example would be to create a 'rural' category in the EU Datathon, which in 2019 had categories for economics and finance.

6. Conclusion

The need for institutionalisation of the commitment to, and appropriate skills for, open data, remains a key barrier to the publication and use of open data in both rural areas and sectors. In these areas it cannot outstrip the wider data economy, but it is to be hoped that increasing focus on the importance of data and data capacity will also give impetus to open data and allow it to grow.

In performing the research for this analytical report, we read many reports (and policies) both on open data, and on rural populations and industries. We found it striking how rarely the two worlds overlapped. We read many rural development reports urging an increase in digital accessibility, hubs and businesses, but very few even contained the word 'data', far less 'open data', let alone views on how the data sharing economy would benefit the rural sectors. It is crucial that rural populations and industrial sectors are not (unintentionally) excluded from the benefits of the data economy. To do this, 'smart rural' must become something more real than a few isolated projects, but also something more than a faint facsimile of smart cities. The greatest opportunity for this lies in 'smart regions', that create the necessary links between urban and rural, while acknowledging the differences.

---

[50] EU Rural Review p31 https://enrd.ec.europa.eu/sites/enrd/files/publi-enrd-rr-24-2017-en.pdf