D2.3.2 Guide to Valuing Data

This is an additional section of the open data guidance package (output 2 - D1.4.1) which provides methods for assessing the value of opened data sets based on the Accelerator cases, and provides strategies for public authority decision-making regarding open data and smart services.

1. Introduction

D1.4.1 states that, “Cities investing in opening up data aim at creating added value out of it.” It identifies 5 specific types of value that can be created. To these 5, we have also added another kind of value: citizen participation.

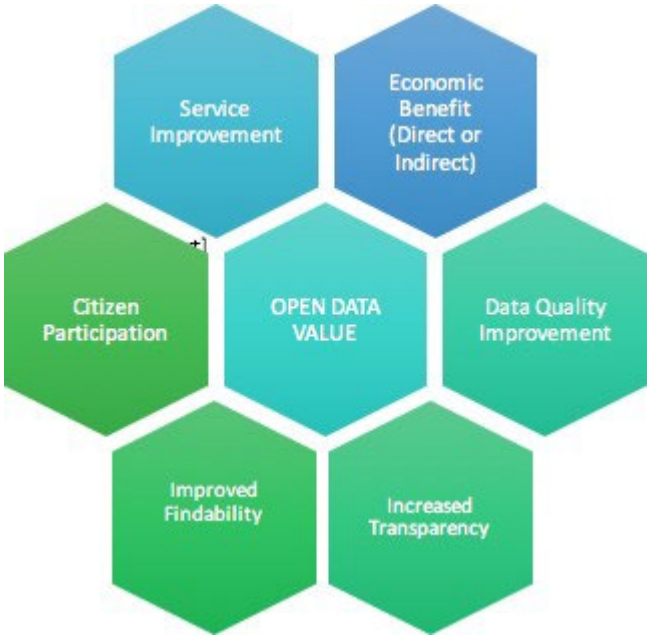

Figure 1: Six Types of Value Created by Opening Data

This deliverable describes how data can be valued for each of these types of value, and in particular explores methods for assessing service improvement and economic value.

2. Data quality improvement

In the SCIFI project almost all cities have stated that poor quality of data is one of the reasons one might not decide to release a dataset as open data (for the project). It has the risk of third parties drawing wrong conclusions based on poor datasets, or wrong usage leading to possible political risks and reputation damage or other negative effects that influence the degree of success for the SCIFI solutions. However, releasing low quality data “*could help identify the dimensions on which the quality of the data is poor, so that governmental data providers can improve these dimensions.*” Higher quality data is more

valuable because it reduces the potential for error when it is utilised, but the required quality of data depends on the purpose for which it is used.

Quality incorporates several criteria. There is a considerable body of research already published on what these constitute. Some key characteristics are explained below.

- Accuracy reflects how well the data describes the situation
- Timeliness indicates how up to date the data is
- Reliability is an expression of how consistent the data is likely to be over time - is it coming from the same source, in the same way, or does it frequently change?
- Completeness is a measure of the data in a dataset versus what should ideally be in there

However, not all lists of data quality characteristics are consistent. Others include auditability, uniqueness, relevance and validity (whether they are collected and stored in a manner that is compliant with any regulation or legal requirement). The characteristics very much depend on why the data is being gathered and what it is being used for. Even when a characteristic is frequently cited - such as 'timeliness' - this can mean many things. Timeliness in a bus arrival app is down to the minute or second; timeliness in other types of data, such as demographic data, can be effective over a much longer timeline.

Greater automation in the collection of data tends to mean higher quality data for several reasons. One is that there is less potential for error to be introduced, but another is that they are more likely to be an exact fit with the data that is required, unlike 'exhaust' data sets (produced as an output of another activity) which often require a 'best fit' approach.

2.1 The SCIFI Experience

Because of this variation in what constitutes quality, the SCIFI project has shown that such pilots can indeed be useful for improving data quality based on use. In the Delft de-icing pilot, the startup discovered missing cells in one of the geographical datasets opened up. Delft went back to the start of the process from gathering to publication to establish why the cells were missing and how they could improve the data. In Delft's waste pilot they were able to establish that a crucial piece of data - the size of the waste bins installed in public locations around the city - was not held.

