Interreg
2 Seas Mers Zeeën
European Regional Development Fund
EUROPEAN UNION

Open data guidance package for cities (d1.4.1)

WP1: Transnational Ecosystem for open data innovation in the public sector

A.1.4. Compile Open Data guidance based on top-level harmonisation approach

Deliverable Nature	**Report (R)**
Dissemination level: (Confidentiality)	**Public**
Contractual delivery data	**01-07-2018**
Actual delivery data	**10-12-2018 Version 1.0** **22-01-2020 Version 2.0 (after first round of pilots)**
Keywords	**open data, open government data, guidance package, smart cities, smart city, SME's, training, advice, data management. SCIFI**
Contact Information	**hello@smartcityinnov.eu**
	htorn@delft.nl

Executive summary

In order to create (added) value out of their data, cities need to open up their data in such a way that the data can be used, reused or republished. The aim of this ‘open data guidance package’ is to enable city partners in the Smart Cities Innovation Framework (SCIFI) project consortium as well as all other cities interested in the publication of open data to use a harmonized approach, increasing possible (re)use and ensuring interoperability and replicability of data.

Many articles on open data have been written and published. This guidance is not reproducing existing knowledge rather than guide the reader through some of the major considerations and hurdles to take for cities as they (strive to) publish open data. Based on four phases cities can publish open data: from the actual datasets of the city to selection (phase 1), preparation of publication (phase 2), publication (phase 3) and reuse of the data by the end-user / reuser (phase 4).

Local contexts, such a politics, organizational structures and technical and data infrastructures influence the possibilities of creating a harmonised publication process. Nevertheless, cities can choose to harmonise some aspects, such as a standard for metadata, the (open) license for reuse and documenting considerations and decisions taken in the publication process. Six recommendations are given to the SCIFI cities in order to create top-level harmonisation.

After the first round of pilots we can draw some first conclusions and give extra recommendations for mid-sized cities. Open data alone is often a small part of the developed solution. And an IoT strategy, sensordataplatform and clear privacy by design principles are basic needs in order to create value with open data in smart city solutions sustainably.

With this second version, the work is not finalized. It will form another point of departure for analysis of practical use cases where gaps are filled and reactions and additions of experts are gathered during the execution of the accelerator and overall project up until 2021.

Table of Contents

Introduction

In order to create (added) value out of their data, cities need to open up their data in such a way that the data can be used, reused or republished. That is, also ready to use as a means in developing innovative concepts or applications. The aim of this 'open data guidance package' is to enable city partners in the Smart Cities Innovation Framework (SCIFI) project consortium as well as all other cities interested in open data to use a harmonized approach, increasing possible (re)use and ensuring interoperability and replicability of data.

Under the leading notion of open data experts and cities in the consortium of the SCIFI project, this state of the art hands-on guidance is developed for cities and addresses open data policies, regulation standards, platforms, formats, and other open data related topics. Based on a questionnaire and multiple (online and face-to-face) meetings an analysis has been done on the differences and similarities between the SCIFI partner cities. Assessed is which existing approaches fit each city's open data status and legacy systems to support an incremental approach to opening up data. The result that lies before also explores linking open data platforms to non-government data, and approaches to privacy issues and Intellectual Property Rights (IPR). The result is not a reproduction of existing knowledge nor is its aim to create new: it is indeed a guidance for cities through the knowledge and practice of open data. A hands-on living document that will be revised and enriched throughout the project. At the end of the project this document is part of the framework of the smart city innovation framework.

First we briefly discuss what open data is and the possible value it might unlock (for cities). Then we present the Open Data guidance including a standardized publication process. The reader will find information about policies, the importance of knowing where the data is and the origins of the data and we guide the reader through the open data publication process. In the third chapter we discuss the differences and similarities of the SCIFI partner cities regarding open data. Lastly we give recommendations for the SCIFI partner cities in order to create top-level harmonisation.

Open Data Guidance in the SCIFI project
Interpreting this deliverable in the light of the total project, the timeline of the project as well as the relationship between this deliverable and Activity 2.1 and Deliverable 2.1.1, the focus of this deliverable lies on a guidance package based on expert knowledge and an analysis of (best) practices. Thus to discover: how do the different cities currently open up their data and how do these approaches differ or fit into best practices as mentioned in literature? The guidance package shows how cities intend to publish data at the appropriate time (i.e. when the challenges are finalized and during the accelerators). Deliverable 2.1.1 will focus on the practical situation, the actual datasets published and needed for the specific SCIFI challenges.

Expert partners in the SCIFI consortium guide city partners in the project in this open data process. Business associations provide the business perspective to ensure the data approach is compatible with and scaled to the technical requirements of businesses (user-driven approach to opening and making data available for innovation).

Expert partners will work with each partner city to ensure the readiness of the data required for the solutions to be developed in the project. The focus lies on the data needed for value creation. Data structures will be analysed and data will be made as compatible as possible without disrupting each city's legacy systems.

This should lead to a, for the project and ecosystem, approach on how cities can open up data and what steps they need to consider. The guidance package will contain both knowledge on the different open data related topics as well as 'services' that may be delivered by expert partners or other partners from the SCIFI consortium. Services are advice or hands-on tools regarding publication of open data.

With this second version, the work is not finalized. It will form another point of departure for analysis of practical use cases where gaps are filled and reactions and additions of experts are gathered during the execution of the accelerator and overall project up until 2021. With this version, the work is not finalized. It will form the point of departure for an extensive round and practical use where gaps are filled and reactions and additions are gathered during the execution of the accelerators and overall project up until 2021.

Foreseen activities will include:

- improving the knowledge base of the addressed topics in this package.
- meetings (online and face-to-face) between cities and expert partners in the project to analyse data structures and data. Output of these meetings provide input for this guidance package.
- enriching the preliminary findings of open data differences in mid-sized cities participating in the SCIFI project by evaluating the publication of datasets during the project.
- reaching out to other cities and knowledge institutes to review the guidance package.
- linking the services better to contextual information, local characteristics and presence of specific conditions.

1. About Open Data

Many articles on open data have been written and published. This chapter is not aiming at reproducing those existing insights and knowledge rather than pointing out where the reader can find relevant insights and knowledge. We will discuss briefly what open data is, open data related to cities and 'smart cities' and the value that might be created with open data published by cities.

1.1 What is open data?

Open data is digital data that are accessible without any restrictions for use, reuse and republish.

Besides European legislation, there are several national laws and regulations that determine whether or not a dataset is (partially) qualified as open data or not. Open data is also not the same as sharing data. The Open Data Institute states: "Groups sharing information with each other is different from opening it up for all to access. Your private data should only be open if you choose to share it. (But if you want to know who's accessing or sharing your data, open data can help.)"[1]

Open data sets also are defined in relation to other types of data, especially big data (Figure 1). Big data consists of large datasets that require specific analysis techniques because they cannot be handled in a conventional way, since they exceed the capacity of the usual technological tools for collecting, managing and processing data.

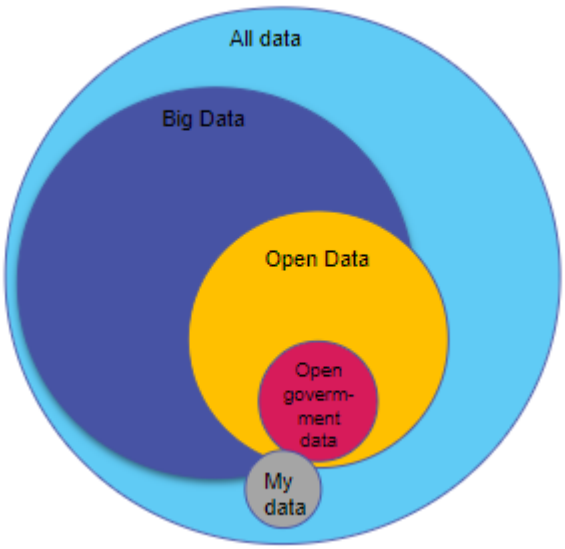

Figure 1: The relationship between open data and other types of data. McKinsey Global Institute analysis. Open data: Unlocking innovation and performance with liquid information.

