Sustainability of (Open) Data Portal Infrastructures

Developing Microeconomic Indicators Through Open Data Reuse

This study has been prepared by the University of Southampton as part of the European Data Portal. 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

Esther Huyer
Email: esther.huyer@capgemini.com

Written and reviewed by:
Jacqui Ayling
Email: j.a.ayling@soton.ac.uk

Elena Simperl
Email: elena.simperl@kcl.ac.uk

Johanna Walker
Email: j.c.walker@soton.ac.uk

Last update: 02.03.2020
www: https://europeandataportal.eu/
@: info@europeandataportal.eu

DISCLAIMER
By the European Commission, Directorate-General of Communications Networks, Content and Technology. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use, which may be made of the information contained therein.

Luxembourg: Publications Office of the European Union, 2020

OA-02-20-167-EN-N ISBN: 978-92-78-42150-2 doi: 10.2830/63195

Note: this document is part of a series of research reports developed on the topic of "Sustainability of (open) data portal infrastructures", all of which are available on the European Data Portal at https://www.europeandataportal.eu/en/impact-studies/studies .

The series is made of the following reports:

1. A summary overview
2. Measuring Use and Impact of Portals
3. Developing Microeconomic Indicators Through Open Data Reuse
4. Automated assessment of indicators and metrics
5. Assessment of Funding Options for Open Data Portal Infrastructures
6. Open data portal assessment using user-oriented metrics
7. Leveraging distributed version control systems to create alternative portals

Table of Contents

1. Valuing Open Data Re-use

Ensuring a sustainable future for open data, and securing governments ongoing commitment to such work, requires evidence of the societal impact open data including (but not limited to) its economic impact. Failure to track the impact of opening data, which demands scarce government resources to prepare, host and maintain, may result in a lack of political will and financial investment in such projects ongoing. ‘Showcasing the economic impact of Open Data is a key element to rallying support for Open Data and to trigger publication and re-use in a country.’ (European Data Portal, 2018a, p. 58) In 2015, an influential report published by the Finnish Ministry of Finance noted that ‘[v]ery little is known about the underlying economic and organizational mechanisms and implications of open data use at the organizational level or at the level of economy as a whole.’[1]

There has been much work on this topic in the intervening years, but there still exists an open challenge in effectively tracking the causal links between the investment in open data programmes and the impact on the economy.[2] To meet this challenge there have been concerted efforts to capture both the macro and micro economic benefits of open government data, with a range of studies being undertaken across the EU (see for example the section ‘Economic Impact’ in the 2018 Open Data Maturity Report (European Data Portal, 2018b, pp. 58–61)).

Global benchmarking is also well-developed, (see the ODIN Open Data Watch Report (Open Data Watch, 2019), the Open Data Barometer (World Wide Web Foundation, 2019), the OECD OURdata Index (OECD, 2019) and GODI (Open Knowledge International, 2018)). Capturing the value chain of open data (rather than just measuring and ranking the state of open data) is required to provide the necessary evidence to policy makers and governments that there is real return on investment in both economic and societal terms to their open data portals. ‘From a government perspective, it is critical to sustain investments to open up government data with a sound business case (i.e. clearly defined value propositions), presenting the potential benefits of embarking on open data efforts.’ [3]

Tracking the path of open data from repository to re-use, to the impact of that re-use is often a broken and circuitous path with a range of actors. Portals are not the only place from which the data they promote can be downloaded; additionally, a number of intermediaries may play a role in finessing the data before the end user creates their product or service. These products and services are often a complex conflation of different data sets which may be both open and closed (i.e. shared or proprietary). Understanding exactly what role a particular data set may play in an application, and teasing out exactly what economic benefit can be attributed to its use is difficult, but can be captured through the use of appropriate metrics.

[1] Koski, 2015, p. 3
[2] Attard, Orlandi and Auer, 2016; Lämmerhirt and Brandusescu, 2019
[3] (OECD, 2018, p. 241)

2. Metrics for Open Data Re-use

A metric is a quantifiable measure that is used to track and assess the status of a specific business process, but it is also an observable characteristic of [X] that acts as a proxy for some other characteristic of interest/attribute which is less easy to observe. Metrics may not be able to capture the precise but can be useful indicators of the impact of a process, and can be used both in a positive manner, and as a counterfactual to illustrate what would *not* have happened, or been more difficult or expensive if open data were not available.

The re-use of open government data creates value for the public sector in delivering services and insights, where the challenges of tracking value are in applying the appropriate measures to data re-use. Secondary and tertiary re-use of data by the private sector also generates value, but provides particular challenges in tracking the uses to which the data has been put, and the economic impact of such re-use. Micro economic indicators have been developed to track value creation and impact from the perspective of government and of the private sector. Ongoing programmes like the annual Infomediary Sector Report produced by ASEDIE track the progress over time of companies who have a data-based business model, using quantitative business metrics to evidence economic value.

