

Towards a taxonomy for Business-to-Government data sharing

Serena Signorelli^{a,*}, Matteo Fontana^{a,b}, Michele Vespe^a, Lorenzo Gabrielli^a and Eleonora Bertoni^a

^aEuropean Commission, Joint Research Centre, Ispra, VA, Italy

^bRoyal Holloway, University of London, Egham, UK

Abstract. The phenomenon of Business-to-Government (B2G) data sharing represents a growing trend, especially in latest years. In fact, research has shown how privately held data could have a huge potential when used to tackle societal policy issues. B2G data sharing initiatives can be employed in different situations: from emergencies to the construction of official statistics and the use in research, just to name a few. In all these circumstances, the quality level required for the data may be different, as different principles could prevail upon others (e.g., timeliness in the case of emergencies is a key parameter). This heterogeneity in possible use-cases motivates the present work. In fact, our objective is to understand and classify the different contexts in which B2G data sharing may happen. The idea is to create a taxonomy of B2G data sharing initiatives, in which we identify all the different instances where B2G data sharing may occur. Afterwards we add as attributes some identified quality principles that characterise the different B2G data sharing situations. The work aims at providing further information that can help clarify specificities and requirements of B2G data sharing in order to enable relevant data flows and make them more dynamic.

Keywords: Business-to-Government, B2G, data sharing, taxonomy

1. Introduction

Business-to-Government (B2G) data sharing is commonly defined as a collaboration between a private company or organisation and the public sector (conceived at different levels: local, regional, national or supra-national), where the former makes available its data to the latter.¹ The aim of the sharing of data should be a public interest purpose, like, for example, the protection of the environment or the response to a public emergency such.

The last few years have seen this phenomenon growing, as research has shown how privately held data of different kinds could have a huge potential when used to tackle societal policy issues. Many examples of B2G data sharing initiatives can be found in literature, especially in the domain of production of official statistics,

and they're often based on one-off voluntary kind of cooperation agreements, more or less explicit, between the actors (private and public sector) (as an example, see [1]).

B2G data sharing is a concept that presents implications in legal terms and will have to be considered by legislators in future proposals that involve the sharing of data. Actually, as of the third decade of the 21st century, B2G data sharing started appearing in the legislative process of the European Union: one of the aims of the European Strategy for Data² is the adoption of “*legislative measures on data governance, access and reuse*”. Along this line, two Regulations have been adopted: the Data Governance Act,³ already applicable in the EU, and the Data Act,⁴ that will be applicable starting from September 2025. In particular, the Data Act con-

*Corresponding author: E-mail: Serena.SIGNORELLI@ec.europa.eu.

¹<https://digital-strategy.ec.europa.eu/en/faqs/business-government-data-sharing-questions-and-answers>.

²<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>.

³Regulation (EU) 2022/868 <http://data.europa.eu/eli/reg/2022/868/oj>.

⁴Regulation (EU) 2023/2854 <https://eur-lex.europa.eu/eli/reg/2023/2854>.

tains specific provisions concerning mandatory B2G data sharing in exceptional circumstances (for example in case of a public emergency) or when needed to implement a legal mandate, if the data availability is not guaranteed through other means. In those specific cases, private companies shall be asked to share data with the public sector to allow a quick and secure response to the public threat, but at the same time minimizing the burden on businesses.⁵

There exist different situations where B2G data sharing can give a valuable contribution in addressing public purposes: from emergencies (among the ones ruled by the Data Act), to the construction of official statistics indicators and to the use in research, just to name a few. Some examples have already been implemented; on the topic of emergency situations, the use of mobile network operators' data by the European Commission's Joint Research Centre to help fighting COVID-19 [2]. In the area of official statistics, the agreement between Eurostat and some collaborative economy platforms (Booking, Airbnb, Expedia and Tripadvisor) allows the statistical office of the European Union to produce some figures and data on short-stay accommodations across the EU.⁶ In all these circumstances, the agreements under which the data are shared may be of completely different kinds (going from one-off collaborations that give access only to a current view on the data, to periodic high frequency updates), depending on the willingness and openness of private companies to cooperate with the public sector. But different situations require different levels of the quality of data, depending on several factors; for example, in the case of an emergency situation, timeliness represents a key principle to guarantee data quality, while if the same dataset is used for research activities, the level of importance of the timeliness principle is not excessively prominent.

There are many elements that motivate our work; first, the heterogeneity concerning the quality principles that characterise all possible B2G data sharing situations. Secondly, as such kind of partnerships for the sharing of data is increasing (not only in the EU context), there exists a scientific interest in understanding better their features. Finally, in light of the provisions of the European Strategy for Data⁷ (and in particular

of the Data Act⁸) B2G data sharing initiatives will become more and more common, and we clearly see the need of an instrument that goes towards this direction. Our objective is to understand and classify the different circumstances in which B2G data sharing may occur and the characteristics that each of them should have. In practice, we aim at creating a taxonomy of B2G data sharing agreements (in all possible forms they may be established), in which we first identify all the different situations where B2G data sharing may occur, and then we add attributes (in the form of identified quality principles) that characterise the different B2G data sharing situations. The last step in this activity is the identification of B2G data sharing situations where the levels of the identified quality principles are the same, in order to reduce the cardinality of the taxonomy and to group together similar situations. As an supplementary step, we propose some additional elements in the form of an information set that we think is needed as a complement to every B2G data sharing situation. This work aims to provide further information that can help clarify specificities and requirements of B2G data sharing in order to ultimately stimulate and enable more dynamic and relevant data flows. The authors would like to stress that the taxonomy aims at being descriptive, trying to map all the possible kinds of B2G data sharing situations that are being implemented, and not at being seen as a prescriptive tool, nor at planning its operationalisation (that indeed remains an interesting topic to focus on, but out of the scope of the present work). Moreover, the authors would like to specify that they do not have a specific mandate to start the development of the taxonomy, but they act independently for scientific purposes.

