ClairCity: official statistics as an enabler in a citizen-led European air quality project

Keywords: air quality, citizen involvement, data management, model based policy making, visualization, gamification

1. INTRODUCTION

ClairCity is a four year European project (2016-2020) working directly with citizens and local authorities in six countries around Europe with the aim to improve air quality. Nine research institutes, six municipalities and one national statistical institute – Statistics Netherlands - work together on models, scenarios and policies to help cities decide on the best local options for a future with clean air and lower carbon emissions. This paper explains the ClairCity project in brief and highlights the enabling role of official statistics in this international, multi-disciplinary project.

Six cities are partners in the project; Amsterdam in the Netherlands; Bristol in the UK; Ljubljana in Slovenia; Sosnowiec in Poland; the Aveiro region in Portugal and the Liguria region around Genoa in Italy. Each city faces different issues and causes of air pollution, but all of them are working to improve their air quality. From an official statistics point of view this creates an interesting challenge. Although local circumstances differ, a generic model based approach is applied which asks for good quality data on demography, citizens behaviour, traffic patterns, energy (both production and consumption) and other aspects that may influence local air quality now and in future.

A key element of the ClairCity approach is to use existing data to drive the modelling activities. It uses data that is already collected by local, national and European institutes to develop new models of urban air pollution and carbon emissions and scenarios to reduce emissions in future. Although initially developed for the six participating pilot cities / regions, the outputs will be generic so that they can be reused in every European city with over 50,000 residents. This will make it easier for cities to identify changes that they can make to reduce emissions and make a positive change in peoples' lives.

A consequence of this approach is that the project has a very broad and extensive data demand. Moreover, the data needed is usually scattered around multiple data sources and usually not very standardised. That is where official statistics come in. Both the identification of data, the management of multiple - maybe incompatible - data sources and their metadata as well as the visualization of scenario effects are fields where official statistics can - and should - add their knowledge and experience.

2. DATA DISCOVERY AND DATA MANAGEMENT

Statistical offices have a long tradition of data collection. These days survey-based data is supplemented and wherever possible replaced by administrative data and new data

sources such as web scraped data and big (sensor) data [1]. The data discovery and data management is one of the core skills of official statistics institutes. These skills can be offered to others as well.

2.1. Data portal

One of the cornerstones of ClairCity is the use of multiple levels of geo data on different regional levels. These data has been brought together by a mix of research partners and city and has to be managed. At the start of the ClairCity project Statistics Netherlands created a data portal, based on the well-known CKAN system [2], to manage the raw, intermediate and production data. Apart from core data management concepts such as organisations, access control, grouping and tagging, a conceptual meta model was added that reflects the specific view of ClairCity on their data. Figure 1 shows the data portal front page using this model. This data portal has been used heavily in the project and this shows the enabling role a statistical institute might have offering its knowledge and experience on (meta) data management in research projects.