Open data is often big data, but small data sets can also be open. Sensors that collect data give often big datasets while a dataset of a city with parking spots for disabled people is ought to be relatively small. Open and big data are distinct concepts. Open describes how liquid and transferable data are, and big describes size and complexity of data sets. The degree to which big data is liquid indicates whether or not the data are open.[2]

[1] The Open Data Institute (3-11-2017). "What is open data and why should we care?" Retrieved from URL: https://theodi.org/article/what-is-open-data-and-why-should-we-care/ on 20-4-2018
[2] Retrieved from SCIFI output D1.5.1. on innovative procurement

Open data is also mentioned often in the light of the concept of Smart City, where it can take up a role of enabler as it places key information into the hands of citizens and those with the ideas and technical knowledge that is required to solve Smart Cities problem. “Open data projects therefore hold great potential to provide citizen-centric solutions, optimizing smart city services according to the needs and preferences of the local people, in alignment with geographically differing customs and policies.”[3]

Why open data is relevant is stated by Halonen (2012): “In terms of government data, the argument goes that data that is created by public funds should be free for all to use and re-use. Data is thus seen as a common resource that does not suffer from scarcity. It is seen as morally wrong to restrict the use of data to just a certain group of people.”[4] Charalabidis et al. (2018) state that by proactively opening up public data or open government data, cities “can create considerable benefits for several stakeholders, such as forms and individuals interested in the development of added digital services or mobile applications, by combining various types of Open Government Data (OGD), and possible other private data.”[5]

Tim Berners-Lee (2010) listed a five-star growth model within basic principles (of re-usability) of open data (and re-usability of data) that helps us come to a definition:[6]

★ Data is available on the web (in whatever format), but with an open licence

★★ Data is available as machine-readable structured data (e.g. in Excel, instead of an image scan of a table)

★★★ As in two stars plus non-proprietary format (e.g. CSV instead of Excel)

★★★★ All the above plus use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at things created by others

★★★★★ All the above, plus linking your data to other people’s data to provide context

Considering the above, the SCIFI project defines open data as:

- accessible without payment or at negligible cost.
- not subject to any copyright, patent, trademark or trade-secret regulation.
- accessible without registration.
- machine readable (at least three stars of the Tim Berners-Lee model).
- foreseen of metadata.
- as raw (and as complete) as possible.
- findable.

[3] Bee Smart City (2017). Benefits of Open Data for Smart Cities. Retrieved from URL: https://hub.beesmart.city/solutions/benefits-of-open-data-for-smart-cities on 20-4-2018

[4] Halonen, A. (2012, 19). Being open about data: Analysis of the UK open data policies and applicability of open data. The Finnish Institute in London, London. Retrieved from URL: http://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf on 20-4-2018

[5] Charalabidis, Y. et al. (2018) The World of Open Data, Public Administration and Information Technology 28, URL: https://doi.org/10.1007/978-3-319-90850-2_9

[6] Berners-Lee, T (2010). Linked Data. Retrieved from URL: http://www.w3.org/DesignIssues/LinkedData.html

1.2 Creating value through open data

Cities, specifically the SCIFI cities, investing in opening up data aim at creating added value out of it. There are different types of value that can be pursued with opening up data. As Halonen (2012) states: “Open data is intrinsically a combination of various different things and thus is associated with different objectives and benefits for different groups of people. There is not any single pattern of goals but various interrelated application areas, which together form a compilation of objectives and potential benefits.”[7] When disclosed, data can stimulate economic growth, better decision-making, more transparency and efficiency of governments, as well as higher quality of life and more inclusive societies, according to the European Data Portal.[8]

Within the context of the launch of the European Data Portal, further evidence of the quantitative impact of re-use of Open Data is measured.[9]

No matter the added value a city in the project strives for, there are three elements to take into account when creating the desired impact: the challenge, the stakeholders and the datasets related.

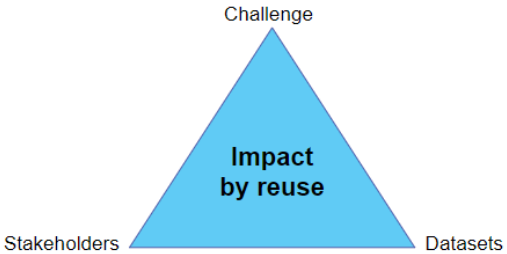

Figure 2: three components to create impact with open data

In The Open Data Goldbook, the following summary of benefits of open data is visualized:

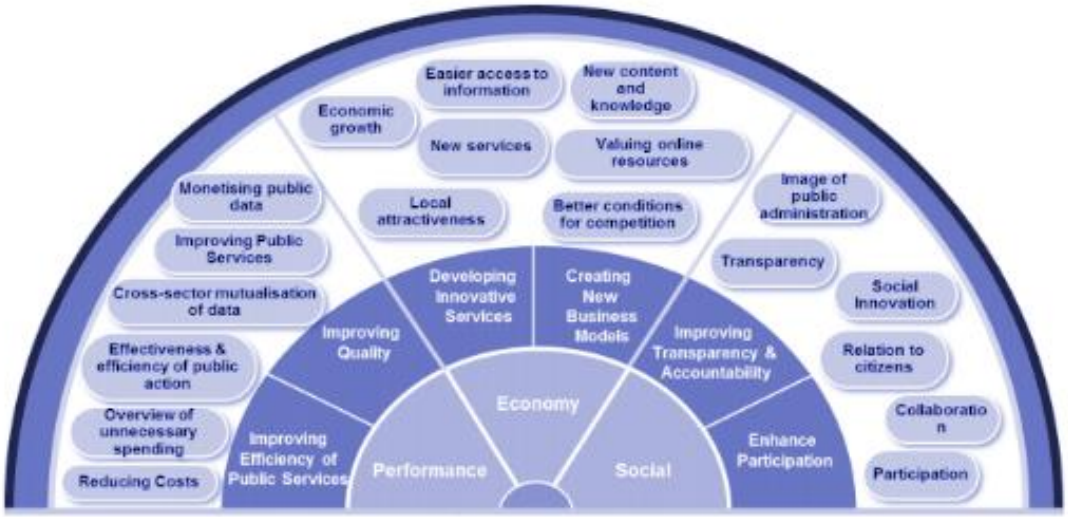

Figure 3: Benefits when Open Government Data is re-used (source: The Open Data Goldbook)

[7] Halonen, A. (2012, 19). Being open about data: Analysis of the UK open data policies and applicability of open data. The Finnish Institute in London, London. Retrieved from URL: http://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf on 20-4-2018
[8] The European Data Portal. The economic benefits of Open Data. Retrieved from URL: https://www.europeandataportal.eu/en/highlights/economic-benefits-open-data on 20-4-2018
[9] European Union, 2015. European Data Portal: Creating Value through open data. Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf on 20-4-2018

We identify five different types of value we strive for with the SCIFI project, depending on the challenge provided and the type of solution a city is looking for (see Figure 3 below).
Cities can strive for creating economic value (direct and indirect) and opening up data might lead to quality improvement of the data. Opening up data might increase findability of information and thus increase transparency for the city as well. The given challenges ask for innovative solutions that also might improve public services.

When opening up data, cities should be aware of the value they are pursuing, for the type of value might ask for different interventions in the publication process or a different focus in the open data policy. The value a city might strive for may also influence the commitment of the organisation to publish open data.

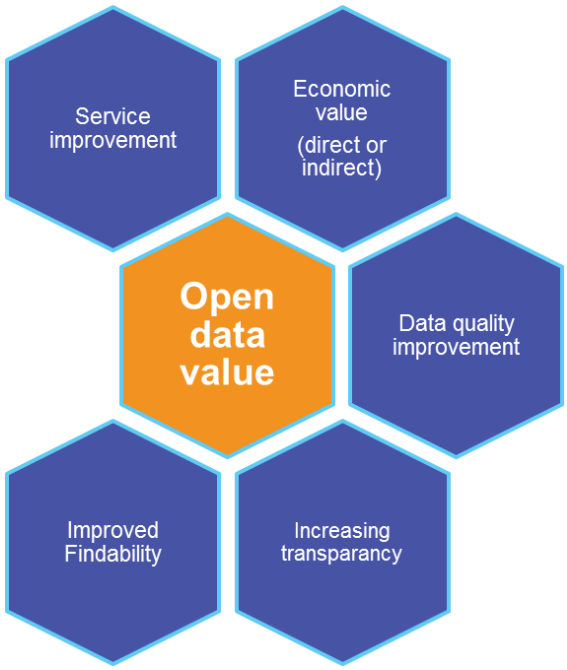

Figure 3: Types of added value with open data that may be pursued by citie in the SCIFI project.

1.2.1 Economic value

According to the EU Commission, the value of the EU data economy is foreseen to pick values beyond EUR 600 billion by 2020, representing 3.17% of the overall EU GDP, more than double of the 2015's value. Within the data economy open data plays a pivotal role, whose value will reach near 70 billion EUR in 2020.[10]

In response the economic opportunities as presented by the EU commission and the increasing availability of open (government) data, several business models have been developed.[11]

[10] Retrieved from SCIFI output D1.5.1 on innovative procurement.
[11] Zeleti, F.M., Adegboyega, O. and Curry, A. (2016) "Exploring the economic value of open government data", Government Information Quarterly Volume 33, Issue 3, July 2016, Pages 535-551

Economic value might be:

- companies that establish in the city
- companies developing new products or services with open data. Those services might lead to better or more efficient services for the city.

1.2.2 Data quality

As the GAP-analyses of SCIFI cities show in chapter 3, the effort a city puts in opening up their data does not always result in the value the city strived for when it started. For example, the city of Brugge organized hackathons that led to new products but many of those turned out not to be sustainable.

Different reasons might be underlying and one of those might be a poor data quality. If the quality of opened data leaves much to be desired, use and reuse might not happen. Let alone that if the data is reused, real impact and value is established. The Open Data Institute even states: “If you don’t think you have a quality problem with your data, you haven’t looked at it.”