The methodology that we followed to develop the taxonomy is the one shown in [3]: we started from the identification of the meta-characteristics (the classification of B2G data sharing situations and the characteristics of each of them), and then we identified the meta-dimensions of the taxonomy through a 'conceptual-to-empirical' approach. These meta-dimensions are structured on two levels (or layers): the first layer contains *themes*, *patterns* and *attributes*, while the latter dimension is in turn characterised by three meta-dimensions (the second layer), namely *spatial and temporal*, *methodological*, and *legal and governance*. Section 2 contains a detailed explanation of all these elements. In Section 3 the three identified meta-dimensions

⁵https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113.

⁶https://ec.europa.eu/commission/presscorner/detail/en/ip_20_194.

⁷<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>.

⁸Regulation (EU) 2023/2854 <https://eur-lex.europa.eu/eli/reg/2023/2854>.

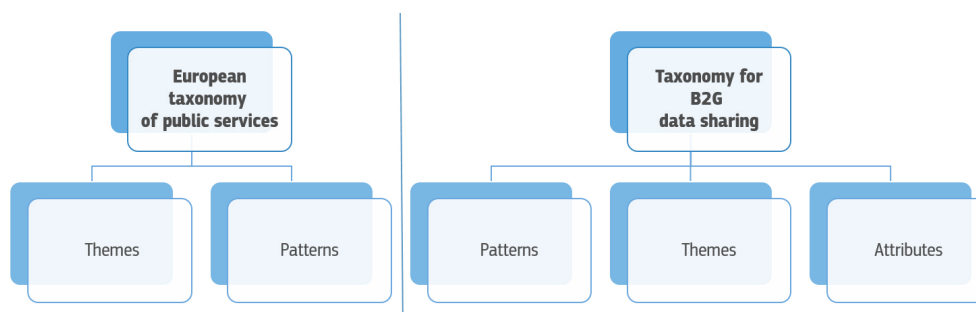


Fig. 1. Elements of the EU taxonomy of public services compared to the elements of the proposed taxonomy. Source: own elaboration.

are put together to concretely build the taxonomy, while Section 4 contains the proposal of an information set that should ideally accompany every B2G data sharing initiative.

2. Elements of the proposed taxonomy

As mentioned in Section 1, our proposed taxonomy first aims at identifying all the situations when B2G data sharing can occur, and, as a second step, defining the quality attributes that are needed in each situation with the final goal to cluster different instances of B2G data sharing into subgroups.

The identification of all possible B2G data sharing situations builds upon the building blocks of a taxonomy [4] developed by the European Commission's Directorate General for Informatics (DIGIT) and that aims at mapping all European public services through two elements: *themes* (also referred to as thematic areas, like 'Defence', 'Education', 'Health care', just to name a few) and *patterns* (that are public service types, or the core services of governments, like for example 'Control and monitoring'). As visualised in Fig. 1, these two elements are adapted to our work's specific needs and become the first two building blocks (meta-dimensions) of our proposed taxonomy. A third building block is then added – the so called *attributes* –, composed of a selected subset of quality principles that characterise the different B2G data sharing situations.

2.1. The European taxonomy of public services

The European taxonomy of public services [4] was published in 2019 by the European Commission's Directorate General for Informatics (DIGIT) to help public administrations in harmonising their catalogue of services. In particular, one of its goals is to make it easy

for users to find public services, as well as to compare them across different Member States.

The EU taxonomy of public services is built as the combination of two elements, *themes* and *patterns*, that together create a 'high-level generic public service' [4]. Public services can be allocated to one *theme* and one *pattern* only. The following two subsections deep dive into the details of those two elements and the choices we made for the proposed taxonomy.

2.1.1. Themes

Themes are the first element of the EU taxonomy of public services [4] and can also be referred to as thematic areas. To understand the meaning of *themes* of the EU taxonomy of public services, DIGIT created a visual explanation available in Fig. 2.

The list of themes from the EU taxonomy of public services is composed by 31 elements; examples of themes are 'Defence', 'Education', and 'Environmental', just to name a few.

For our work, we decided to consider the full list of themes as they appear in the EU taxonomy of public services because they cover all thematic situations in which B2G data sharing could occur (even if some cases could appear more theoretical than others – see for example 'Religious'). The complete list of themes is available in Table 1.

The EU taxonomy of public services does not foresee any sub-theme, as its authors state that "*Some sub-themes were identified, but this was done on an ad hoc basis and should be further refined.*" [4, Page 42]. For the construction of our taxonomy, this first level of themes appears to be enough, but it could be worth considering possible future extensions of them. For example, a proposal of a possible sub-theme for 'Manufacturing' is the sub-classification of industrial ecosystems foreseen in the European Industrial Strategy,⁹

⁹https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-industrial-strategy_en.

Table 1
List of themes from the EU taxonomy of public services

Themes		
Agriculture and food	Animal	Border control
Culture, sport and leisure	Defence	Digital
Education	Emergency	Environmental
Family	General business	General government
Health care	Housing and building	Legal
Life event and identity	Manufacturing	Media
Monetary policy	Money and debt	Natural resources
Public space management and heritage	Religious	Retail
Stock market	Tourism and travelling	Transportation and transportation infrastructure
Utilities	Voluntary organisation and charity	Welfare and social care
Work		

Source: [4].