Metrics deployed by ASEDIE	
Subsectors of infomediary companies	Technical consulting, culture, directory service, economic and financial, publishing, market research, meteorological, geographic information, infomediation technology, tourism
Turnover	average, total, by subsector
Employee	total, by subsector, average turnover per employee, average expenditure per employee, average wager per employee
Share capital analysis	total, by subsector, average social capital
Profit and Loss	total, by subsector
Analysis of commercial risk	total, by subsector
Long term companies	sales evolution, employee evolution
Delisting	by motive (e.g. closure), community, subsector

Figure 1 Metrics for measuring value of infomediary companies in Spain. [4]

It is of course, not only private sector companies who can be expected to benefit from open government data – the public sector itself, and the citizens it serves, have frequently been identified as beneficiaries of economic benefits from innovation in open data re-use. Efficiency and productivity gains for public sector services directly benefit citizens as well, who additionally benefit from potential

[4] Adapted from (ASEDIE, 2018, p. 11)

time and cost savings from innovative services running over open data, as well as improved public services both in levels of service and in reduction in costs to the taxpayer.[5]

The value chain that links portal activity (views, downloads, user feedback, queries), to final use and impact, can be complex, with bare portal analytics not giving any sensible indication of the actual use to which the data might have been put, or its economic impact. Open Data portals are not an end in themselves, but act as facilitators in the discovery and use of (primarily government) data. The metrics applied to portals (as in the large national level studies above) are important indicators of how useful the data might be (quality, relevance, currency, ease of access, format etc.), but these measures do not capture how and to what effect the data is used down the pipeline. The aim of the open data agenda is to promote a wide variety of societal objectives including economic growth and the creation of new innovative products, services and apps. A failure to understand how this is happening and at what rate or volume will mean portals are liable to be unsustainable going forward as they lack a clear return on investment for public bodies.

Measuring the economic impact of open data falls into the classic problem set that plagues evaluation of any complex project; identifying the chain of causal links; devising appropriate tools and methodology for measurement, and having sufficient resources to carry out the evaluation. As Innovate UK note, '[r]igorous evaluation can be resource-intensive, so it is important to consider the proportionality of any activity.'[6] The literature suggests six criteria for 'good metrics' of open data (see table below.)[7] As well as valid and reliable, a desirable metric should be discriminatory, efficient, transferable and reliable. An ideal metric would rate highly on all of these criteria: an efficient assessment (e.g. automated) that could be quickly run against a large group of datasets with high validity and reliability giving results that are comparable for a wide range of contexts.

Metric Assessment	Description
Valid	The metric is closely correlated with the attribute of interest
Reliable	The metric gives consistent results over time and between observers
Discriminatory	The metric should be sensitive enough to discriminate between common values of the attribute
Efficient	The less time and resource required to use it the better. In some contexts, poor efficiency can lead to poor validity and reliability. If the aim is to measure a large number of datasets for a large variety of users (e.g. the Open Data Barometer) then poor efficiency may force the assessors to use a small convenience sample which potentially introduces both bias and sampling error.
Transferable	The same metric can be used in a variety of different contexts and across cultural and economic variation.

[5] Koski, 2015; OECD Digital Government Studies, 2018

[6] Innovate UK, 2018, p. 8

[7] eg Walker, Frank and Thompson (2015)

Comparable	If a metric is comparable not only is the metric transferable to a wide variety of contexts but the results can be meaningfully compared. Ideally this would result in a universal standard that transcends cultures and applications.

Figure 2 Data Metric Attribute Assessment Framework[8]

3. Data for Re-use Challenge

In the vast majority of open data use cases, the data is opened and used without a clear understanding of how the impact will be measured. This is most often done post hoc. The Re-use Challenge is designed to facilitate the implementation of measurement of impact from the beginning of the open data re-use process.

This project follows four mid-sized cities who are opening and reusing their own data to address issues in the city, across the Smart City sectors of environment and mobility. They are doing so by running seven pilot projects in partnership with SMEs. By following the process ethnographically from the inception to the end result, it is hoped to devise a range of impact attributes for each of the projects. By comparing the impact attributes at the end, we can create a suite of transferable attributes, with metrics or indicators for each, and methods for obtaining these metrics.

The cities range from having a substantial number of datasets open, to having only a few, to having none open at all. Each city defined two or three challenges that outlined problems the city was experiencing, and outlined currently and potentially open datasets with which the problems might be addressed. By means of an open competitive call, they invited small to medium enterprises to propose solutions. The most attractive solutions were selected for 6 month pilots in the cities.

4. Methodology

The metrics are derived from the original purpose of re-use in the following manner. The first three sections are completed from publicly available documents regarding the pilot project. The second three sections (overlapping in the Expected Impact field) are developed through interviews and discussions with the cities and the private companies developing the pilots. Semi-structured interviews were conducted with key personnel responsible for delivering the pilot projects in the city administration, with further sample interviews carried out with the companies delivering the solutions.