2.2 Value Assessment Recommendation

A good place to start is with the 5 dimensions of the Quality Assurance Framework of the European Statistical System. Assessing datasets against these 5 criteria can help public authorities think about how they collect, store and use their data as well as valuing it.

3. Improved findability

"*To find the right data within the organisation itself is already a challenge,*" one city stated during one of the first gatherings of the partners in the SCIFI project. If cities strive to publish open data, they need to be able to identify datasets that might be published as open data. Despite the advent of Google Search, a central data platform or portal is still key to increasing dataset discoverability. Traditionally, portals have been seen as an externally facing tool that focuses on publishing.

However, more recent work has focused on moving the value proposition to portals as enabling a community of data users, including public authorities, citizens, developers and other technology businesses and experts[1]. There are many ways to engage staff in using a central portal both for publishing and finding data. Implementing a request system can help public authority staff to understand what data could potentially be available to them.

3.1 The SCIFI Experience

In the de-icing pilot, Delft was able to establish that key data was not being stored in the most reliable and accessible manner - in other words, it was available to only a small number of people in a single department. St Quentin found that although the data might be theoretically available, finding where it was located, or the correct version, was challenging. With data sets being newly published to a variety of standards, they also experienced problems harmonising their data.

Delft found itself at the other end of the availability issue, when the city and SME teams identified a potential useful dataset from the Dutch passenger railway operator, but were unable to obtain it as it had yet to be opened.

3.2 Value Assessment Recommendation

Internally, value can be measured by an increased number of departments accessing the data/accessing it through the platform instead of other channels, or other metrics such as departments accessing data that is not created by/managed by their department. However, this requires that the central data platform is well-promoted throughout the organisation.

4. Increasing transparency

Outstandingly relevant are the potentials and opportunities of additional transparency in government. "*Organisations increase transparency when they expect valuable external*

[1] https://european-data-portal.gitlab.io/future-open-data-portals/

influences and are interested in a more intensive interlinking with their surroundings, without the risk of getting damaged."[2] Open data offers free access to government data and information to all and thus it increases the transparency of cities.

The key facilitator of increasing transparency is 'open by default', which is the number 1 principle of the Open Data Charter,[3] created in 2015 and signed by over 100 governments to date. This states that governments must justify the data they are keeping closed, for security or data protection reasons.

However, there are occasions on which simply being transparent, without an associated policy, can be less than helpful to citizens. Such an area is air quality. Releasing data on air quality in a city or region will inevitably attract (important) questions on how the relevant authority is dealing with air quality issues. Thus, the city must be ready to engage with developing an informed strategy at the same time as releasing the raw data. This will help citizens engage with the strategy.

4.1 The SCIFI Experience

The cities of Mechelen and Bruges collaborated on an air quality challenge. They defined part of the challenge as the necessity to close the loop between measuring air quality levels, implementing a scenario tool, executing the advised scenario in practice, and finally, measuring the air quality again to identify eventual impact. In this way, data was not 'released into the wild' without the city being able to respond.

4.2 Value Assessment Recommendation

The Open Data Charter has recognised that *"opening up data in isolation is less effective than it can be if targeted at solving specific policy problems."* They call this "Publishing with Purpose." Cities aiming to increase transparency should prioritise opening the datasets that will help them develop robust strategies and policies alongside other stakeholders. Value here can be assessed based on the use of datasets to inform policy.

5. Increasing citizen participation

At the heart of the value of increased citizen participation lies two truths: the first is that data is not valuable until it is used, and the second is that citizen engagement can create additional value and opportunities for all.

[2] Geiger, C. & Lucke, J. von (2012) Open Government and (Linked) (Open) (Government) (Data). JeDEM 4(2): 265-278.
[3] https://opendatacharter.net/principles/

Social initiatives often fill important vacuums that the private market does not have the incentive to fill. These might be run by individuals, volunteer groups or campaigning groups, but essentially amplify the range of solutions available to address city challenges.