What is a theme?	
Public service description	A university student grant is a public service that allocates grants to university students to increase opportunities for students with less financial power.
Filtering	If we remove the sub user role (student), the core/subpattern (grant) and the goal of the service, the only thing that remains is a 'university'.
The theme	This is a detailed service (subtheme) of a generic public service theme, education.
Good to know	A theme can be used with any pattern. A public service (sometimes vaguely defined by the public administration) can contain multiple themes.

Fig. 2. Explanation of a theme. Source: [4].

that contains 14 industrial ecosystems that span from 'Aerospace and defence' to 'Agri-food' and 'Construction', just to name a few.

2.1.2. Patterns

The second element of the EU taxonomy of public services [4] is composed by *patterns*, that are public service types, or the core services of governments. When public services are broken down to their core by removing any context, the detail of a core service remains. The overarching concept of all these details is defined as a pattern.

As done for themes, to understand the meaning of patterns in the EU taxonomy of public services, DIGIT created a visual explanation available in Fig. 3.

The EU taxonomy of public services is conceived with the following nine patterns: 'Framework', 'In-

formation', 'Registration', 'Certification', 'Financing', 'Production', 'Feedback', 'Control & monitoring', 'Taxation', and each of them is proposed with a list of sub-patterns [4] that can help understanding better the various cases (see Table 2 for the complete list).

Starting from our own considerations and reflections on B2G data sharing, we imagined the following four patterns as fit for our purpose: 'Control & monitoring', 'Official statistics', 'Planning & management', 'Research'.

The reasoning behind these choices relies on the answer to the question '*for what public purpose aims could a public entity use privately held data?*'. We think that a public entity may need some data to plan in advance some interventions (like the impact of a new policy) but also to check their impact afterwards; addition-

Table 2
List of patterns from the EU taxonomy of public services

Patterns	Sub-patterns						
<i>Framework</i>	Procedures	Measures	Law definition	Management (of a bureaucratic structure)	Asset management	Collective infrastructure	Schemes and plans
<i>Information Registration</i>	Information	Advice					
<i>Certification</i>	Registration						
<i>Financing</i>	Certification	Licensing	Permission	Authorisation			
	Financial support	Material support	Provision of free services	Provision of discounted services			
<i>Production</i>	Governmental service at normal price	Maintenance of property	Maintenance of infrastructure	Provision of infrastructure or a charge	Waste management		
<i>Feedback</i>	General complaints	Injury/damage complaints	Feedback	Appeals against decisions	Mediation		
<i>Control and monitoring</i>	Control	Monitoring	Testing	Assessment	Law enforcement		
<i>Taxation</i>	Taxation						

Source: [4].

What is a pattern?	
Public service description	A university student grant is a public service that allocates grants to university students to increase opportunities for students with less financial power.
Filtering	If we remove the sub user role (student), the subtheme and the goal of the service, the only thing that remains is a 'grant'.
The theme	This is a detailed service (subpattern) of a generic public service pattern, financing.
Good to know	A pattern can be used with any theme. A public service (sometimes vaguely defined by the public administration) can contain multiple patterns.

Fig. 3. Explanation of a pattern. Source: [4].

ally, a public entity may need data to be able to produce official statistics at different levels of detail, or to carry out some research activities (that may afterwards lead to the other three patterns). The EU taxonomy of public services does not supply a definition for each pattern, but it rather explains them through a series of sub-patterns. The analysis of those sub-patterns helped us find a correspondence between our own ideas of patterns and those already present in DIGIT's taxonomy.

'Control & monitoring' is a pattern already defined in the EU taxonomy of public services, and the sub-patterns that explain its meaning are *control, monitoring, testing, assessment* and *law enforcement*; it refers

to all those activities that a public entity may carry out to perform an ex-post monitoring check.

For the other three patterns that we envisioned we need a deeper investigation on the EU taxonomy of public services' sub-patterns to check whether some of the patterns have already been identified and included in DIGIT's taxonomy. First, 'Planning & management' could be considered as part of the 'Framework' pattern, as the sub-patterns identified in DIGIT's taxonomy are *procedures, measures, law definition, management (of a bureaucratic structure), asset management, collective infrastructure* and *schemes and plans*. This pattern relates to a public service type which activities are carried out ex-ante (in a planning phase). Second, 'Official

statistics' may be included into the pattern 'Information', as its sub-patterns are *information* and *advice*. It can be seen as a public service type that supports the other ones, both ex-post and ex-ante.

Finally, 'Research' is the only pattern that is completely brand new and that we have decided to add, as a match in DIGIT's taxonomy could not be found. It relates to experimental activities that could be carried out in support of the other three patterns, but as a preliminary step that could afterwards lead to the establishment of a consolidated public service.

The final list of patterns of our taxonomy is then the following:

1. 'Control & monitoring', that includes all possible monitoring activities to be carried out ex-post.
2. 'Information', that includes official statistics and, more in general, all those activities aimed to provide information and advice to public entities (but not under an experimental framework, that is indeed covered under 'Research').
3. 'Framework', that includes all activities that may help public entities in the planning phase of their day-by-day work (ex-ante).
4. 'Research', where the experimental activities (before becoming structural) are included.