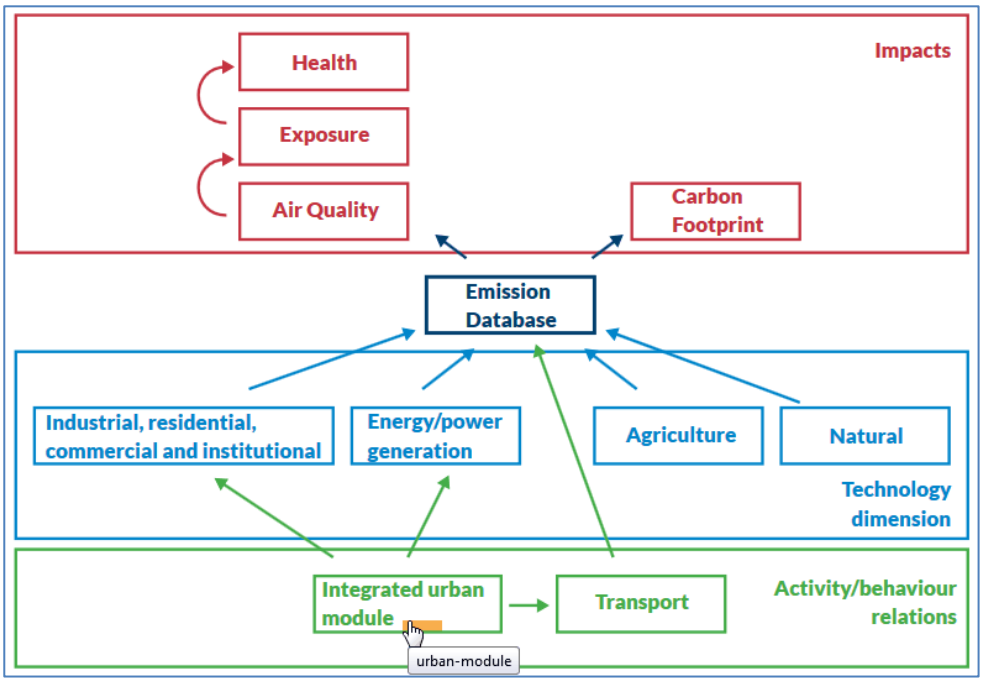

Figure 1: ClairCity data portal

2.2. Best practices on practice activity data

Without access to basic demographic urban statistics, data on the presence and use of city transportation, estimates on energy production and consumption of households and enterprises and data on industrial, agriculture and natural resources, it is would not be possible to develop a clear view on future air quality scenarios for a city. More generally, data on the behaviours and activities of people in a city are needed. This type of data is known as practice activity data [3]. During the project it became clear that these data are not always readily available, and if available, they are not always of sufficient quality. Ideal data requirements are almost never fully met and therefore in many cases data being used in practice is the result of a compromise.

One of the advantages of the involvement of six very different pilot cities each having a different data landscape is that a lot of experience has been gained identifying, analysing and processing available data sources. These experiences will be collected and written

down in a deliverable on best practises for collecting practice activity data, to be used by any European city with over 50,000 residents. The statistics office plays a leading role in this data discovery guide.

3. DISSEMINATION

Air quality improvement scenarios are complex to explain to the public. Some scenarios may show quick wins but only for a limited period while others may show positive effects for a longer period but need longer time to become visible. ClairCity needed ways to visualize such complex dependencies in an understandable way.

3.1. Dotmaps visualizations

Statistics Netherlands experimented in earlier projects with dotmaps to visualize demographical variables on a detailed regional level while preserving privacy aspects [4]. This technique appeared to be useful in ClairCity. This technique was used to visualize variables that were relevant for scenario development. Figure 2 shows a dotmap for Amsterdam. Among the variables that can be chosen are household type, household income, dwelling type, the use of district heating, number of cars, the existence of a wood stove or fireplace and emissions from gas consumption. It is planned to extend these maps with air quality layers and a time dimension so that dependencies among practise activity data, emissions and scenarios can be shown graphically.

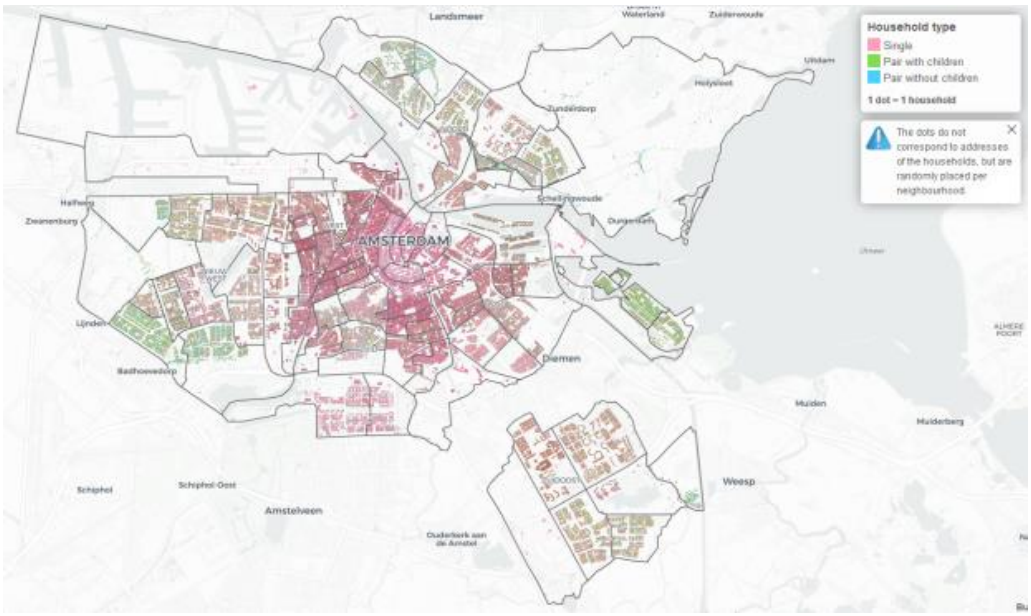

Figure 2: Dot map visualization of household types in Amsterdam

3.2. Gamification

Participating in a European project such as ClairCity is not only a matter of bringing knowledge and expertise, we as a statistics community can also learn from other communities. One interesting example in ClairCity is the use of gamification to improve citizens' engagement. The University of the West of England (UWE) has a lot of experience using gaming concepts in policy oriented projects and applied this experience in ClairCity [5]. The player, in the role of the mayor of the city, decides on the actions to take to save the city. Not only is this a nice way to explain – maybe complex – model