We selected the portals on the following criteria:

- The portals are early in their journey and have not yet developed metrics for measuring impact in depth;
- Replicability – there are hundreds of portals of similar sizes at a similar (or earlier) stage in their journey who will be able to immediately use similar metrics. The process itself can be additionally be used by larger and more mature portals;
- Accessibility to the full journey.

[8] User-Centred Methods for Assessing the Value of Open Data (Walker, Frank and Thompson, 2015)

This process works for portals that are early in their journey and have not yet considered measuring impact in depth; but also for more advanced portals that wish to establish criteria for opening data that takes into account possible impact, rather than simply availability or ease of opening.

5. Results

The results tables below provide a description of each case study, detailing the project purpose and expected outcomes and impacts, with particular attention to the economic impacts of the proposed smart city solutions. From the specific examples in the case studies, transferable metrics are derived that can be added to the portfolio of measurement tools to build evidence for the economic impact and benefits of open city data.

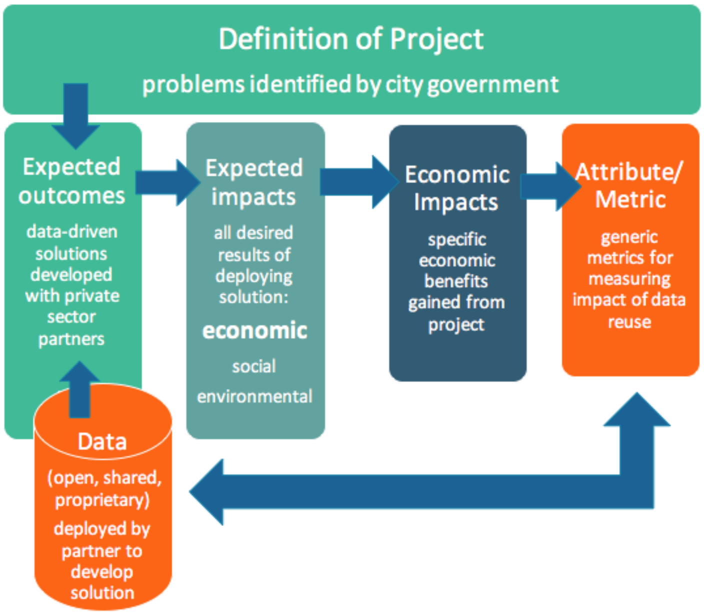

Figure 3 Information captured in Case Studies. This can also be used to set up impact tracking at the beginning of an open data reuse project.

5.1 City 1 - De-icing and Waste Management pilots

City 1 - De-icing Pilot	
Definition of Project	The city's maintenance department is responsible for de-icing slippery roads and snow removal in winter. This task is carried out following a designated protocol, which is predefined and static, and prioritises larger roads over smaller, regardless of particular traffic routes (especially those used by bicycles). However, by implementing an approach that allows for more flexibility, the city can provide more customization in order to deliver a higher service level to citizens. The city is looking for a solution that enables the maintenance department to customize de-icing routes and planning, but in the future this may also be extended to other route prioritisation challenges, such as waste

	collection.
Datasets	Major road intensities http://www.ndw.nu/pagina/nl/4/databank/31/actuele_verkeersgegevens/ Traffic accidents http://www.ndw.nu/pagina/nl/4/databank/31/actuele_verkeersgegevens/ Biking accidents https://fietsongevallen.oververkeer.nl/ Heightmap http://pdokviewer.pdok.nl/ Regional biking network http://pdokviewer.pdok.nl/ Local road intensities Priority bikelanes Weather data Salt spreading routes Major road traffic densities Traffic accidents Cycling accidents Regional cycle network Heightmap Local road traffic densities Priority cycle lanes
Expected outcomes	Tools that provide smart solutions for treatment prioritisation Crowdsourcing apps that encourage user reports Improved dataset combinations for prioritization
Expected impacts	Reduction in citizen complaints Improved citizen perception of city services Reduction in ice-related accidents Improved road conditions Improved service delivery Cost reductions for transport, staff time and resources
Operationalisation of impacts (ie, how will costs be reduced?)	Predictive analytics to anticipate conditions and inform smart routing De-icing measures delivered only to necessary sites via the most efficient route Reduction in staff time, vehicle use, energy and salt/grit use Reduction in accident related costs (emergency services, medical treatment, insurance)
Economic Impact	**Labour costs** – efficiency gains in de-icing teams **Resources** – mileage, fuel savings, targeted use of grit/salt for de-icing trucks **Road traffic and accident incidents** – reduction in treatment costs, insurance claims, emergency services call outs, days lost from work
Attribute/Metric	**Staff costs/work hours** **Mileage** **Fuel Consumption** **Grit/salt consumption** **Citizen complaint levels** **Road traffic accident statistics** **Emergency call out statistics** **Hospital admission/treatment records**