To exemplify the identified patterns, we may take one specific source of data to showcase the different situations: mobile network operators data. In fact, they may be used to plan the introduction of limitations on the number of tourists in cities (ex-ante, the pattern in this case is 'Framework'), or after the introduction of these measures to monitor how effective they may be (ex-post, pattern 'Control & Monitoring'). But the same source of data may also be used by National Statistical Institutes as an alternative to census data to identify the population on a territory in a faster and less costly way (pattern 'Information'), but also as an attempt to study a phenomenon during an emergency situation (as it was done by the Joint Research Centre of the European Commission during the COVID-19 pandemic [2] – pattern 'Research').

For each situation, in principle only one pattern applies (the same should be valid also for themes), but we acknowledge the fact that some B2G data sharing agreements may cover more than one theme and pattern. In such cases, the agreement should be split into multiple parts, in such a way that each of the single parts can be categorised under one theme and one pattern only.

2.2. Attributes

After having tailored the themes and patterns from the EU taxonomy of public services [4] to our needs,

we add a third meta-dimension to our taxonomy for B2G data sharing: *attributes*. The first two elements have set the scene for all possible B2G data sharing situations, identifying their topics (*themes*) and purposes (*patterns*), and we now need to list all the characteristics (*attributes*) that such data sharing initiatives could present.

The approach we follow to set up attributes is a top-down one (and conceptual-to-empirical) [3]: we identify three main categories (or meta-dimensions) that it is worth covering (spatial and temporal, methodological, legal and governance), and we then enrich each of them with some selected attributes.

Some sets of principles that cover a few of these identified dimensions are available in literature, but they are not specifically targeted at B2G data sharing. As an example, consider the United Nations (UN) quality principles of official statistics [5], which are composed by six elements: relevance, accuracy, timeliness, accessibility, interpretability and coherence. Similar approaches have been followed by other statistical organisations (UNECE and OECD, among others), adding a few components (like reliability and clarity) to the UN set of principles (for a detailed review, see [6]). Additional examples of principles that could be applied are the FAIR [7] and/or the CARE [8] ones, but as already anticipated, they do not cover the whole set of aspects that we envisioned.

Some transversal approaches appear in literature, like an attempt to map some quality principles traditionally coming from the statistical domain (e.g., timeliness, completeness, etc.) over the four FAIR dimensions [9]; in that case the aim of the work was to provide a set of quality principles on the quality of data in general, not specifically in the context of B2G data sharing.

To our knowledge, the only example that was developed with the specific aim to identify principles for B2G data sharing is contained in [10]. The report is the result of a working group named "Facilitating the use of new data sources for official statistics" set up by Eurostat, the statistical office of the European Union. Its work was specifically targeted at the use of privately held data for official statistics, and in Chapter 3 of the report a list of general and specific principles for the use of this kind of data is suggested.

For our proposed taxonomy, we decided to take profit from some existing principles and to adapt them to our own needs, adding some brand new ones that were missing. According to the three broad dimensions previously introduced, we identified a list of 12 principles that are available in Table 3. The idea behind this list is

to assign a level of importance (high, medium or low) to each of the attributes. In the next subsections we dig into each of the broad dimensions and the single attributes' characteristics.

2.2.1. Spatial and temporal dimension

The first dimension we identified is related to the space and time references of the data. Depending on the specific situation, these attributes could present more or less relevance; indeed, as an example, after an earthquake (emergency) the spatial coverage needed in terms of data may be very limited (namely the specific area where the disaster happened), differently from a B2G data sharing partnership established to produce official statistics at the EU level, where the spatial dimension should cover all the 27 Member States.

We identified the following four attributes that relate to the spatial and temporal dimensions:

1. Timeliness;
2. Continuity over time;
3. Availability of time series;
4. Coherence and consistency over time and space.

We dig into their details in the following paragraphs.

2.2.1.1. Timeliness

This principle has been specifically defined in the official statistics domain; in particular, [5] defines it as “[...] the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available [...]”.

It represents a very important principle when it comes to the need of having access to data in nearly real-time. But its importance has not the same relevance in every B2G data sharing situation; for example, if the aim of the data sharing agreement is research, receiving the data with a bit of delay does not represent a big issue. Indeed, when monitoring a phenomenon with high time frequency (for example, something that happens weekly or even daily), receiving data promptly is key.

2.2.1.2. Continuity over time

Still related to the time dimension, continuity in a B2G data sharing situation represents the ability to guarantee that the sharing of data continues over time, without any break. This feature proves fundamental in the long term for some circumstances, like the monitoring of a phenomenon over time, while in other situations continuity is not perceived as that important (think, for example, to the need of having a one shot picture at a specific point in time of a specific situation). The lack

of continuity over time, in specific B2G data sharing situations where this principle has high importance, may lead not only to discontinuity in monitoring activities, but also to possible interruptions, as it is often complicated to find a substitute for a specific dataset when dealing with privately held data (in many cases this proves nearly impossible).

2.2.1.3. Availability of time series

Depending on the pattern the specific B2G data sharing falls into, the length of time series that is needed may vary. A phenomenon could potentially be analysed over a short or a long period; for example, some recurring patterns may need to be identified over a week while others happen over several months or years. Moreover, the public entity could aim not just at a one-off analysis of the phenomenon, but also at its comparison at different points in time; in this case, an extended temporal coverage would be needed.

The availability of time series principle could be seen as very similar to continuity over time, with the difference that the former specifically focuses on the past, while the latter focuses on the future. The two principles do not necessarily go hand in hand in different B2G data sharing situations. For example, in the context of a research project the availability of long time series may be important to derive conclusions on a specific issue, but the guarantee of having the data available over time (continuity) may have less relevance. On the contrary, in order to monitor a specific situation, continuity of the data flow in the future may be key, but the availability of long time series from the past could be not necessary at all (for example, in the case of the monitoring of efficiency of a newly implemented service, where no historical time series are available).