based predictions to the public, it also attracts people not interested in data or models to the subject. We think this principle could also be applicable to the dissemination of official statistics. Figure 3 shows the game for Bristol and Amsterdam. In a cartoonish scene the user faces the ever growing problem of air pollution and needs to take decisions that influence environment, health, society and economy (meters on top) to save the city.

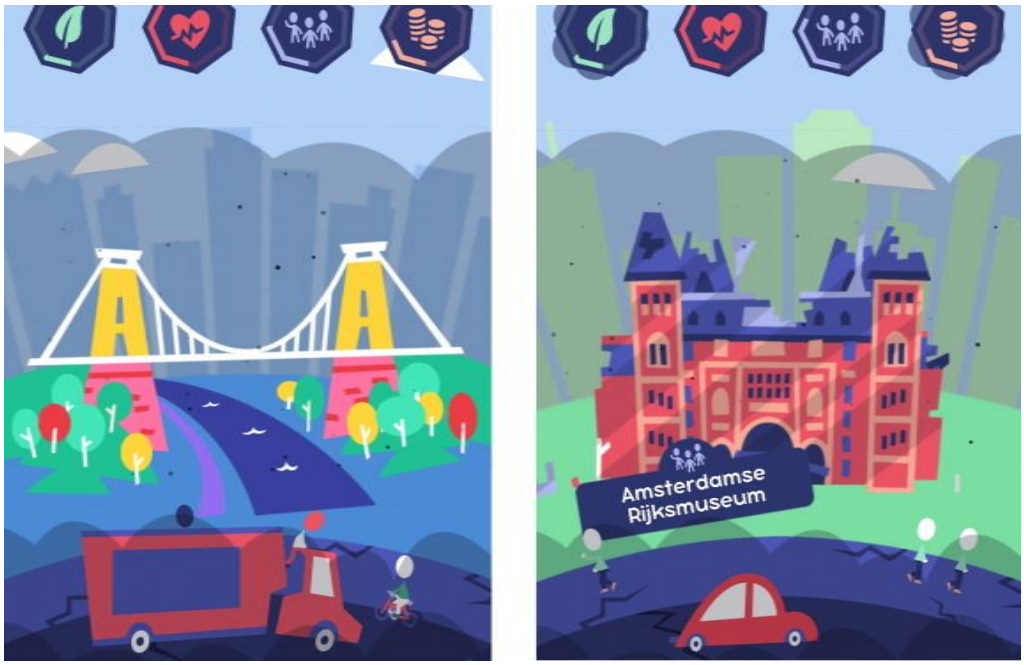

Figure 3: ClairCity game for Bristol and Amsterdam

4. CONCLUSIONS

The role of official statistics in society can further increase by participating actively in environmental projects such as the European ClairCity project for improving air quality. Statistical organisations can offer their knowledge and experience on data retrieval, data and metadata management, data processing, data analytics and dissemination. Moreover statistical organisations can learn from best practices in dissemination in such projects to apply them in official statistics. The use of gamification to explain complex models and attract different use groups is an example.

REFERENCES

[1] M. Sebille et al, Final Technical Report, ESSnet Big data, 30 June 2018

[2] Open Knowledge Foundation, User Guide, CKAN 2.8, https://ckan.org

[3] T. Chatterton et al., Putting people at the heart of air quality management (2018). Air Quality: Science and Application, Barcelona, Spain, 12-16 March 2018

[4] M. Tennekes, Aggregation of dots (2018). Paper presented at the seminar Methods for Big Data in Official Statistics, 4-5 October, Heerlen, The Netherlands.

[5] E. Hayes et al., ClairCity Project: Citizen-led scenarios to improve air quality in European cities, Proceedings WIT 26th Int. Conf. on Modelling, Monitoring and Management of Air Pollution, 19-21 June 2018, Naples, Italy

ClairCity is funded through the European Union’s Horizon 2020 programme.