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.”[12]

1.2.3 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. And publish those datasets in such a way that they are accurate and timely. Publishing the data on a central website or findable via a central website makes it easy to find not only for third parties (end-users) but also for those within the municipality to find the data. Resulting in higher efficiency. As written in the Open Data Goldbook: “When data is open, none of your colleagues will have to go through an internal process to receive particular data. Many organisations have encountered the benefit of having their data open, simply because it takes less time to find data.”[13]

[12] Charalabidis, Y. et al. (2018, 57) The World of Open Data, Public Administration and Information Technology 28

[13] The Open Data Goldbook. Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/goldbook.pdf on 13-3-2018.

1.2.4 Transparency

Outstandingly relevant are the potentials and opportunities of additional transparency in government. “Organisations increase transparency when they expect valuable external influences and are interested in a more intensive interlinking with their surroundings, without the risk of getting damaged.”[14] Open data offers freely access to government data and information to all and thus it increases the transparency of cities.

1.2.5 Improving services

With the publication of open data it’s reuse might lead to improving services as third parties may develop new solutions or tools that may improve the way a city provides a service or 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 the delivers services.

[14] Geiger, C. & Lucke, J. von (2012) Open Government and (Linked) (Open) (Government) (Data). JeDEM 4(2): 265-278.

2. From open data to value: guidance for cities

There are many books, tutorials and other literature available about open data. As a part of the SCIFI project, several have been read. Although many contain important insights, none of them seem to be practical (and concise) enough for (SCIFI) cities who want to publish open data. The ones read and reviewed that come close to desired practicality and applicability are the Open Data Goldbook by The European Data Portal[15], information from the Open Data Institute and the 'Roadmap Linked Open Data' published by TNO.[16] The latter one explains nine steps for data providers on how to publish linked open data but it is also possible to apply these steps for open data. Another practical approach is that of the Australian Open Council Toolkit.[17] The recently published book 'The World of Open Data' also contains many insights on different aspects of open data.[18] These and others have been used to shape this guidance for cities.

When it comes to publishing data as open data, alike other areas, cities have many differences. Differences that cannot all be overcome for the sake of creating value with open data in a harmonized approach for SCIFI. Local organisational structures, political contexts, data infrastructures and data quality influence the way cities publish open data as well as the durability and sustainability of a publication process or selection of datasets. These local contexts and circumstances may limit the possibilities for open data or give opportunities for publication. Moreover, cities should be aware of the value they are pursuing with open data, as the desired value might ask for different interventions in the publication process or open data policy.

This guidance-chapter provides cities a generic though hands-on guideline for publishing open data in order to ensure a high (quality) level of open data and to manage expectations of cities and users towards the data published. Thus, this chapter helps cities in shaping their specific processes and policies for open data, taking into account every city's unique contexts.

2.1 Open data policies and governance

The trigger for cities to start with releasing open data highly differs. Some cities start with a political trigger, some reply to a request from citizens, others to a demand of a business that sees an opportunity for adding value to their company's products or services. In some countries it is necessary to release (open) data by law. And some cities started publishing open data because public servants thought it might be relevant as an end-product from information processes. The GAP-analyses of the SCIFI cities show these different triggers as well.

[15] European Dataportal, open data goldbook, URL https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf
[16] TNO, Stappenplan Linked Open Data, URL: http://www.pilod.nl/wiki/BoekTNO/stappenplan
[17] URL: https://opencouncildata.org/how-to-publish-data/
[18] Charalabidis, Y. et al. (2018) The World of Open Data, Public Administration and Information Technology 28

Use case: Delft

In Delft the city's board of aldermen (2014-2018) strived to start publishing open data and unlock economic value, improved data quality, better services and higher transparency. The responsible alderman gave orders to develop a strategy for open data in 2016 and after that the city explored publication via a pilot in order to develop a publication process.[19]

The trigger to start with open data often determines the first steps taken to publish open data as well as the ambition. That is, whether or not a city starts from writing a policy or start with a pilot, or both or differently. The initial trigger also influences the approach chosen, the focus on creation of value and the commitment of the organisation. To build a open data strategy, the European Data Portal offers in the Goldbook a step-by-step guide that is highly recommended.[20]

No matter the trigger to start with open data, a well-written policy defines the commitment of the city to publishing data. Commitment in terms of organisation, effort and investment. The Open Data Institute published a guide for organisations with considerations when it comes to open up data, from context to specific topics such as licensing, technique and measuring reuse.[21] Besides (organisational and political) context, a good policy should incorporate the following topics according to the Open Data Institute:

- approach to identifying and prioritising data for release.
- privacy considerations
- data licensing and reuse rights
- data publishing standards
- engaging with reusers
- measuring success
- approach to consuming open data
- concrete commitments of the organisation
- policy transparency

The level of detail of the listed topics differs as cities sometimes focus on delivering general principles that should be followed, like the Smart Flanders region in Belgium.[22] These principles may indeed lead to more detailed guidance or process for those involved in opening up data. The city of Brugge for example, has defined a detailed policy.[23]

[19] City of Delft, Open Delft Strategy' (2016) Retrieved from URL: https://ris.delft.nl/document.php?m=1&fileid=295228&f=5ea3d4b26ab0a923489f30d8e356e336&attachment=0&c=61762 on 20-4-2018
[20] The Open Data Goldbook (,14). Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/goldbook.pdf on 13-3-2018.
[21] The Open Data Institute. "How to write a good open data policy". URL: https://theodi.org/article/how-to-write-a-good-open-data-policy/
[22] Smart Flanders Open Data Charter. URL: https://smart.flanders.be/open-data-charter/
[23] Brugge Open Data. Retrieved from URL: https://www.brugge.be/opendata

We will discuss briefly approaches to open data (by design and by default) and their influence on the governance of open data, data licensing and platforms for publication before we present the open data publication process.

2.1.1 Approaches to open data

Key to being able to trust data is knowing by what means it has been disclosed and where data originates. If (re)users do not trust the data, it is unlikely that they might use the data and thus it is unlikely to gain value out of it. If end users do not trust data, they are unlikely to believe they can rely upon the information for accountability purposes. Publishing open data without being certain about quality and having a process and organisation in place decreases the chances of creating value out of it.

By Design and By Default

One way to approach open data is to consider open data as a resource in solution development for challenges of the city. This open data 'By Design' indicates that cities, in all their processes, projects and developments take into account the possibility of open data as a means. Open data thus, is not being published for the sake of releasing open data but releasing open data with a goal, creating the possibility to connect to specific end users and stakeholders.

Next to By Design, some cities publish open data By Default. By Default means that open data is a goal on its own and that value creation or impact comes from serendipitous reuse of the data. Cities publish the open data as a standard outcome of an (information)process (if the data is in fact open). If one is interested in data, he will come and look for it in the city's data portal. Some cities may publish hundreds of datasets but not all will be linked to the city's goals or projects and developments. Effort of the city lies more in releasing the data and less in connecting with end-users,

Data, data, where are thy?

As well as with By Design, a successful By Default approach requires that cities are well aware of where there data is stored, where it originated, have a clear view on quality and a good understanding of data management. They take open data into account when buying (data related) services from third parties, preventing themselves from losing the control over and ownership of the data. This helps cities to achieve their goals with the help of (strategically published) open data and create value more rapidly.

Not all cities however, can check these boxes. This means that cities often also need to invest in good data management before being able to release (qualitative) open data (in a sustainable way). Not only may this help cities improve quality of data, it also enhances the chances of developing a proper business case for SME's or other organisation wanting to reuse the data for the data is considered to be more trustworthy. The questionnaire presented in Appendix 1 is a tool cities could use to map their data landscape.

2.1.2 Licenses

"Data which is shared with a license becomes Open Data."[24] In order to stimulate re-use of data, cities should apply a clear Open License to the open data they are providing. As stated in the report of the Open Knowledge Institute about the state of open data, open licensing can be a problem. "On the one hand, more governments implement their unique open data license versions. Some of them are compliant with the Open Definition, but most are not officially acknowledged. On the other hand, some governments do not provide open licenses, but terms of use, that may leave users in the dark about the actual possibilities to reuse data. There is a need to draw more attention to data licenses and make sure data producers understand how to license data better."[25]

The European Data Portal has a license assistant that gives an overview of possible licenses and guides end-users on how to implement these licenses or combine several different ones.[26]

2.1.2 Platforms for publication

Before starting publication, cities need to be able to publish the data at a central point, increasing the findability of the data. In the Open Data Toolkit of the World Bank different options have been proposed for releasing open data on platforms (with a catalogue). They state that there are a few common characteristics in the different possible platforms.[27] All platforms have:

- Easy access.
- Search possibilities.
- Machine readable data access.
- Metadata available.
- Clear data licenses mentioned.
- Data preview or visualization.
- Standards compliance.
- Application Programming Interfaces (API's).

It is not always necessary to invite tenders, benchmark or purchase a platform (as a service). Some cities might already have applications that offer functionalities that meet the city's demands for an open data platform. Cities should take into account their ambition, their budget, end-users and the sustainability of a platform in their consideration. For example: to what extend does the city want to offer visualization of the data for the end-users?