2.2.1.4. Coherence and consistency over time and space

The concepts of coherence and consistency allow us to make a step further from pure B2G data sharing, as they both relate to the possibility of combining together different data sources. This combination of different datasets may involve traditional data as well as innovative sources (among which we usually find privately held data).

The concept of coherence “[...] reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time [...]” [5], while consistency relates to the principle that data do not have to contain contradictions [9].

Table 3
List of attributes of the proposed taxonomy

Meta-dimensions	Attributes
<i>Spatial and temporal</i>	Timeliness Continuity over time Availability of time series Coherence and consistency over time and space
<i>Methodological</i>	Accuracy Transparency Interpretability
<i>Legal and governance</i>	Coherence and consistency over data provider Degree of data reusability Distribution of data products derived from the original data Presence of a governance framework Involvement of data subjects and data holders

Source: own elaboration.

These two concepts must apply both over time (on different periods, for example on two distinct years) as well as over space (for example on different Member States). The specific situation and aim of the analysis that needs to be carried out determine the level of importance of these two principles, that may be high in case the analysis covers different geographic areas (eventually with data provided by different data providers) compared to an analysis focused on a specific moment in time and a specific limited geographic area.

2.2.2. Methodological dimension

The second dimension that we identified is related to some methodological aspects that must be considered in B2G data sharing, and it contains the following three principles:

1. Accuracy;
2. Transparency;
3. Interpretability.

Both accuracy and interpretability represent attributes taken from quality principles of official statistics, complemented by transparency, a principle more targeted to the peculiarities of privately held sources of data.

2.2.2.1. Accuracy

Citing again the work of Brackstone on quality principles in official statistics [5], accuracy can be defined as “[...] the degree to which the information correctly describes the phenomena it was designed to measure [...]”. Accuracy has always been cited together with timeliness, in a debate that lasts since years (the so-called *timeliness vs accuracy trade-off*) [11]. It is undoubtedly true that in every situation, included B2G data sharing, high levels of accuracy are something to aim at, often putting aside timeliness; but it is also true that in peculiar situations, like for example emergency ones, it is the timeliness principle that prevails to the detriment of accuracy.

2.2.2.2. Transparency

This principle has been added to the list of attributes with the specificities of B2G data sharing situations in mind. Transparency would theoretically invest high importance in any data sharing situation (not only B2G ones), to allow many stakeholders (including citizens) to get informed about how data for public purposes are acquired and used. Moreover, this goes alongside the path towards open source that appears since many years in various fields, including the statistical domain [12]. But some confidentiality issues as well as the reluctance to share contractual details by private companies (when they are not legally binded to do so) has created a number of situations where transparency is far from being really implemented. Depending on the specific B2G data sharing situation, the importance of the transparency principle may vary; think for example at the use of privately held data to monitor (pattern) taxation (theme): in this case citizens would be very interested in the terms and conditions under which this activity takes place. On the contrary, in other situations (like those categorised under the ‘Emergency’ theme) the focus on transparency lays into the background, leaving the main stage to other more prominent attributes.

2.2.2.3. Interpretability

The other attribute that has been adapted from principles of official statistics is interpretability, that according to Brackstone in [5] “[...] reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately [...]”.

Depending on the theme and/or pattern of the specific B2G data sharing situation, this principle may have a higher or lower level of importance, for different reasons. In fact, if we consider activities under the ‘Research’ pattern, the level of interpretability required to understand the data may be lower, as these data are

handled by highly skilled individuals used to deal with non traditional sources of data and specifically trained to interpret very complex data. Under the ‘Emergency’ theme, instead, this attribute may require a low level of importance as well, but for a different reason: the public body that requires those data will need to act fast and the provision of supplementary information and metadata may take away precious time from the main critical task.

2.2.3. Legal and governance dimension

When dealing with innovative sources of data that were originally collected for one specific purpose and then reconverted to another one, legal issues must be considered, as well as aspects more related to the practical governance of the databases and of the subjects whose data are participating in this re-purposing process. In detail, the following five principles have been identified to cover these issues:

1. Coherence and consistency over data provider;
2. Degree of data reusability;
3. Distribution of data products derived from the original data;
4. Presence of a data governance framework;
5. Involvement of data subjects and data holders.

The following paragraphs will dig into the details of each attribute.

2.2.3.1. Coherence and consistency over data provider

We already mentioned coherence and consistency under the spatial and temporal dimension (coherence “[...] reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time [...]” [5], while consistency relates to the principle that data do not have to contain contradictions [9]), but in the specific case of B2G data sharing situations it is worth highlighting the importance of these attributes in terms of the data provider. It may happen that one single provider of data is not capable of covering a whole market (as an example, the mobile network market is composed by a number of individual operators), so, in order to achieve full coverage, many providers need to be involved, and the data need to be merged into one single dataset. In this case, the above-mentioned concepts of coherence and consistency apply to each of the data providers that are involved.

Depending on the specific B2G data sharing situation, the importance of these two principles may vary; for example, consider a research project that will serve as a pilot to check the feasibility of the use of novel

data sources. In this case, the pilot may involve just one single data provider, with a low level of importance on coherence and consistency. On the contrary, a control and monitoring activity (pattern) carried out on the topic of taxation (theme) through privately held data would require a high level of importance of coherence and consistency over the different data providers in order for this task to be effective and fair.