'The Bristol Approach' is a framework for ensuring technologies and programmes for smart cities align with the needs and priorities of the people who use them. The project found that by enabling policy contributions from citizens, engaged citizens were satisfied that they could have impact, and discontented citizens tended to have a more positive attitude.

Lastly, while citizens are the source of much data for the city, often this cannot be used due to privacy concerns. Some kinds of data can only be used if citizens specifically and knowingly consent, and other types can only be collected by crowdsourcing the data or its being volunteered by citizens. The first implementation of the Bristol model was a project where citizens collected, shared and used data to address the problem of humidity in rented homes.

5.1 The SCIFI Experience

St Quentin ran a hugely successful pilot which looked at watering their green spaces in a more effective way, both financially and environmentally. But this challenge wasn't on their initial list to focus on. It was only when they went back to present their long list of challenges to groups of citizens that they discovered more effective watering of public parks and playing fields was high up on the list of citizen concerns. By running this pilot using their data on weather and usage, the city was able to meet the needs of citizens.

Bruges and Mechelen both engaged with their citizens to provide data to their cycling pilots. It's important to note that most of this data was not appropriate as personal data. Therefore the only way the city was able to obtain and use this data, and remain GDPR compliant, was by engaging with citizens to gain consent.

When Mechelen's air quality pilot did not achieve what was hoped due to technical problems with sensors and data collection, the city was able to pivot towards working with a citizen group which used less technically advanced (and therefore easier to install and manage) sensors.

5.2 Value Assessment Recommendation

There are a number of ways to identify this value, both in terms of numbers of citizens engaged and initiatives created but also in reduction of total numbers of complaints, reduced controversy regarding policy implementation and so forth.

The next two sources of value are reviewed in conjunction with each other.

6. Service improvement

The reuse of open data such as that in SCIFI is intended to lead to improved services, via the development of new solutions or tools by third parties (either citizen groups or the private market). These tools may improve the way a city provides a service, or enhance the service itself. Open data might also cause gathering of new insights by the city itself or third parties that may influence and improve the service the city delivers.

7. Economic value

Economic value might be:

- companies that establish in the city, employ staff and pay rates and taxes
- companies developing new products or services with open data that benefit business or consumers
- Cost reductions created by (directly or indirectly) by better or more efficient services for the city

Often economic measures are down the line and very much based on the commercial market mechanisms, such as increased employment, and increased tax revenues. However, these are a. Extremely long term, b. Can be hard to isolate and c. not applicable where the supplier is based in another country. Therefore, it is crucial to develop a way to identify more localised and short term metrics of value.

8. Economic value through service improvement

Data can only be valued by what is done with it (or what can potentially be done with it)[4]. Therefore it is hard to value prospectively, generically and in isolation. The key thing is to understand what can be done with data and how that can be measured

Microeconomic studies have similar econometric methodologies to macroeconomic studies, but focus on specific publishers and data. An example is the report 'Assessing the Value of TfL's Open Data and Digital Partnership[5]'. This identified direct benefits, realised in the form of revenues from market transactions and indirect benefits of positive externalities, for example, increased engagement with municipality and services. These are very reliable and comparative metrics. However, the key challenge with microeconomic metrics is that they usually are seen as only applicable to private reusers, and are generally limited to metrics

[4] https://www.bennettinstitute.cam.ac.uk/publications/value-data-summary-report/
[5] http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf

such as sales, turnover, profit, jobs, and so on. However, there is a vast range of potential indicators that can be developed.

Below is a process for identifying economic impacts based on the use of data for the service delivery pilots in SCIFI. This begins with identifying a problem, or challenge, that the public authority needs to solve. The authority then describes the expected outcomes of solving the problem, and the impact of these outcomes. These impacts can then be described economically - for example, 'fewer cars in the city centre' might be beneficial by increasing productivity through decreased journey times; reducing spending on road repairs; or losing fewer days of work or school because of increased air quality. These impacts can then be measured as metrics. The public authority may already have quantified their desire to reduce the number of cars in the city centre as part of their strategy, and this metric can be used.