[24] The European Data Portal Licensing assistant. Retrieved from URL: https://www.europeandataportal.eu/en/content/show-license
[25] Global Open Data Index (2017) The State of Open Government Data in 2017. Retrieved from URL: https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf
[26] The European data Portal Licensing assistant. Retrieved from URL: https://www.europeandataportal.eu/en/content/show-license
[27] The World Bank, Data Technology Options. Retrieved from URL: http://opendatatoolkit.worldbank.org/en/technology.html on 3-3-2018.

Again, local context, regulation and (data)infrastructure influences the choice for a platform. However, in order to strive for harmonisation and create better interoperability, cities may consider certain techniques. For example platforms build with open source components.

Note that some national governments offer a catalogue where one can find all the open data (sources) that are offered by government organisations. It improves the findability of the open data of the city if they also connect with this platform.

2.2 The open data publication process

Four generic phases of the publication process can be defined. Those are: selection of the dataset, preparation of publication, the actual publication and shortly after publication and lastly, the (re)use of data. Within those phases, 12 steps can be taken and considerations made by cities in order to ensure (re)usability and durability of releasing qualitative data for value creation.

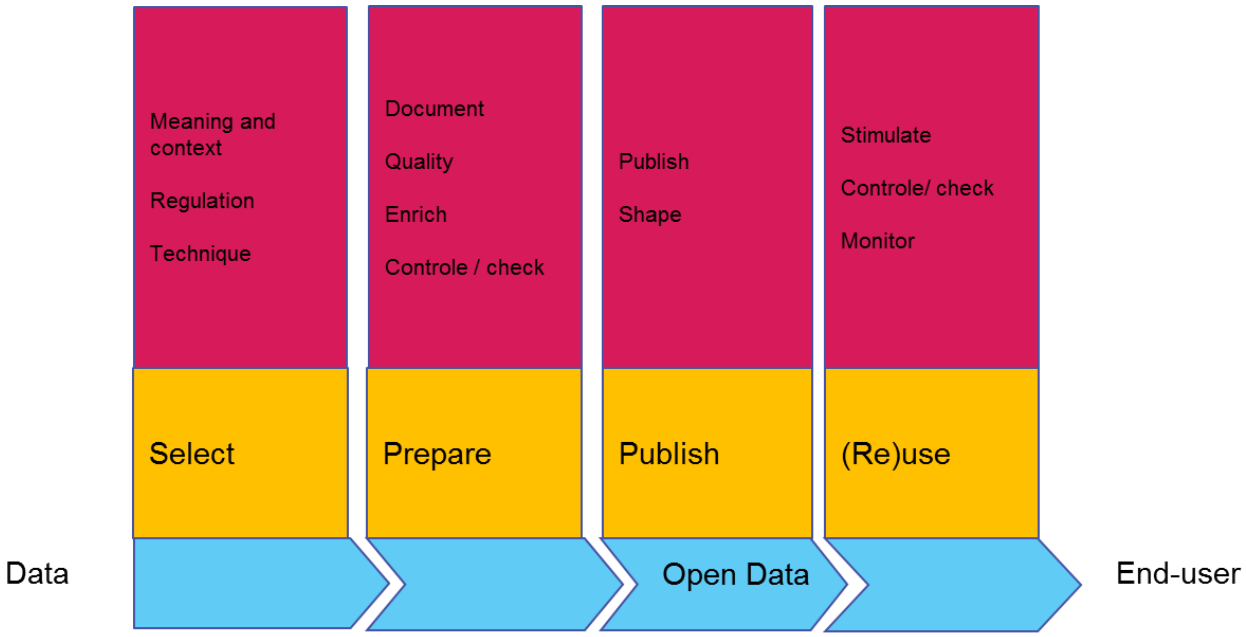

Figure 1: Open data publication process, based on the one presented by Mulder, B. (2016).[28]

The image above might suggest the process ends with the end-user. The end-user, however, needs to be able to give some sort of feedback, creating a loop in the publication process that helps the city improving the data(quality), the process of publication and even the organisation. In the last phase of the process, this interaction with the end-user is covered.

[28] Mulder. B. (2016). Presentation LOD workshop ESI HHS, eSociety Institute of The Hague Institute of Applied Sciences. Retrieved from URL: https://www.slideshare.net/eSocietyInstituut/presentation-lod-workshop-esi-hhs on 20-11-2018

2.3 Phase 1: select datasets

The first phase of the open data publication process, the selection of datasets that might be published as open data, sets the boundaries for and indicates the work that needs to be done in the other phases of the publication process. Based on this phase a city knows if the dataset may be published as open data and what it (possibly) takes to do so.
This phase is the most timeconsuming at first sight but if done right, saves time during the other phases and enhances the chances of reuse of the open data. It may also lead to the decision of the city to not proceed with the publication process.

Within the selection phase, one needs to take three steps:

1. meaning and context.
2. regulation.
3. technique.

We will address these steps in the following subparagraphs. If cities have an overview on where there data is, for example by using the questionnaire of Appendix 1, they may already have these topics (partially) addressed.

2.3.1 Meaning and context

Specifically in a governmental (related) organisation and/or political organisation, one should first have a clear view on the meaning and context of the trigger (and dataset) before heading towards actual preparation and publication of the dataset(s). Thus, a first analyses of stakeholders should be done as well as contextual analysis. In defining a challenge (for SCIFI), the meaning and context should be addressed by the city already.

Cities can answer the following questions in order to get that understanding:

1. Who and / or what is the trigger to start this process? It is important to understand why a person or event caused a trigger, for it gives insights in how a city might shape the publication process and the involvement of stakeholders and/or end-user. If a trigger comes from businesses, it values if a city understands the possible added value a business is looking for. It could also determine the desired format in which a city publishes data.
2. What is the political context? Is there a political chance or risk when releasing these data?
3. Is there a specific organisational context that should be taken into account? To what extent is there organisational commitment to the publication of (these) data.
4. Where can we find the data? Is it situated at internal servers, external servers? Is it a Data related context ; history of data collection, application or not? And how does this influence the possibilities of publication?

Cities could think of several use cases of the data besides the one that caused the trigger for publication. Having use cases in mind, helps the city to take hurdles in the publication process. The Open Data Institute created a tool that helps mapping data ecosystems, thus helping cities and other (public) organisation to understand and explain where and how the use of data in specific use cases creates value.[29] An additional benefit of this tool might be the identification of multiple stakeholders (that could be involved in the publication process..

2.3.2 Regulation: is the data open or not?

As written in the first chapter, it mainly depends on regulation both national and international (EU) whether or not data is considered to be open. Cities have to consider (1) ownership of the data, (2) personal information in relation to the GDPR and (3) other legislation that might prohibit the publication of data.

If data is open or not can be found by asking the following questions:

- What rules and legislation apply to the specific dataset and / or context?
- Is the dataset to be considered as 'open' considering the legislation?

These questions might be answered based upon a short questionnaire, as to be found in Appendix 4 of this guidance package.

GDPR and privacy

As far as open data are personal data, i.e. related to an identified or identifiable natural person, any processing of personal data must comply with the applicable legislation on the protection of personal data. Privacy considerations should have been given a place in the open data policy of cities. We want to ensure that personal information is not released by mistake and recommending steps to mitigate, e.g. by undertaking privacy impact assessments or approaches to anonymization. Coming straight from the European Dataportal, a city should address the following privacy concerns releasing data:[30]

1. Understand the data. Consider potential use cases, the value of the data and potential risks.
2. Consult. Engage stakeholders about the publication programme, be mindful of additional risks that are identified.
3. Remember the three pillars of privacy, data protection and public confidence.
4. Be very sure of the grounds for publishing personal data.
5. Anonymise well and thoroughly. Follow guidelines for anonymizing personal data.
6. Remember utility. There is no point publishing data which has been denuded of serious content.
7. Don't release and forget. Anonymization and Open Data are not cheap options.
8. Have a plan in place in the event of a problem. Be not only transparent, but also transparent about your transparency.

[29] The Open Data Institute (2018) Mapping Data Ecosystems. Retrieved from URL: https://theodi.org/article/mapping-data-ecosystems/ on 4-4-2018.
[30] European Data Portal: How address privacy concerns when opening data? Retrieved from URL: https://www.europeandataportal.eu/en/highlights/how-address-privacy-concerns-when-opening-data

There are several guidelines available on anonymizing data. Sometimes, just removing fields from a row is insufficient. Cities need to choose a right strategy for anonymizing. Such as aggregation, dithering, hashing and coding.[31] However, existing guidelines on privacy sensitivity still require a lot of interpretation effort by the data provider.[32] Defining the context clearly as well as defining potential use cases help interpreting these sensitivities and choosing the right strategy.

Ownership and Intellectual Property
Make sure to know if the data is owned by the city and if there are no other rights, such as intellectual property rights, that form a constraint for releasing the data. Unlike for material such as text, music or film the legal situation for data varies widely across countries but most jurisdictions do grant some rights in the data (as a collection).[33] It may also be the case that some intellectual property rights are owned by a software or hardware provider.

Open data when released, as stated in the Open Data Goldbook, is free of intellectual property. It is free to download, manipulate and re-use for any purpose.