2.2.3.2. Degree of data reusability

When accessing privately held data, data reuse is always an issue that takes the spot, especially in the research realm where the access to these sources of data is often hard to achieve and researchers are always looking for new datasets. In some particular B2G data sharing situations, the possibility of reusing the data is very important, giving the possibility to others to explore the same dataset for similar purposes. At the same time, data providers may be reluctant to sign such permissive agreements, given the sensitive content of their data in terms of trade secrets, but also given the increasing commercial importance of the sale of datasets; moreover, permissive agreements may be way too expensive compared to less permissive ones, implicitly leading public administration to give up on the possibility of data reuse. In other cases public administrations requiring data may consider the degree of data reusability as less important like, for example, in cases where privately held data are used to monitor activities for taxation purposes.

2.2.3.3. Distribution of data products derived from the original data

This principle is linked to the degree of data reusability, as it is among the contractual clauses that usually are proposed to data providers during data acquisition processes. If data reusability implies the possibility of reusing an acquired dataset by other users in (usually) the same public entity, the distribution of data products derived from the acquired data implies the possibility of publishing tools that the public administration has created using that specific procured dataset. These products may be simple charts or visuals, but they may evolve as well into more advanced tools like interactive dashboards and/or aggregated datasets. As seen for data reusability, openness on this attribute could prove costly in financial terms, and could be hindered by data providers to protect trade secrets or their business interests. Again, in order to push towards the need for open science, B2G data sharing under the ‘Research’ pattern may consider this attribute as highly important, while activities under the ‘Control & monitoring’ pattern may consider it as less important.

2.2.3.4. Presence of a data governance framework

As Abrahams et al. [13] state, data governance “*specifies a cross-functional framework for managing data as a strategic enterprise asset*”. It is not compulsory for entities at the moment to implement such a framework, but it represents a good practice that it is advisable to have. Depending on the type of B2G data sharing situation, the need for such a framework may invest more or less importance, considering the previously introduced attributes as well. In fact, when a data provider agrees on the possibility of reusing the data, the importance of having a data governance framework becomes key to ensure that the rules surrounding the use of data are respected. The importance of this attribute may be high also in cases where the B2G data sharing agreement implies a continuous and periodic data flow, while in other situations (as in emergency cases) other attributes may be more important (like, for example, timeliness) and the lack of a data governance framework does not represent a blocking feature.

2.2.3.5. Involvement of data subjects and data holders

The last attribute that we identified concerns the involvement of data subjects (as defined in Article 4, point (1) of GDPR¹⁰) and data holders (as defined in Article 2, point (8) of the Data Governance Act¹¹) in B2G data sharing situations. This requirement may appear particularly important in some cases, where it is required that data subjects and data holders are aware of what it is done with the data that pertain to them (for example when the public entity performs a monitoring activity in the taxation sphere), while for other more exploratory situations (like research activities) or during emergency ones there is less need or time to get them involved.

3. The structure of the taxonomy

After an overview of the elements that are the core parts of the taxonomy of B2G data sharing, it is time to merge them. The taxonomy is a combination of the identified themes and patterns, and users can classify B2G data sharing agreements combining one theme with one pattern (for example, ‘Control & monitoring’ – pattern – and ‘Health care’ – theme). This way, the first layer of the taxonomy is composed by 124 dif-

ferent situations, that is the result of the four patterns multiplied by the 31 themes.

Each single B2G data sharing initiative has to be characterised by the 12 identified attributes – the second layer of the taxonomy – by assigning different levels of importance to each of them. As an example, in an emergency situation (pattern ‘Control & monitoring’, theme ‘Emergency’) timeliness would constitute one of the most important features for the data, while continuity over time could be considered marginal. The latter, by contrast, could result in being fundamental in other situations (like in all those that fall under the ‘Information’ pattern – to be understood as official statistics). Figure 4 shows how the two layers are combined in a specific B2G data sharing situation, namely research (pattern) in education (theme).

The next step of this work would be to assign a level of importance (high, medium, low) to each of the attributes in the complete set of identified B2G data sharing situations. After this activity is completed, some similarities will appear in different combinations of patterns and themes; this will allow to group some of the B2G data sharing situations and will constitute the final step of development of the taxonomy.

4. Information set for B2G data sharing

While fine-tuning the list of attributes, we identified some essential information that accurately characterises the specific B2G data sharing situations. In the context of B2G data sharing initiatives, we recognized the critical need for accompanying the agreements with some details. As a result, we propose an information set to enhance B2G data sharing; it aims to provide comprehensive context information and to facilitate effective communication between private companies and public entities.

To the best of our knowledge, at the moment there exists no specific archive or inventory of available B2G data sharing initiatives; a couple of tools may be considered as a sort of inventory/repository for B2G data sharing initiatives to some extent, but their primary aim is broader than this. The first one is the Tenders Electronic Daily (TED),¹² the online version of the ‘Supplement to the Official Journal of the EU’ that contains data about public procurement in the EU (above a certain threshold). This database does not specifically address

¹⁰Regulation (EU) 2016/679 <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

¹¹Regulation (EU) 2022/868 <https://eur-lex.europa.eu/eli/reg/2022/868/oj>.

¹²<https://ted.europa.eu/en/>.