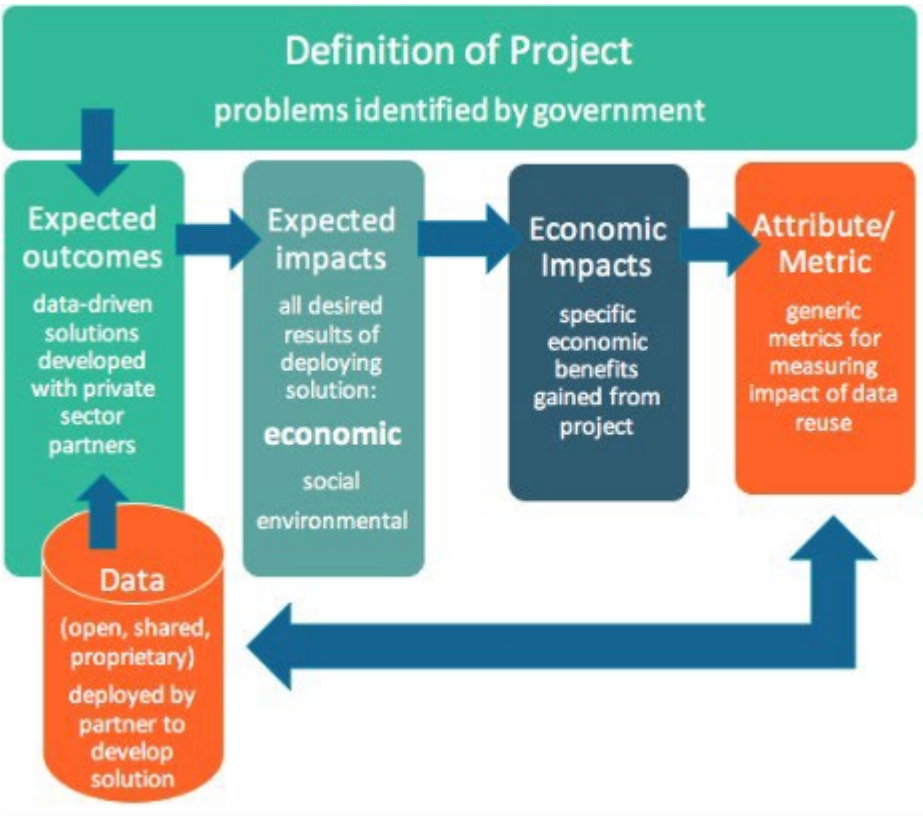

Fig 2: Outline of the process for valuing economic and service improvement data

When thinking about the value of data, it is also important to recognise that different stakeholders may be able to access different types of value. It is important to note that the benefits that accrue to technology suppliers may well differ to those that accrue to public authorities. For instance, these might include more obvious measures such as increased turnover or employee count, but also having a reference implementation to

enable them to recruit other customers, or having a testbed (and test data) to create new, valuable features for existing products.

Below is a table showing how value can be assessed based on the impacts of the products and services created with data. Some of these are direct economic measures (eg wage costs) and some are indirect, such as reduced journey times. Many of these examples are transferable to other contexts.These metrics can be added to the portfolio of measurement tools to build evidence for the economic impact and benefits of open city data.

Economic Impact Domain	Value Metric	Data Source
City Operations Data		
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 build up 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)
Commercial Data		
Start ups	SME no. SME turnover SME profit/loss SME sustainability SME employment Data Products Data Services Customers	1. Self-reporting 2. Industry surveys (data services sector) 3. Financial reporting data 4. Investment data 5. Sector employment figures 6. Market survey data (products and services)

Fig 3: Potential Data Value Metrics for Valuing Data [The original version of the table above was created for the European Data Portal based on the SCIFI project].

9. Appendix

A. Methodology for developing the metrics assessment process

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, March - June 2019.

These were combined with Expected Impacts taken from Business Case and Challenge Documents (available on www.smartcityinnovation.eu)

Delft 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

- 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

Bruges/Mechelen 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

- 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.

St Quentin Expected Impacts

- Energy savings by optimizing water consumption
- Reduction in water and labour costs
- Replicability, interoperability and transferability of the solution