2.3.3 Technique

The last step in this first phase of publication is to take a look at the more technical side of the dataset, the origins of the dataset and how the dataset is released in a way that is sustainable, timely and accurate. If a platform and process are in place, a city should use techniques that fit the demands of the chosen platform and the process. For example: the city of Delft publishes the geographical data at a Geo Information Server (GIS) from where it is automatically harvested to the open data platform. This also applies to a data landscape or data model or datawarehouse, if cities have these in place.

Cities need to consider the following:

1. The origin of the dataset and how the data is collected. This helps with determining the right and possible formats to publish the data (XML, JSON, as linked data, etc.). Note that raw data is more likely to be produced using formats customized to the specific data, the tools used, or industry standards,[34] and that the desired format might defer from the customized one.
2. What would be the best timing of release of the data? Real time, every hour, every day, every month? The context in which the data is collected or published may influence this as well as in what timeframe the data is collected. It might happen that the requested release time to meet the city's challenge c.q. solution development does not match the actual timeframe in which the data is collected. If this is the case it might be considered to change the actual data collection or collect new data that

[31] Open Knowledge Foundation. The Open Data Manual: Making personal data anonymous. Retrieved from URL: https://odm-test.readthedocs.io/en/latest/appendices/making-personal-data-anonymous.html
[32] Charalabidis, Y. et al. (2018, 60) The World of Open Data, Public Administration and Information Technology 28
[33] Open Knowledge Foundation. The Open Data Manual: What Legal (IP) Rights Are There in Data(bases). Retrieved from URL: https://odm-test.readthedocs.io/en/latest/appendices/what-legal-ip-rights-are-there-in-databases.html
[34] Bennett, D. and Harvey, A. (2009). Publishing Open Government Data. Retrieved from URL: https://www.w3.org/TR/gov-data/ on 8-11-2018

matches the requested timing. In the words of the Open Data Institute and Open Data Support: adapt the update frequency of data to the nature of the data and its intended use.[35]

3. Is it possible and desired to publish the data directly from the source or do you publish copies derived from the source? In real time, harvest, copies of data, etc.

2.4 Phase 2: prepare publication

In the second phase cities deal with any constraints and take hurdles before processing and releasing the open data. By this time, the city should know whether or not the data is open, if it is possible to publish the data as open data and in which formats it could be (or is desired to be) published. The constraints cities might find in this preparation phase lie in those areas that were indicated in the first phas but seem to lie mostly in in privacy, data quality and technique. These will be covered in the steps of documenting, the data quality, enriching the data and control the steps before the actual publication.

2.4.1 Document

In phase 1, cities have determined whether or not the data is public and if so, if it is allowed to publish the data as open data. In the preparatory phase of publication the first step is to document the process and taken decisions regarding the data and upcoming publication. This includes the desired value, context and legal aspects. Especially if a city decides not to go ahead with the open data publication process, it is important to argue why. Furthermore, by documenting the considerations made (in a standardized format), the city builds an archive that it can use when end-users ask certain questions about decisions made in the process. The archive of documentation can also be used as examples in new possible open data releases.

2.4.3 Data quality

Before we mentioned that key to being able to trust data is knowing by what means it has been disclosed and where data originates. These are two elements of data quality cities should take into account. “The overall quality of data is not only important in terms of reusability but also towards credibility when it comes to open governmental data.”[36] It requires effort to release relevant and qualitative data but it serves multiple purposes.

In the TNO nine steps handbook on linked data, they mention the following aspects of data quality that should be taken into account by cities when checking the quality of the data they want to release. We have enriched these aspects with examples and definitions as given by the European Data Portal and Open Data Support.[37]

[35] The European Data Portal. (2014) Presentation metadata: Open Data & Metadata Quality. Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_2.2_open_data_quality_en_edp.pdf on 16-4-2018

[36] Charalabidis, Y. et al. (2018, p.101) The World of Open Data, Public Administration and Information Technology 28, URL: https://doi.org/10.1007/978-3-319-90850-2_9

[37] The European Data Portal. (2014) Presentation metadata: Open Data & Metadata Quality. Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_2.2_open_data_quality_en_edp.pdf on 16-4-2018

Metadata	*Name of the dataset, description of the dataset, name of the publisher, locations, release (date), potential use, compliance, production date, provenance.* The metadata describes important information about the data set. Having the metadata in place, is having a part of your communication towards end-users in place. When creating the metadata, cities should have these end-users in mind. There are several metadata standards cities can apply, such as DCAT, INSPIRE or CKAN attributes.[38] SCIFI cities will use the DCAT-standard. Note that the European Data Portal sees metadata as a type of data and that quality aspects of data thus also apply to metadata.
Dataset	Accessibility, format of the dataset, kind of data, identifiers, use of vocabularies, semantics, data model, links, size of the dataset, concise, complete, believability, reputation. The European Data Portal and Open Data Support give the following recommendations: • Make appropriate attributions so that re-users can determine whether or not they can trust the data. Thus: is the dataset complete or not? If not, why? Etc.
Data records	*Validity, complete, consistent, unique, timely accurate, precise* The European Data Portal and Open Data Support give the following recommendations: • Balance the accuracy of your data against the cost in the context of the application; it needs to be good enough for the intended use. Make sure that there is organisational commitment and investment in procedures and tools to maintain accuracy. • Process all data before publication to detect conflicting statements and other errors (in particular if data is aggregated from different sources). • Monitor the update mechanisms on a continuous basis to ensure completeness.
Process	*Issues documented, process described, update frequency, support (official, community, tools)* The first step of this phase as part of the data quality.
Availability	*Rights, license/ fee, SLA, authenticity, security* The European Data Portal and Open Data Support recommend to make sure that responsibility for the maintenance of data is clearly assigned in the organization.

[38] Open Data Monitor (2015). Metadata standards. Retrieved from URL: http://knowhow.opendatamonitor.eu/odresearch/metadata-standards/

2.4.4 Enrich

Besides data quality, it is important to view the dataset in the light of the value and end-user. Is it rich enough to be of added value for the end-user (in solution development) or does it need be enriched with more data? Is it perhaps too difficult to use and does it need enriched metadata or examples of possible reuse?

Data enrichment is a quite general term that refers to processes used to enhance, refine or otherwise improve (raw) data.[39] Cities can think of different data enrichment activities:

- correct typographical errors in a database or likely misspellings. This could be done by precision algorithms.
- extrapolating data, where based on different possible methodologies more data is generated from the given (raw) dataset.
- change the process of data collection. For example: data may originally be required for a smaller goal than it is now. A dataset that is created to maintain green spaces in the city might demand enriching the data if the maintenance demands of a city develops. Changing the process of collection of data requires more effort of the city as you change the whole process.

2.4.5. Control / check

After documenting the considerations, decisions made regarding the publication process and undertaken activities for preparing publication, the last step of this second phase is to check if all was done correctly and exhaustive. Thus, preventing any difficulties or risks for after publication.

2.5 Phase 3: Publish data

As cities arrive in the third phase, the data is ready to be processed and released as open data. This phase as a will the last one, require little effort if the city has done the first two phases well.

2.5.1 Publish & Shape

Before pushing the ‘red button’ of publication, cities may consider the following:

Timing	What is the best timing to publish the data in order to stimulate reuse? Is there a political window of opportunity to stimulate reuse? Or perhaps the data is best published in a specific season. For example: a dataset with data about de-icing of roads probably has a higher chance of reuse if published autumn and winter than in summer.
Stakeholders within the organisation	Wo needs to be informed about the publication? Internally: who are the stakeholders that need to know about the publication or possibly are being reached out to be end-users?
External stakeholders	How do you want to inform or reach out to your possible stakeholders? Do you present the data in the format desired for them to reuse?

39

Contact information	Make sure contact information is accurate and clear. The contact person(s) should be available after release. Make sure as well that it is clear who within the organisation decides upon corrections if feedback has been given on the dataset.
Prevention of wrong use	To prevent possible misuse of the released data, cities might want to add information when releasing the data. Specifically if data has a possible political risk.
Findability	Is the data findable on the platform? And is clear where to find it from the first page? For example: some cities offer 'the latest datasets' on the first page of the platform. Others release data under categories, like 'mobility'.
Visualization	To what extent does the city want to offer the end-user a visualization of the released data? Some platforms offer visualisation on a map of the city. Others do not put a lot of effort in visualisation for they see visualizing the data already as an application.

2.6 Phase 4: (Re)use of data

As stated in the first part of this chapter and visualized in Figure 2, cities can create impact with three ingredients. Those are the datasets that are opened, a challenge the city has that the datasets relate to and thirdly involving stakeholders. "When data producers [cities in this case] publish new datasets, they have to make the datasets known to potential end users to optimize the chances that end users will take advantage of datasets and to facilitate the creation of new services."[40]

In the reuse phase, when the data has been published, cities reach out to end-users. In order to monitor the reuse and impact, they should consider what exactly they want to know about (re)use and how they could monitor it. For example: if a city wants to measure what the end-user thinks of the quality of the data, they might consider implementing a rating functionality. Perhaps a city wants to know how many views a dataset has, as well as the download rates.