Figure 4 Results - City 1 De-icing Pilot

City 1 – Public Litter Bin Pilot	
Definition of Project	Public litter bins, ashtrays and dog waste bins provide an important function in keeping cities clean, hygienic and attractive. Unavailable or over-filled bins mean that citizens are inconvenienced, and also encourages littering. This particular aspect is highly visible to citizens, and is frequently a priority issue for them. Consequently this is an area of investment for cities. However, current collection timings are based on routine and experience. This leads to wasted time and money as under-filled bins are collected, or over-flowing bins where an event or other extra-ordinary occurrence has taken place. Therefore, there is a requirement for a solution that enables customized and optimised route scheduling, that allows city staff to decide when and where it is necessary to intervene.
Datasets	Events calendar Location of green spaces Location of bodies of water Street furniture Yet to be opened User reports Inspection reports on public spaces Garbage hotspots Number of people in city by date
Expected outcomes	Develop tools that will help to optimize and reduce the tours required for the maintenance of existing cleanliness facilities, including monitoring levels, identifying litter spots (particular dog waste bags); Provide a solution that helps survey which cleanliness facilities need a maintenance in order to schedule collection operations and optimize daily tours. Find an innovative way to raise awareness of citizens concerning street cleanliness or incentivize them to use waste facilities. Linked with the citizen reporting system Ongoing determination of whether the public bins are correctly placed and if there are a sufficient number Replicability, interoperability and transferability of the solution
Expected impacts	Better internal efficiency in the maintenance of cleanliness facilities Reduction in time spent by the relevant department Litter facilities are always functional and ready to use Raised public awareness concerning the issue of street cleanliness
Operationalisation of impacts (ie, how will costs be reduced?)	**Labour/Productivity** – remove need for visual checks of environment (city), smart routing, targeted work flows (contractor) **Resources** – mileage, fuel savings **Procurement** – better data for contract negotiation for waste services
Attribute/Metric	**Staff costs/work hours** **Mileage** **Fuel Consumption** **Citizen complaints** **Waste contract pricing**

Figure 5 Results - City 1 Public Litter Bin Pilot

5.2 City 2 and City 3 – Safer Cycling and Air Quality pilots

City 2 and City 3 – Cycling Pilot	
Definition of Project	The aim is to have more people biking in/to the city, with a strong focus on children and their safety, particularly on journeys between home and school. The cities also want to encourage cycling for all in the cities: inhabitants, schools, employees, workers, visitors and parents as well as children. Current issues that prevent the uptake of cycling in the cities include the fact that people still perceive car transport as being faster and easier. Working parents in particular find their car the best solution when combining commuting with journeys to and from childcare and school, and the car is seen as the best option when shopping. Multi-modal switching is not always possible. The urban bicycle infrastructure is far from ideal, cities lack integrated information on the level of safety in specific (school) areas, and there is also a difference between perceived safety and objective, infrastructural safety. Therefore, the aims are: How to get more people biking in our city; Ways to make biking to school and work safe and convenient; Data-driven methods for behaviour change and education; Safe infrastructure (in school environment); Ways to get information about quality of roads and environment; Ways to improve traffic safety (with the focus on bicycles); Safety gear that is attractive, popular and fun.
Datasets	City 3: Road Register: information on road infrastructure. City 3: Intersection Register: information on road intersections. City 3: Roadworks that affect traffic. City 3: Traffic measures: information on the amount of motorised and non-motorised traffic on specific roads. City 3: Demographic measures. City 3: Geolocation of popular citizen city destinations. City 3: Traffic accident information. City 2: Road infrastructure (street width, gradient, …) City 2: District borders City 2: Location popular destinations City 2: Number of school-age children City 2: Bike paths City 2: Cyclists by counting loops City 2: Location schools City 2: Number of accident reporting (police) City 2: Route2school (helps schools and municipalities by thoroughly analyze road safety on school routes and collect information about the travel behaviour of pupils) City 2: Bike counts
Expected outcomes	Apps and services delivering: New forms of hardware; (e.g. bike infrastructure in public domain) Tools that persuade/help people access cycling for all purposes; (e.g. gamification tools, apps, …)

	Tools to assist cities screen/evaluate safety and traffic circumstances; Smart tools, signage; Smart route planners; Tools to identify conflict areas and reduce conflicting priorities; Enabling conditions (bike sheds, bike lanes etc); Educational tools.
Expected impacts	An increase in cycling in the relevant city, especially amongst the target groups of school children; Reduction in accidents especially on school journeys; An increase in the perceived safety of cycling amongst citizens; Fewer cars in city centres An increase in the number of children who know how to cycle safely and choose to cycle.
Operationalisation of impacts (ie, how will costs be reduced?)	**Road traffic and accident incidents** – reduction in treatment costs, insurance claims, emergency services call outs, days lost from work/school **Health and wellbeing** – increased activity levels, reduction in poor air quality impacts on range of health indicators e.g. respiratory disease, asthma, cardio-vascular disease, childhood obesity T**raffic congestion** – better traffic flow for city, reduction in pollutants
Attribute/Metric	**Road traffic accident statistics** **Traffic flow/congestion measures** **Emergency call out statistics** **Hospital admission/treatment records** **Health statistics** **Cycling app statistics (journey length, route, frequency)** **Improved air quality**