THEMES / PATTERNS	Control & monitoring	Information	Framework	Research
Agriculture & food				
Animal				
Border control				
Culture, sport & leisure				
Defence				
Digital				
Education				
Emergency				
Environmental				
Family				
General business				
General government				
Health care				
Housing & building				
Legal				
Life event & identity				
Manufacturing				
Media				
Monetary policy				
Money and debt				
Natural resources				
Public space management & heritage				
Religious				
Retail				
Stock market				
Tourism & travelling				
Transportation & transportation infrastructure				
Utilities				
Voluntary organisation & charity				
Welfare & social care				
Work				

META-DIMENSION	ATTRIBUTE	LEVEL OF IMPORTANCE
Spatial and temporal	Timeliness	
	Continuity over time	
	Availability of time series	
	Coherence and consistency over time and space	
Methodological	Accuracy	
	Transparency	
	Interoperability	
Legal and governance	Coherence and consistency over data provider	
	Degree of data reusability	
	Distribution of data products derived from the original data	
	Presence of a governance framework	
	Involvement of data subjects and data holders	

Fig. 4. Combination of the first and second layer of our proposed taxonomy. Source: own elaboration.

B2G data sharing initiatives, as it is established to increase transparency in public procurement, but it may contain examples of data acquisition procedures from private companies to public entities. The second example is the Data Collaboratives Explorer¹³ created and maintained by The GovLab;¹⁴ data collaboratives are defined as a new form of collaboration that goes beyond the public-private partnership model, and the explorer contains a set of examples where public value is generated by exchanging data in different sectors, categorising initiatives as data cooperatives or pools, prizes and challenges, research partnerships, intelligence products, Application Programming Interfaces (APIs) and trusted intermediaries.

As these examples are not specifically targeted at B2G data sharing initiatives and due to the limited amount of information contained (as those tools were created with different purposes in mind), we suggest

the following information set to be provided for B2G data sharing initiatives in possible future inventories/repositories. We identified the following list of features to be part of this basic information set:

- Spatial coverage;
- Temporal coverage;
- Population coverage;
- Type of data;
- Partnership scheme;
- Control over the use of data;
- IT infrastructure used;
- Metadata on high importance attributes.

The spatial and temporal coverage information to be provided depends on the type of analysis carried out and on the public entity that performs it.

Indeed, if for example a regional tourism office would like to analyse the flows of foreign tourists within its borders over a summer period through mobile phone data, the spatial coverage would be limited to its territory, and the temporal coverage to some months over the year, possibly comparing the same time frame with

¹³<https://datacollaboratives.org/explorer.html>.

¹⁴<https://thegovlab.org>.

previous years. In case the same analysis would be carried out by a National Statistical Institute, while the temporal coverage could remain the same, the spatial coverage of the dataset should refer to the whole country, not only to one specific region.

Another information to be specified is the statistical unit that is subject to analysis (namely, the population coverage). Considering the above-mentioned example, if the analysis of tourism flows is limited to foreign tourists, the statistical units considered would only be foreign SIM cards whose presence was registered in the region/country during an identified time frame. At the same time, in case the analysis should cover all tourism flows (national and foreign), all SIM cards registered at a given moment in the region/country should be included, without discriminating on their country of release.

A fundamental feature that should be part of the information set are some details on the type of data; without delving into the specific details that a proper codebook should contain, for the purpose of the information set it would be sufficient to indicate the broad typology of the data part of the B2G data sharing agreement (for example, mobility data, consumer prices and behaviours data, energy consumption data, etc.) and whether the B2G data sharing involves the actual data, or only an anonymised subset or a synthetic version of the full dataset.

A detail that is specifically targeted at B2G data sharing situations is the type of partnership scheme that has been agreed between the private company and the public entity; on this specific issue, different kind of collaborations have been identified in literature. For example, the Publications Office of the European Union [14] identifies five different types of collaboration: multi-party data sharing agreement, data donorship, data partnerships, data intermediaries and data sharing by regulation. Another proposal of models for B2G data sharing comes from the European Commission [15]: data donorship, prizes, B2G data partnerships, intermediaries and 'civic data sharing'. Finally, Micheli [16] identifies a series of models for B2G data sharing specifically targeted at cities: data donorship, public procurement of data, data partnerships and pools, and data sharing obligations. These examples show how literature still not completely agrees on well defined types of collaborations that may be established, but even if the topic is evolving, it would be important for B2G data sharing situations to be accompanied by this kind of information.

Concerning the control over the use of data, it is essential to specify in the information set whether the

data provider has retained some control over the data that are part of the B2G data sharing agreement or not; this aspect is important to assess the independence over the use of the data.

Moreover, the information set should contain details about the IT infrastructure used; in fact, the data may be processed at the provider's premises (without physically moving the data) or the datasets may be shared with the public entity and processed on its own infrastructure. Depending on the choices made on whether the data are moved or not, some implications may arise in order to ensure reproducibility and auditability of the analytical work, and also in terms of security and privacy.

Last but not least, concerning the attributes of the taxonomy that present 'high' importance rankings, the information set should include the metadata referring to those specific attributes. For example, a B2G data sharing agreement that presents a high importance level of the degree of data reusability attribute should already include its contractual terms in the information set, in a way to reiterate the importance of the attribute.

5. Conclusions

This paper presents a first proposal of a taxonomy for B2G data sharing, therefore it is expected to be further consolidated and integrated with new insights, especially in the 'attributes' part. Moreover, after the structure of the taxonomy will be consolidated, the following step of the work will be to identify the levels of importance of each attribute in the different B2G data sharing situations, in order to group cases where situations are different, but the quality principles required are similar.

The paper aims at providing a basic information set that should accompany any B2G data sharing situation, to start setting the scene for possible future inventories or archives of this type of initiatives (something that at the moment is not included in the two broad inventories mentioned in Section 4). This work aims to provide insights on the differences of B2G data sharing settings, commonly grouped as a single instance of data flows whereas they exhibit a broad and relatively diverse range of goals, contexts and therefore requirements.