In order to stimulate reuse, cities might also want to add examples where the data has been used already or where similar datasets from other cities have been used. Examples of reuse might inspire others. Publishing which processes use the data, end-users might indicate what public services are connected with it.

Cities might advertise on virtual community channels the information and reach potential reusers for their datasets.

[40] Foulonneau M, Turki S, Vidou G and Martin S. (2014) "Open data in Service design" Electronic Journal of e-Government Volume 12 Issue 2 2014, (pp99-107), Retrieved from URL: www.ejeg.com at 09-06-2018.

3. Analysis of open data maturity of partner cities

To shape the guidance package and to strive for harmonisation in the publication process, cities have been asked to fill in a questionnaire which is based on an existing questionnaire of a framework for comparison of open data policy maturity.[41] The questionnaire used can be found in the Appendix (1) of this guidance package. In addition a workshop was given during one of the partner meeting of SCIFI in 2018 and several teleconferences have been made. In case the city publishes open data, the open data platform has been taken into account. The results have been analyzed and this chapter provides a summary of and insights in the similarities, differences and challenges the SCIFI cities have regarding open data. We will also discuss these in the light of policies and background information provided by the cities as well as general open data challenges as indicated by the Global Open Data Index.[42] Thus with this chapter we give an insight in the 'open data maturity' of partner cities. This chapter leads up to the recommendations given to the SCIFI cities in order to create harmonisation in the publication process.

National differences

A first difference can be identified at national level. Every country has its own specific legal regulations when it comes to open data. The Directive on the re-use of public sector information provides a common legal framework for a European market for government-held data (public sector information). Specific European countries should have transposed the 2013/37/EC PSI Directive into the legislation of their country.[43] For France, The Netherlands and Belgium (the SCIFI city partners) there are differences in this transposing. Belgium has taken some specific measures at federal and regional levels, France and the Netherlands both have adapted their legislative framework for access to documents to include re-use of PSI. This difference in adopting the PSI Directive (may) have influenced the approach of open data of the different partner cities as well as their legislative frameworks.

General similarities and differences

In general we see the following similarities:

Ambitions and goals	Goals and ambitions for open data publication are similar: transparency, economical value, better public services. There is one city with a local open data strategy (Delft). Others have a regional strategy or principles. Except for one city, all cities publish open data. About 60 datasets have been published by Brugge, while Mechelen pubglished 0 datasets.

[41] Zuiderwijk, A. and Janssen, M. (2013). Open data policies, their implementation and impact: A framework for comparison. © 2013 Published by Elsevier Inc.
[42] Global Open Data Index (2017) The State of Open Government Data in 2017. Retrieved from URL: https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf
[43] Note that the European Parliament as from April 25th 2018 adopted a proposal for revision of this Directive. More information can be found on the website of the European Commission.

Organisation	The organisation and governance is arranged differently in each city.
Policy & Process	Only one of the cities has a clear, detailed and defined process for publication in place (Brugge). Others have a publication process, but not as detailed.
Licenses	Brugge has a IPR limitation on the platform, the other cities do not.
Platforms	GIS (Geo Information Server) is popular as a platform for publication. Brugge[44], Delft[45] and Saint Quentin[46] use GIS as a basis for publication.
Data	Geographical data is popular as type of data to publish. Cities state that this is the case because it is one of the most structured and standardized types of data they have. It takes therefor less time and effort to publish the data. Privacy considerations are also small with geographical data. Metadata is published differently. Brugge published a dataset with all the metadata, whereas Delft adds the metadata to the different datasets.
Standards	2 cities work with DCAT as the standard for metadata.
Formats	Cities publish open data in different formats. Brugge and Mechelen strive to publish open data in URI's, as stated in the Smart Flanders open data principles. Delft and Saint Quentin strive to publish data in minimum of CSV and KML.
Stimulate reuse	The three cities that publish open data, release them under a category like 'mobility', increasing the findability. Very little effort is put into stimulation of reuse. The biggest effort a city made is to organize a hackathon (Brugge). Because Brugge, Mechelen and Delft use GIS platforms as basis, there are (limited) visualisations of the datasets possible.
Monitoring reuse	There is no monitoring of reuse of data. That is, cities do not monitor how often a dataset is downloaded or used in another way. One of the cities monitors the amount of views (Delft). Only one of the cities offer ratings or direct feedback possibilities of end-users (Delft).

3.1 Approaches, policies and governance

Although not part of the questionnaire, city partners were asked during the workshop in Cambridge about their approach towards open data, the governance, the organisation and process of publication. The analyses of SCIFI partner cities shows that there are many differences in approach and operations regarding open data.

Brugge for example, has opened up more datasets than all other SCIFI partner cities and organises theme-based hackathons to stimulate reuse. The city doesn't have a clear local

[44] City of Brugge. Open Data Platform. URL: https://www.brugge.be/opendata
[45] City of Delft. Open Data Platform. URL: https://delft.dataplatform.nl/
[46] City of Saint Quentin. Open Data Platform. URL: http://open-data.saint-quentin-numerique.fr/

strategy for publication, but applies to the regional cooperation that supports in becoming a smarter city by providing knowledge and striving for standardization in order to stimulate reuse and replication (Smart Flanders). The Smart Flanders' approach to open data is that of By Default.

One of the Belgium cities stated that it takes a lot of steps in order to achieve that ambition: "They want to run, but we first need to walk." Mechelen does not have a policy in place there where Brugge has a detailed one. Brugge also released several datasets as open data and Mechelen yet needs to release the first one.

Delft on the other hand, choose the approach of By Design. They have a local strategy for open data, 'Open Delft' that is user-driven. The strategy however, does not contain a detailed process of publication like the one of Brugge.

3.2 Challenges for SCIFI cities

A number of challenges have been identified in general open data publication. They are outlined hereunder, based upon general findings op challenge at global level that relates to local level as well as upon the (similarities found in the) analysis of the maturity of partner cities in the SCIFI project.

The following challenges are identified by the SCIFI cities:

- Data gaps exist. Most of the SCIFI cities do not have a clear view over their data. They do not have in sight what data could be possibly opened up. This will require effort during the SCIFI project in order to execute the harmonised approach. On the other hand this offers a chance to identity the needs for tools and support in the process.

- Open licensing can be a problem. On the one hand, more governments implement their unique open data license versions. There is a need to draw more attention to data licenses and make sure data producers understand how to license data better. For SCIFI cities a similar license for reuse is necessary in order to stimulate reuse under the same conditions.

4. Recommendations for SCIFI cities

Based on this guidance package and analysis of the SCIFI cities open data maturity in chapters 2 and 3, recommendations can be given to the SCIFI partner cities for opening up data for the SCIFI project in order to strive for top-level harmonisation.

Principles of SCIFI for open data, cities will:

1. Follow the four phases as given in this open data guidance and use the tools as provided.
2. Use an open license for open data (Creative Commons zero if possible) to ensure reusability of the data.
3. Use DCAT as the standard for publishing metadata. If possible all cities provide their metadata both in their own language as well as in English.
4. Strive to publish data in machine readable format with a minimum of the 3 stars of the 5-stars open data model of Bernes Lee (CSV, JSON, XML).
5. Data of cities will first be published on the own platforms for cities and from those platforms will be taken and published. If not possible they will make sure the data is published directly on the SCIFI open data platform based (established by Fabourg Numerique).
6. Strive to publish data as timely as possible and if necessary in real time.

Besides these, the following recommendations are given:

- If cities work together in one challenge, it is recommended they analyse the datasets with similar data and strive to harmonize the datasets as much as possible for solution development.
- Any linguistic challenges in solution development will be overcome via the use of tools provided by Fabourg Numérique.

4.1 Learnings and recommendations after the first accelerator

The first round of pilots gave the SCIFI partner cities the opportunity to publish open data and test the proposed publication process. The learnings gathered during the accelerator are (briefly) discussed below and will be used to refine and enrich the process during and after the second round of pilots.

	Call 1	Call 2
City Open Data Sets[1]	33	18
City Not Yet Open Data Sets	16	10
Third Party Open Data Sets	25	13
Third Party Information	0	9
Third Party Not Yet Open Data Sets	2	1
Other (Sensor data to be collected/websites)	2	3
Total	78	54

Lesson and recommendation 1: sharing data versus open data
Some datasets were considered easier to share than to publish according to the proposed process. That led to the following questions: why is sharing easier than opening up data? Is sharing a natural first step to open up data? Or is there something that we are missing with opening up data that makes it easier to share?

When it comes to a pilot cities consider sharing data as a way to speed up the solution development. To some extent cities share data of which they are not (yet) entirely sure if to make the dataset open data. This does not mean that cities do not have the responsibility towards the data anymore.

Based on the experience of sharing data during the pilot phase and opening up the data at a later stage, the following recommendations can be made:

1. Sharing data during a pilot requires an adjustment of the (SCIFI) contract with business wherein one specifies how to deal with shared data (in terms of security, privacy and storage). It depends however on the country specific legislation to what extent these adjustments should be made.
2. Cities that share data for the pilot should go through the selection phase of the process at least before sharing the data for solution development. The answer to this phase is whether or not the data can be opened up in a later stage and does not have any legislative restrictions.
3. Cities should strive to share data in machine readable formats with a minimum of the 3 stars of the 5-stars open data model of Bernes Lee (CSV, JSON, XML).