Figure 6 Results – City 2 and City 3 Cycling Pilot

City 2 and City 3 – Air Quality Pilot	
Definition of Project	Air quality has a serious impact many aspects of city life. These include the health of citizens, who may suffer lung problems or heart disease and the environment, where poor air quality may cause smog and acid rain. The challenge is even stronger in highly urbanised regions in the area and historic city centres face their own particular issues regarding narrow 'canyonised' streets and the degradation of historic buildings. Currently, it is difficult for cities to have an integrated evidence-based view on air quality and link these to specific circumstances. These may be general, such as time, weather, traffic and events, or particular to the built environment issues such as specific building materials or street layout, where pollutants become trapped by air recirculation between narrow buildings. Once these are identified, how can local authorities have an impact on these to improve the air-quality for citizens?
Datasets	City 3: Road Register: information on road infrastructure. City 3: Intersection Register: information on road intersections. City 3: Traffic measures: information on the amount of motorised and non-motorised traffic on specific roads. City 3: Particle measures: information from a mobile installation. City 3: Particle measures: information from a fixed installation. City 3: Demographic measures.

	City 3: VMM, Flemish environmental society City 2: Road infrastructure (street width, gradient, …) City 2: Public infrastructure inventory City 2: VITO MeetMeeCity2 (independent Flemish research organisation in the area of cleantech and sustainable development) City 2: District borders City 2: Location of forests City 2: Curieuze neuzen (big citizen scientific research on air quality, specifically NO2) City 2: Pollutant Release and Transfer Register (PRTR reporting)
Expected outcomes	Integration of air-quality measures, e.g. temporal measures, real-time measures, other open data sources; different measure like NOx, particle measures; from different organisations, suppliers or sensors. Scenario tool for air-quality improvement Gauging the impact of specific measures; What-if scenarios. Real-life feedback integration Closing the loop between using the scenario tool, executing the advised scenario in practice, and measuring the eventual impact on air-quality; Urban planning tool Where to locate specific sites (schools, factories) to minimise the air-quality impact.
Expected impacts	Root causes for air-quality degradation have been identified; Remedial actions have been identified and have been shown to work; In the long run, air quality measures are reduced below World Health Organisation limits on a daily basis.
Operationalisation of impacts (ie, how will costs be reduced?)	**Policy and intervention** - identification of effective pollution abatement measures **Health and wellbeing** – reduction in poor air quality impacts on range of health indicators e.g. respiratory disease, asthma, cardio-vascular disease
Attribute/Metric	**Traffic flow/congestion measures** **Emergency call out statistics** **Hospital admission/treatment records** **Health statistics**

Figure 7 Results – City 1 and City 2 Air Quality Pilot

5.3 City 4 – Green Space Watering pilot

City 4 – Green Space Watering Pilot	
Definition of Project	The digital roadmap of City 4 retains as a top priority the development of sustainable cities by acting on two levers: optimization of energy consumption and improvement of the (internal) efficiency of public service. City 4 has a focus on encouraging the development of autonomous solutions in public services. In this framework, the city wants to deploy an intelligent system in green spaces that can control watering in order to optimize water consumption and make the system completely autonomous. The main difficulties lie in the ability to develop an autonomous solution to control the triggering and stopping of watering heads and to produce this solution at a reasonable cost for a middle-sized city. In particular, the lack of data on the

	prevailing conditions in green spaces limits the capability of the city to optimize water consumption and realize energy savings. Currently, the only data used to monitor watering heads is a day-night cycle. Further, municipal staff have an increased workload since the introduction of a new national legislation that forbids the use of chemical products to manage green spaces. Therefore, this is also question of internal efficiency in order to limit the time of the city's staff on the field with the support of a full autonomous system. City 4 has already explored some existing solutions on the market, but they are not aimed at the medium-sized city market.
Datasets	Events calendar related to occupation of public green spaces; Location of green spaces; Weather datasets from weather station owned by of the local government or from collaborative weather network if needed to gather forecast weather; Prevailing environmental conditions of public green spaces through sensors (air humidity; ground humidity)
Expected outcomes	Optimise water consumption in green spaces by using or / and combining different parameters: weather data and forecast; water needs in function of green space's profile / typology; Monitoring tools for city staff such as a dashboard that gives an overview of historical and present conditions of green spaces; Solutions linked with local weather stations; Remote and autonomous controls for watering systems.
Expected impacts	Energy savings by optimizing water consumption Reduction in water and labour costs Replicability, interoperability and transferability of the solution
Operationalisation of impacts (ie, how will costs be reduced?)	**Resources** – mileage, fuel savings, targeted use of water **Labour** – reduction in visual checks, site team visits, running watering systems **Productivity** – automated work flow **Monitor service delivery** – set targets
Attribute/Metric	**Staff costs/work hours** **Mileage** **Fuel Consumption** **Water consumption**

Figure 8 Results – City 4 Green space Watering Pilot

5.4 Summary of Applicable Metrics

Figure 10 below summarises the evaluation metrics suggested by participants to track and evidence the impact of the pilot projects in the cities.