Before proceeding with the next steps, the authors would like to get feedback from the scientific community concerning the validity of the approach and the proposed structure of the taxonomy, in order to refine and readjust some elements that may appear weak. Moreover, specific quantitative techniques could be ap-

plied to determine the best way to rank the identified attributes, like the Analytical Hierarchy Process (AHP) [17,18]. This ranking process can also be extended to involve a panel of experts, using, for example, DELPHI surveys [19]. Next steps would also include the testing of the validity of the taxonomy on existing B2G data sharing inventories/repositories (see Section 4). A possible solution for this could be the extraction of a list of data acquisitions from the Supplement to the Official Journal of the European Union (TED)¹⁵ database, that includes all EU public tenders above specific contract values and it is aimed at increasing transparency in public procurement.

Finally, as the paper presents a proposal of a taxonomy for B2G data sharing, the next activity could also step towards the proposal of a taxonomy for Government-to-Government (G2G) and Government-to-Business (G2B) data sharing that could be interoperable with the one proposed here – once finalised. While the interoperability of the two taxonomies would bring great value, a G2G data sharing taxonomy may be something difficult to create, as public entities may exchange data between them without a notification to the outside world. Additionally, in Europe the open data directive¹⁶ sets the obligation for governments to publish some specific datasets as open data, with no specific agreement behind the actors (as it is foreseen by law), and no specific purpose for the use of this kind of data; this way, the introduction of a G2G / G2B data sharing taxonomy may be less useful than a B2G data sharing one.

Acknowledgments

The authors would like to thank the participants of the New Techniques and Technologies in Statistics 2023 conference for the useful comments and suggestions. This project has been funded through the JRC Centre for Advanced Studies and the project Computational Social Science for Policy (CSS4P).

Conflict of interest

No potential conflict of interest was reported by the authors. The views expressed are purely those of the

authors and may not in any circumstances be regarded as stating an official position of the European Commission.

References

- [1] UN Global Working Group on Big Data for Official Statistics. Handbook on the Use of Mobile Phone Data for Official Statistics; 2019. Available from: <https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf>.
- [2] Vespe M, Minora U, Iacus S, Spyrtos S, Sermi F, Fontana M, et al. Mobility and Economic Impact of COVID-19 Restrictions in Italy Using Mobile Network Operator Data. LU: Publications Office; 2021.
- [3] Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. *European Journal of Information Systems*. 2013 May; 22(3): 336-359. Available from: doi: 10.1057/ejis.2012.26.
- [4] DIGIT Directorate-General for Informatics, Programme I. European taxonomy for public services. European Commission; 2019. Available from: https://joinup.ec.europa.eu/sites/default/files/news/2019-09/ISA2_European%20taxonomy%20for%20public%20services.pdf.
- [5] Brackstone G. Managing Data Quality in a Statistical Agency. *Survey Methodology*. 1999; 25(2): 139-149. Available from: <https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999002/article/4877-eng.pdf?st=h4M7snWA>.
- [6] Vale S. Statistical Data Quality in the UNECE; 2010.
- [7] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016 Dec; 3(1): 160018. Available from: <https://www.nature.com/articles/sdata201618>.
- [8] Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, et al. The CARE Principles for Indigenous Data Governance. *Data Science Journal*. 2020 Nov; 19: 43. Available from: <https://datascience.codata.org/articles/10.5334/dsj-2020-043>.
- [9] Publications Office of the European Union. Data.europa.eu data quality guidelines. LU: Publications Office; 2021. Available from: <https://data.europa.eu/doi/10.2830/79367>.
- [10] European Commission Statistical Office of the European Union. Empowering society by reusing privately-held data for official statistics: a European approach : final report prepared by the high level expert group on facilitating the use of new data sources for official statistics, 2022 edition. LU: Publications Office; 2022. Available from: <https://ec.europa.eu/eurostat/web/products-statistical-reports/-/ks-ft-22-004>.
- [11] Skaliotis M. Timeliness and Accuracy in Official Statistics 2.0. Malta; 2010.
- [12] Grazzini J, Lamarche P, Gaffuri J, Museux JM. “Show me your code, and then I will trust your figures”: Towards software-agnostic open algorithms in statistical production; 2018.
- [13] Abraham R, Schneider J, Vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*. 2019 Dec; 49: 424-438. Available from: doi: 10.1016/j.ijinfomgt.2019.07.008.
- [14] Publications Office of the European Union, Capgemini Invent, European Data Portal. Business-to-government data sharing. LU: Publications Office; 2020. Available from: <https://data.europa.eu/doi/10.2830/078126>.

¹⁵<https://ted.europa.eu/en/>.

¹⁶Directive (EU) 2019/1024 <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>.

- [15] European Commission. Commission Staff Working Document: Guidance on sharing private sector data in the European data economy; 2018. Available from: <https://digital-strategy.ec.europa.eu/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy>.
- [16] Micheli M. Public bodies' access to private sector data: The perspectives of twelve European local administrations. First Monday. 2022 Feb; Available from: <https://firstmonday.org/ojs/index.php/fm/article/view/11720>.
- [17] Saaty RW. The analytic hierarchy process – what it is and how it is used. *Mathematical Modelling*. 1987; 9(3-5): 161-176.
- [18] Vargas LG. An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research*. 1990; 48(1): 2-8.
- [19] Beiderbeck D, Frevel N, von der Gracht HA, Schmidt SL, Schweitzer VM. Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*. 2021; 8: 101401.