Lessons and recommendation 2: basic strategy on IoT (data)
When it comes to creating sustainable solutions with open data, cities need to think beyond the pilot phase. During the pilots questions popped up about mostly the use of sensors, connectivity and sensor data as almost all cities experimented with sensors. Instead of struggling with these questions during the pilots and with the risk of opting for ad hoc solutions, cities are recommended to have a basic strategy on the use of IoT solutions (e.g. sensors) and IoT data in advance that should include at least:

1. How the city sees its role when it comes to sensors in development of smart city solutions (ownership, dataresponsibility, etc.)
2. Connectivity preferences (is the city open to any network or does it require one specific network, for example LoRa or SigFox?).
3. Privacy demands for IoT solutions (depending on the role the city sees for itself regards to ownership, etc.).

Cities are strongly recommended to create a policy or agreement on the maintenance of the hardware used during and after the piloting phase.

Lessons and recommendations 3: integration with existing infrastructure
Perhaps not solely a lesson learned for open data but more in terms of replicability of the solution: if a dump or selection of a dataset is used for solution development, it is necessary to understand the integration aspects to determine the usability, sustainability and replicability of the solution.

1. Cities are recommended to investigate integration aspects during the piloting phase to prevent any constraints or problems when the city wants to procure the solution. For example: if you share a dataset with static citizen reports, understand what it takes to share the citizen reports in real time if you want to implement the solution.

References

In this guidance package we referred to the following articles, projects, institutes and other sources. We also like to share where the reader can find more insights on open data.

Smart City (2017). Benefits of Open Data for Smart Cities. Retrieved from URL: https://hub.beesmart.city/solutions/benefits-of-open-data-for-smart-cities on 20-4-2018

Bennett, D. and Harvey, A. (2009). Publishing Open Government Data. Retrieved from URL: https://www.w3.org/TR/gov-data/ on 8-11-2018

Berners-Lee, T (2010). Linked Data. Retrieved from URL: http://www.w3.org/DesignIssues/LinkedData.html

City of Brugge. Open Data Platform. Retrieved from URL: https://www.brugge.be/opendata

City of Delft. Open Data Platform. URL: https://delft.dataplatform.nl/

City of Delft, Open Delft Strategy' (2016) Retrieved from URL: https://ris.delft.nl/document.php?m=1&fileid=295228&f=5ea3d4b26ab0a923489f30d8e356e336&attachment=0&c=61762 on 20-4-2018

City of Saint Quentin. Open Data Platform. URL:

Charalabidis, Y. et al. (2018) The World of Open Data, Public Administration and Information Technology 28.

Geiger, C. & Lucke, J. von (2012). Open Government and (Linked) (Open) (Government) (Data). JeDEM 4(2): 265-278. URL: https://jedem.org/index.php/jedem/article/viewFile/143/115

Foulonneau M, Turki S, Vidou G and Martin S. (2014) "Open data in Service design" Electronic Journal of e-Government Volume 12 Issue 2 2014, (pp99-107), Retrieved from URL: www.ejeg.com at 09-06-2018.

Global Open Data Index (2017) The State of Open Government Data in 2017. Retrieved from URL: https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf

Halonen, A. (2012). Being open about data: Analysis of the UK open data policies and applicability of open data. The Finnish Institute in London, London. Retrieved from URL: http://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf on 20-4-2018

Hartog, B. & Mulder, B. et al. (2014). Open Data within Governmental Organisations. JeDEM 6(1): 49-61. Retrieved from URL: https://jedem.org/index.php/jedem/article/view/291/269

Mulder. B. (2016). Presentation LOD workshop ESI HHS, eSociety Institute of The Hague Institute of Applied Sciences. Retrieved from URL: https://www.slideshare.net/eSocietyInstituut/presentation-lod-workshop-esi-hhs on 20-11-2018

Smart Flanders Open Data Charter. URL: https://smart.flanders.be/open-data-charter/

The European Data Portal. Licensing assistant. Retrieved from URL: https://www.europeandataportal.eu/en/content/show-license

The European Data Portal. The economic benefits of Open Data. Retrieved from URL: https://www.europeandataportal.eu/en/highlights/economic-benefits-open-data on 20-4-2018

The European Data Portal. How address privacy concerns when opening data? Retrieved from URL: https://www.europeandataportal.eu/en/highlights/how-address-privacy-concerns-when-opening-data

The European Data Portal. (2014) Presentation metadata: Open Data & Metadata Quality. Retrieved from URL: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_2.2_open_data_quality_en_edp.pdf

The Open Data Incubator. URL: https://opendataincubator.eu/

The Open Data Institute (3-11-2017). “What is open data and why should we care?” Retrieved from URL: https://theodi.org/article/what-is-open-data-and-why-should-we-care/ on 20-4-2018

Open Council Data. How to publish data step by step. Retrieved from URL: https://opencouncildata.org/how-to-publish-data/

Open Data Monitor (2015). Metadata standards. Retrieved from URL: http://knowhow.opendatamonitor.eu/odresearch/metadata-standards/

Open Knowledge Foundation. The Open Data Manual: Making personal data anonymous. Retrieved from URL: https://odm-test.readthedocs.io/en/latest/appendices/making-personal-data-anonymous.html

Open Knowledge Foundation. The Open Data Manual: What Legal (IP) Rights Are There in Data(bases). Retrieved from URL: https://odm-test.readthedocs.io/en/latest/appendices/what-legal-ip-rights-are-there-in-databases.html

Open Knowledge International (2017). The State of Open Government Data. Retrieved from URL: https://index.okfn.org/

Open Data Charter. Principles of Open Data. URL: https://opendatacharter.net/principles/

The World Bank. Open Data Toolkit. Retrieved from URL: http://opendatatoolkit.worldbank.org/en/starting.html on 3-3-2018

The World Bank, Data Technology Options. Retrieved from URL: http://opendatatoolkit.worldbank.org/en/technology.html on 3-3-2018.

TNO, Stappenplan Linked Open Data, URL: http://www.pilod.nl/wiki/BoekTNO/stappenplan

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly, 31*(1), 17–29 https://doi.org/10.1016/j.giq.2013.04.003

The Open Data Institute. “How to write a good open data policy". Retrieved from URL: https://theodi.org/article/how-to-write-a-good-open-data-policy/ on 20-4-2018

Appendixes

1. Mapping the city's data landscape
2. Open Data questionnaire SCIFI city partners
3. Intake form open data (example)
4. Are the data open and may I publish them as open data?

Appendix 1: Mapping the city's data landscape

The following list of questions can be used by cities in order to map their data landscape, including the relation between the data and applications and processes and the open data readiness.

Topic	**Mapping question**
About this mapping method	Contact person
	Was there an instruction for the contactperson to fill in this form?
General characteristics of the application	Name of the application
	Description of the application (registered?)
	Is data processed or kept in the application?
	Inventory of datasets
	Who is contact person for the dataset(s)?
	Comments on the application (for example: will be replaced by or is the same as…)
	What is the goal of the application?
	Who is supplier of the application?
	Who is producer of the application?
	Which department owns the application?
	Who is application manager?
	Who is functional application manager?
	Who is the contact person internally for the application?
	What is the position of the internal contact person for the application?
	What is the department of the internal contact person of the application?
	What is the mail address of the internal contact person of the application?
	Who uses the application? Name recognizable user groups.
	When is the next moment the contract with the supplier expires or can be adjusted?
	Does the application contain a Record Management functionality (RM)?
	Does the application offer the opportunity to close down a record, document or file in such a way that they cannot be adjusted anymore?
	Is logged when, what and by whom adjustments have been done (mutations and use)?
	Are roles and rights formally defined? Refere to / explain

Topic	Mapping question
	these rolls and rights.
General characteristics dataset(s)	Brief summary of the dataset
	Describe the most important data within the dataset.
	With what goal / goals is the data being collected?
	On which grounds are the data collected?
	Name the process/processes wherein the dataset is used.
	From when is the data being collected? Year + month.
	Frequency of collection.
	Name the source organisations (if relevant).
	Department responsible for the dataset.
	Who is the contact person for the dataset within the organisation?
	Position of the contact person within the organisation.
	E-mail address of the contactperson within the organisation.
Links - data in and out of the set	Is the data added manually and/or is the data partially or in total coming from another source (not manually)? If not manually: please specify the data and which sources the data comes from. Name the applications and / or services that take care of distributing the incoming data.
	Are the data in total or partially distributed to other applications, databases or services? Y/N.
	If Y, list the applications, databases or services.
	Name the applications or services that take care of the distribution of the outgoing data.
Data administration and –management	Are the data archived? Y/N
	If Y, where are the data archived?
	Are the data (after an xx period of time) destroyed?
	In the process of archiving and destroying: is a formal process in place? Refer to this process.
	Specify laws and regulations that relate to the use, processing, archiving and destroying terms of the dataset.
	Are there any other developments that are of influence on the processing and management of the dataset?
Openness of the data (related to e.g. privacy)	Does the dataset contain any personal related information or information that could lead to a person? (GDPR)
	If Y, list the exact data that relates to a person. Describe the nature and volume.
	Does the dataset contain process related personal information of the employees of the city?
	If Y, describe the nature and volume of these personal data.
	Does the dataset contain process related personal information of representatives of external parties / organisations?
	If Y, describe the nature and volume of these personal data.
	Are there any other judicial grounds to refuse publication of the data?
	If Y, describe.
	Conclusion publicity / openness of the dataset.
Intellectual Property Rights	Does any form of copyright apply on the dataset or any other database rights / rights of third parties?
	If Y, describe.
	Conclusion reusability.