Economic impact	Description	City activity	Metric	Data source to be used
Labour	Reduction in City staff time spent on monitoring activities	C1 – visual checks for waste C4 – monitoring green spaces	Staff costs/work hours	City workforce data
Labour **Productivity**	Increased efficiency in routing	C1 – waste contractor C1 – de-icing teams	Staff costs/work hours	Waste management contract data

				City workforce data
Labour Productivity	Automated work process	C4 – watering	Staff costs/work hours	Workforce data – will compare pilot site with comparable managed without sensors
Resources	Transport – reduction in trips for staff	C1 – visual checks C1 – refuse trucks C4 – watering team	Mileage Fuel consumption	Vehicle/plant records
Resources	Water	C4 – rationalise water use	Water volume	Water meter data
Road traffic accidents	Reduction in costs to emergency services, hospitals	C3 and C2 – safe cycling C1 – effective de-icing	RTA statistics	Local Road Traffic Accident Statistics Treatment costs for RTAs Emergency services call out costs
Congestion	Reduction in school run journeys by car	C3 and C2 – safe cycling	Transport type Journey type Congestion levels	Annual schools transport survey App data tracking journeys Road traffic reports
Health and Wellbeing	Increased activity levels	C3 and C2 – safe cycling	No. of children cycling to school Distance/mins of exercise (app)	Users activity on app (e.g. mins activity per day, distance travelled, transport type) School Transport Surveys Health outcomes – local data e.g. obesity Health impacts of exercise – can infer cost savings in medical treatment
Health and Wellbeing	Reduction in exposure to poor air quality	C3 and C2 – healthier cycling routes Air quality monitoring - open AQ data for citizens	Local Health and Wellbeing outcomes Route tracking (app) Air quality Working hours lost to sickness	Health statistics – e.g. Respiratory disease, asthma, cardio-vascular disease, children's fitness/obesity Distance cycled (activity levels) (app) combined with Air quality data to infer exposure to pollutants

				Can infer cost saving in health care/reduction of lost hours at work
Procurement	Data insight used to negotiate contracts	C1 – waste contractor	Waste contract pricing No/type of citizen reports	Insights from smart waste bins to measure cleanliness and collection process used to leverage contract renewal negotiations Citizen reporting/complaints
Contract management	Data insight used to check fulfilment of contract duties	C1 – waste contractor	Street hygiene levels No/type of citizen reports	Insights from smart waste bins for cleanliness and collection process used to monitor contract outputs and costs Citizen report/complaints
Monitoring service delivery	Data insights used to improve service delivery	C1 – de-icing C4– automated watering	Improved quality of green spaces No. of citizen complaints (ice)	Management reporting data City records of citizen complaints/feedback
Internal data management - efficiency	Using open data as catalyst for improving data processes	All	No. of data sets opened Data quality Data collection/ update frequency	Portal analytics Portal user feedback
Compliance	Open by default legislation	All	No. of data sets opened	Portal analytics

Figure 9 Summary of Metrics and Measurement for Cities

6. Discussion

6.1 Transferable Metrics

The evidence gathered from the City case studies presented in Figure 11 below, extracts generalizable metrics categorised according to the accessibility project owners and developers have to the data for measurement – **operations data** (in-house), **related public services data** (held by other public services) and **private sector data** (financial and business data).

The operations data is held in-house by the cities as it is derived from their own management and financial records. This data is readily applicable across organisations and domains as it is routinely collected for the purposes of financial and budget reporting, and for monitoring and delivering services. Of course, criteria like reliability are dependent on the robustness of an organisations own data collection and management processes. Surfacing this data via a dashboard would have a range of benefits internally for the organisation, and allow analysis and visualisation of data for a variety of purposes, including specific project evaluations. Aggregated data could also be surfaced for a citizen-facing dashboard to enable transparency in budgets and spending, and business intelligence insights.

Drawing on related public services data presents more of a resource challenge for evaluation purposes as it requires more time, effort and what could be complex interpretation and analysis of data sources, and building models for inferring impacts of open data reuse projects. This is a barrier to a data owner engaging in more complex project evaluation due to restrictions on resource.

City governments themselves are unlikely to engage in local assessments of the data market place and SMEs, but this kind of analysis is of interest at a regional and national level, with countries such as Spain having business associations like ASEDIE, the Spanish Multisectorial Information Association. They conduct longitudinal analyses of the information and data market, and the outputs of trade and governmental economic assessments can be useful to local assessments of the impact of open data in a local market place.