Topic	Mapping question
	Are there ohter non formal grounds for not publishing the data?
	If Y, describe.
Internal reuse of data	Is it possible to reuse the dataset internally for another goal then where the dataset is collected for?
	Is there a policy or set of rules available on sharing these data internally?
	If Y, describe or link.
Open Data Readiness	Are the data publicly available?
	If Y, URL Open Data
	URL in another way then Open Data
	Are there any other cities or organisations that publish this kind of data as open data?
	If Y, URL / URI
	Are there, within the city (municipality and outside), concepts or applications that use these data? Are there, outside of the city, applications or concepts known that use these data. For example: the election app of OpenStateFoundation? URL (when yes)
Classification information security	Risk availability: what is the biggest damage thinkable caused by this applications disfunction at the most unconvenient time?

Appendix 2: Open data questionnaire SCIFI city partners

Each SCIFI partner city answered the following questions in order to determine the similarities and differences in open data. These questions are based upon a framework of Zuiderwijk, A., & Janssen, M. (2014).

Open Data Policy and Regulations

1. How does the city define open data?
2. Does the city have an “information strategy”? E.g. an overall strategy for information / data for the city that relates to open data? (please insert link if it is publicly available)
3. Does the city have an “open data policy” or “open data strategy” or a document that is used as such?
4. If there is a policy, since when does it apply?
5. What does the city hope to achieve with open data (as might be stated in the open data policy)?
6. Are there any other policies related with open data or the open data policies, such as smart city strategies. (please insert link if it is publicly available)
7. "Briefly describe what the city aims at achieving with the open data policy.
8. E.g. what are the underlying principles?"
9. Target group(s) for the open data are:
10. What communicative instruments about policy are being used?
11. Encouragement of data use, promotion?
12. Are there any fines and rewards as instruments used for this policy? (for example a sum of money for a developer after a hackathon)
13. Are there any metrics, e.g. indicators for output steering?
14. What national legal grounds are there for publishing data
15. Which (national) laws, except for the GDPR, restricts the city in publishing data?
16. Are there any restrictions regarding the use of open data?
17. Are there any regulations on Intellectual Property Rights that apply to your open data policy at the moment?

Process of publication

1. Is there a clear publication process within the city? (insert link or share if possible)
2. Which functions/ positions play a role in the process?
3. Who decides whether or not data is published? Is there one standard or does this differ? And why?
4. Does the city use any formats within the process? For example for an intake? If so, please mention which and upload in drive
5. "Which technical standards and formats are used for open data? Can the city indicate what level of the '5 star level' open data theory of Tim Berners-Lee it reaches? "
6. Provision of metadata: does the city use a metadata standard for publishing open data? If so, which one(s)?
7. In what languages does the city publish their metadata?
8. Does the city use a standard geographical coordinates system for their data? If so, which one? If not, name all used.
9. Does the city publish any real time open data?
10. Does the city publish their data straight from the source (another application) or does the city publish an extract (manually)?

11. Does the city have a clear process for processing data before publication when needed? (for example when a dataset needs to be anonymized)
12. Has the city published open data yet?
    If yes, how many datasets?
13. What types of datasets have you published? E.g. demographical, social, geographical?
14. Data available without registration?
15. What platform does the city use to publish their data? If more, please state why.
16. Way of presenting data: is it user need driven?
17. If using a platform, when was it launched?
18. Is the platform used based on open source? If so, please mention here.
19. Are there any proclaimers on the publication platform? What do they state?
20. Is there a privacy statement regarding open data available on the platform?
21. Technical support for the use of publicized data
22. Relationship data provider - user
23. Are there any insights on usages of publicized data?

Reflection on the current approach of partner cities

1. Do you have 'ambassadors for open data' in the organisation management and in the political arena?
   if yes; how have you managed to do so?
   if not; can you elaborate why?
2. At a scale from 1 - 10, where do you think your city stands with its open data approach and policy?
3. What has been the biggest challenge so far in the process of publishing open data?
4. How are you dealing with this challenge?
5. Are the open data sets being used as was anticipated at the time of publishing?
6. If not, can you identify underlying reasons?
7. Is the current policy framework that you have regarding open data sufficient?
8. if not, what aspects are missing?
9. Benefits of publicizing data: what was the biggest value created out of the approach? E.g. a concept or application. Please provide us with a summary.
10. Are there specific subjects or questions you would like to address in the guidance package? And why?

Appendix 3: Intake form open data (example)

INTAKE OPEN DATA

We do not publish open data just for the sake of publishing. To create value for the users of the data and in order to reduce risks with the data (political, technological, judicial, etc.), the first step is an intake.

This document contains… questions to be asked during the intake in order to find out whether or not the dataset(s) have a good chance of being published as open data and adding value. This intake takes place when:

- A direct question from a coworker may lead to publishing a dataset. For example because the data will be used in a specific project. Or because coworkers get signals from other organisations or inhabitants in the city regarding specific data.
- A direct question from the city council or the board of aldermen to publish a specific dataset. In this case the intake is not held with the demanding party but with the coworker of the city responsible for the dataset and/or another stakeholder from within the organisation.
- Demand from citizens or businesses asks for opening up data. In this case the intake is not held with the demanding party but with the coworker of the city responsible for the dataset and/or another stakeholder from within the organisation.

The intake is held by… a coworker from the city (for example an open data steward) that has a responsibility and certain mandate in the open data process. Involved should be: the owner of the relating process and/or the owner or manager of the dataset (the person entitled to decide). Depending on the trigger and the context, it is possible to invite also others. For example a policy advisor or manager.

Goal of the intake is… to determine whether or not open data:
(1) is a mean that contributes to the goals of the project or the development
(2) if there is a demand (in the city, by citizens, by organisations)
(3) if the data may be opened up
(4) what the status of the data is and how much effort it will take to publish

1. Introduction

1. Why are we here? Who is who?
2. What is open data, what could be the goals of open data? (management of expectations)

2. Frameworks, underlying goals and scouting possible usage of data

3. What data(set) are we talking about and within what context?
4. What do we collect the data for?

5. Are there any developments outside of the city's direct influence that make it possible to publish this data or are there parties that already publish the data (as open data)? E.g. a national organisation or national platform.
6. Can you estimate what would be the value that the city or target groups may derive from publishing the data?
7. Do you know if there was any form of contact with the target groups regarding the possible publication of the dataset(s)?
8. Are there any other organisations that publish similar datasets as open data? E.g. other cities.

3. About the dataset

9. Is the de data public (judicial) and who owns the data?
10. May the data, in case public, be published for reuse?
11. Where are the data processed? (Excell, application, etc.)
12. What is the status of the metadata?
    12.1 is there a clear description of the dataset?
    12.2: who is responsible / who is the datamanager?
    12.3: do you know how up-to-date the data is and what is the is the cycle of updating?
13. What can you say about the quality and completeness of the dataset?
14. Is processing of the data needed before publishing? Can you estimate what is needed?

Appendix 4: Are the data public and may I publish them as open data?

The following questions may help cities in knowing whether or not the data they want to publish is public data and if that data is publishable as open data. The questions are based on a Dutch example and thus may defer from the other member states.

The questions 1-11 refer to the question: are the data public?

1. Does the data contain personal data as mentioned in the General Data Protection Regulation (and national personal data regulation) or does it contain any data that might be related to a person?
2. Is there a possibility that publication of data might cause violation to the personal life of a person or group of people?
3. Does the data contain business- or manufacturing data that has been transmitted to the government based on confidentiality?
4. Is there a possibility that releasing this data might frustrate the process of detection and prosecution of illegal acts?
5. Is there a possibility that releasing this data might obstruct the inspection or supervision of public bodies / authorities?
6. Is it possible that releasing the data might lead to disproportional benefits or disadvantages parties?
7. Is there a possibility that releasing the data causes a threat for national security?
8. Is there a possibility that releasing the data causes threats for international relations with other states or international organisations?
9. Is there a possibility that releasing the data has a negative effect on economic or financial interests of the state or another public body?
10. Does the data contain personal views on policies of public servants?
11. Is it possible that any other special regulation applies to the data?

The questions 12-14 refer to the question: may the data, if public, be released as open data?

12. Is the copyright of the data owned by the city?
13. Does the city own the right on the database or datawarehouse?
14. If no in 12 and or 13: are there other license or agreements/contracts that effect the release of the data as open data?