Economic Impact	Attribute/Metric	Project Evaluation Did the data-driven solution work? How can we tell?
Operations Data - collected in-house by government		
Labour costs/ productivity	**Wage costs** **No. of hours worked**	1. Management data from service delivery teams to track impact on staff hours 2. Workflow efficiencies (e.g. smart routing)
Service delivery	**Service level outcomes** **Service delivery costs**	1. Management reporting data 2. IoT data – e.g. smart bins 3. Citizen complaint levels – data gathered from citizen app, website, email and telephone to City re: services 4. Citizen reporting - data gathered from citizen feedback app, civic website, email and telephone re: conditions (e.g. weather, road conditions, cleanliness)
Resources	**Mileage** **Fuel Consumption** **Grit/salt consumption** **Water consumption**	1. Management data from service delivery teams to track impact on vehicle use (mileage, hours of running, repairs) 2. Purchase records - fuel, salt, grit 3. Water meter readings
Procurement	**Contract pricing**	1. Management data – service levels and costs
Traffic congestion	**Traffic monitoring** **Journey times**	1. CCTV and traffic light data 2. Travel app data 3. Citizen reporting
Related Public Services Data		
Road traffic accidents	**Frequency of accidents** **Emergency call outs** **Injury statistics**	1. Road traffic accidents statistics 2. Emergency services records 3. Hospital admissions/treatment records 4. Insurance claims

Health	**Health statistics** **Air quality** **Exercise levels**	1. Hospital admission/medical treatment records for respiratory disease, asthma, cardio-vascular disease, children's fitness/obesity 2. Air quality monitoring data from national data collection and local monitoring 3. Travel app statistics (journey length, route, frequency)
Private Sector Data		
Data services market place	**SME no.** **SME turnover** **SME profit/loss** **SME sustainability** **SME employment** **Data Products** **Data Services**	1. Industry surveys (data services sector) 2. Financial reporting data 3. Investment data 4. Sector employment figures **5.** Market survey data (products and services)

Figure 10 Transferable metrics for measuring impact

6.2 Open Data Re-use Impacts

Populating and managing open data portals at the municipal level presents challenges to Cities, as do implementing innovative data-driven projects. Core challenges are those of their own internal data management processes, which are often scattered and siloed across different departments, with little if no interoperability between systems collecting or storing data. Open data projects encourage cross-departmental working, as reported by the cities in our case studies, and open data encourages the breaking down of siloes within local government. This enables more holistic and innovative solutions to be initiated to solve the 'wicked problems' cities face. Open data also supports the data services and products market place, as noted by our SME respondents, who are able to provide exciting new products and services, alongside building their businesses and employee numbers. National level data is equally valuable (if not more, in some cases) for this purpose.

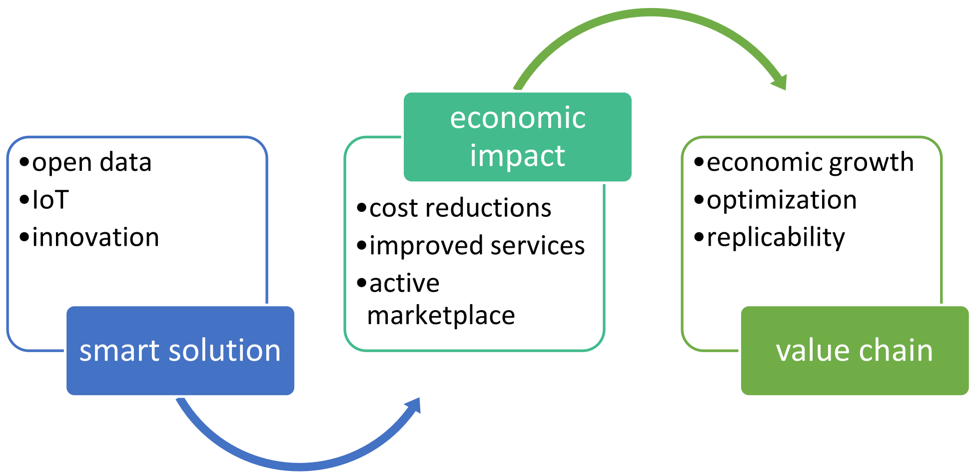

Figure 11 Relationship of elements in open data value chain

Innovation in services in one city presents the opportunity for replication by other cities, for example when the small Italian city of Merano was able to create a transport app based on the use of open data in such an app in the much larger city of Helsinki[9]; this reduces the costs (and time to market) of innovation in the longer term. Evidence can be provided for the value chain illustrated in Figure 12 above, by ensuring that appropriate metrics and evaluation techniques are considered for open data project, and that these are made part of the project planning at the outset of projects to enable impacts to be assessed. These serve the interests of the procurers of these products and services (cities), and the service providers and product developers who utilise these to provide evidence to new customers of their effectiveness. The wider impact this has on the economy and the improvement in public services and data insights can then support further developments and investment. Finally, establishing business cases and a clear value chain to impact is key to convincing private companies of the benefits of sharing their data (whether openly or in a more closed fashion) with cities[10], and the above methodology is of use to national, regional and municipal authorities who wish to facilitate this.

7. Recommendations

There are 4 key recommendations for use of this method:

- The Impact>Attribute>Evaluation method can be applied to any public sector and/or smart city open data project, large or small;

- This is not a final list of indicators and metrics. By applying this method to other data re-use projects the list of possible indicators can be extended;
- New metrics and indicators should be published and shared, to create the opportunity for replication by other portals;

- Building the assessment of these metrics into the original project plan reduces the cost of collecting them.

[9] Digitransit & bus.meran.eu
[10] Ryan and Gregory, 2019

8. References

ASEDIE (2018) ‘Vi Infomediary sector Report | datos.gob.es’. Available at: https://datos.gob.es/en/noticia/vi-infomediary-sector-report-asedie.

Attard, J., Orlandi, F. and Auer, S. (2016) ‘Value Creation on Open Government Data’, in *2016 49th Hawaii International Conference on System Sciences (HICSS)*. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI: IEEE, pp. 2605–2614. doi: 10.1109/HICSS.2016.326.

European Data Portal (2018a) *Creating Value through Open Data*, *European Data Portal*. Available at: /en/highlights/creating-value-through-open-data (Accessed: 27 February 2019).

European Data Portal (2018b) *Open Data Maturity in Europe 2018*, *European Data Portal*. Available at: /en/news/open-data-maturity-europe-2018 (Accessed: 4 March 2019).

Innovate UK (2018) *Evaluation framework*, *GOV.UK*. Available at: https://www.gov.uk/government/publications/evaluation-framework (Accessed: 14 March 2019).

Koski, H. (2015) ‘The impact of open data – a preliminary study’. *The Research Institute of the Finnish Economy*. Available at: https://www.w3.org/2013/share-psi/wiki/images/6/67/Impact_of_Open_Data_in_the_Public_Sector_Koski_2015.pdf (Accessed: 27 January 2019).

Lämmerhirt, D. and Brandusescu, A. (2019) *Issues in Open Data - Measurement*. Available at: https://www.stateofopendata.od4d.net/chapters/issues/measurement.html (Accessed: 11 June 2019).

OECD (2018) *OECD Digital Government Studies Open Government Data Report Enhancing Policy Maturity for Sustainable Impact: Enhancing Policy Maturity for Sustainable Impact*. OECD Publishing.

OECD (2019) *OURdata Index*. Available at: http://www.oecd.org/gov/digital-government/open-government-data.htm (Accessed: 29 March 2019).

OECD Digital Government Studies (2018) *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*. Available at: https://www.oecd-ilibrary.org/governance/open-government-data-report_9789264305847-en (Accessed: 28 March 2019).

Open Data Watch (2019) *ODIN World Map - Open Data Inventory*. Available at: https://odin.opendatawatch.com/ (Accessed: 29 March 2019).

Open Knowledge International (2018) *Global Open Data Index*. Available at: https://index.okfn.org/ (Accessed: 29 March 2019).

Ryan, M. and Gregory, A. (2019). Ethics of Using Smart City AI and Big Data: The Case of Four Large European Cities. *ORBIT Journal, 2 (2)*. Available at: https://doi.org/10.29297/orbit.v2i2.110 (Accessed: 18 September 2019)

Walker, J., Frank, M. and Thompson, N. (2015) ‘User centred methods for measuring the value of open data’, in. *Open Data Research Symposium 2015 (27/05/15)*. Available at: https://eprints.soton.ac.uk/375700/ (Accessed: 9 July 2019).

World Wide Web Foundation (2019) *Open Data Barometer*. Available at: https://opendatabarometer.org/?_year=2017&indicator=ODB (Accessed: 9 July 2019).

Annex 1 – Interview Questions

City Government

Interview Purpose: The what (is supposed to change) and how (are you going to measure it)?

Introduction

Introducing the research project purpose and interviewer.

1. Providing the interviewees the opportunity to introduce themselves, their job roles.

Description of project

1. What challenges are you trying to solve in your city/for the city? (problem space)
2. What solutions are you working on with your partner organisation? (describe)
3. How is data being deployed in these solutions?

Measuring impact and outcomes

1. What do you hope to achieve by using the data? (overall expectations/outcomes for the project)
2. How are you planning to measure the success of the project? Be specific – see Q3.
3. What specifically will you measure/record/assess in order to evidence the impact of the project?
4. What resource (reporting structures/personnel/finance) do you have/intend to put in place to ensure recording and communication of impacts?

Closing discussion

1. Any further thoughts or comments on the opening of data in your city and/or the reuse of open data not covered in our discussion above?