A method for assessing the degree of openness of Semi-Open Data initiatives: applied to the justice domain

Mortaza S. Bargh* and Sunil Choenni

Research and Documentation Centre (WODC),
Ministry of Justice and Security,
The Hague, 2596AA, The Netherlands

and

Rotterdam University of Applies Sciences (RUAS),
Research Centre Creating 010,
Rotterdam, 3011WN, The Netherlands
Email: m.shoae.bargh@wodc.nl
Email: m.shoae.bargh@hr.nl
Email: r.choenni@wodc.nl
Email: r.choenni@hr.nl
*Corresponding author

Ronald Meijer

Research and Documentation Centre (WODC),
Ministry of Justice and Security,
The Hague, 2596AA, The Netherlands
Email: r.f.meijer@minvenj.nl

Sharvin Choenni

Cegeka, Business Solutions,
Zoetermeer, 2719 EK, The Netherlands
Email: sharvin.choenni@cegeka.com

Abstract: Some public organisations struggle to meet the Open Data requirements fully because their data often have sensitive (personal) information, are of low quality, or have interoperability issues in terms of format and semantics. These restrictions quite often hold for justice domain datasets. In order to be transparent, nevertheless, many of such public organisations do share their data in a way that it partially satisfies the open data requirements. These partially opened datasets do not count as Open Data and, therefore, the efforts that organisations put behind these initiatives are not acknowledged adequately and appropriately. To acknowledge such data opening initiatives, we advocate and describe a method to assess the degree of data openness, as a first step for recognising such so-called Semi-Open Data initiatives. We carry out eight case studies, not only to validate the proposed method, but also to show how the method can be deployed in practice.

Keywords: assessment; indicator; measurement; method; open data; openness; Semi-Open Data.

Reference to this paper should be made as follows: Bargh, M.S., Choenni, S., Meijer, R. and Choenni, S. (2022) 'A method for assessing the degree of openness of Semi-Open Data initiatives: applied to the justice domain', *Int. J. Electronic Governance*, Vol. 14, Nos. 1/2, pp.207–235.

Biographical notes: Mortaza S. Bargh is a Scientific Researcher at the WODC and a Research Professor on Privacy and Cybersecurity at the RUAS. His current research interests include privacy and security by design, privacy preserving data mining and data publishing, responsible machine learning, access and usage control, usable security, and risk management.

Sunil Choenni is the Head of the Statistical Information Management and Policy Analysis Department of the WODC. Furthermore, he is a Research Professor in the field of Future Information Technologies at the RUAS. His research interests include data warehouses and data mining, databases, e-government, cyber security, and human-centred design.

Ronald Meijer is a Scientific Researcher at the WODC. His current research interests include information management, data archiving, privacy preserving data sharing (namely for scientific research, datamining and data-labs) with a focus on open data in the judicial domain.

Sharvin Choenni is a business solution developer at Cegeka. Currently, he is working on topics related to CRM and ERP Systems. His interests include cybersecurity and data processing.

This paper is a revised and expanded version of a paper entitled 'A method for assessing the openness of semi-open data initiatives: applied to the justice domain' presented at the *8th International Conference on e-Democracy*, Athens, Greece, 12–13 December.

1 Introduction

Gaining public trust, achieving transparency, stimulating innovations, and delivering economic growth are considered as some of the driving-forces behind Open Data initiatives for public organisations. Open Data initiatives have also gained momentum in the justice domain. Within the justice domain, the notion of Open Justice has been recognised even long before transparency (via Open Data) became an important aspect of governance (McLachlin, 2014). Open Justice refers to making courts and their proceedings open and public so that what is done in the name of justice can be scrutinised and criticised. Nowadays, however, Open Data within the justice domain extends the scope of Open Justice to beyond simply court proceedings and judgements and also includes the data gathered within the administration processes and the judicial procedures of the whole justice branch of government.

Opening data, particularly in the justice domain, is subject to a number of challenges like privacy violation, misinterpretation and misleading. These challenges, if not addressed appropriately, can have adverse impacts on individuals, the public and society. Therefore, many public organisations, particularly those in the justice domain (Open data

trend report, 2015, 2016), have been hesitant or unable to open their data in a way that it satisfies all Open Data requirements, e.g., the data being opened as raw as possible (or as they are), for everybody, in a timely way, and together with primacy, permanence and appropriate metadata. Actually, a large number of public organisations, particularly those in the justice domain, do share a modified form of their data with the public after putting lots of effort to address the privacy, misinterpretation and misleading issues of the data (Bargh et al., 2015, 2016b; Meijer et al., 2014; van den Braak et al., 2012). For example, the shared data are processed, aggregated, and offered to specific data consumers (e.g., scientists) in order to protect the privacy of data subjects or to enhance the quality of data. Actually, fostering democratic processes (i.e., monitorial, deliberative and participatory processes) asks for transforming raw Open Data into meaningful information collectively by public administrators and citizens (Ruijer et al., 2017).

Despite all efforts, the abovementioned data sharing initiatives are not classified as Open Data and the involved organisations cannot position themselves as Open Data compliant. This problem seems to be widespread among countries and neither is it limited to the justice domain. For example, out of 28 EU national Open Data portals, 15 and 16 national portals in 2019 and 2020, respectively, allowed the public to see which data exist but could not made them available as Open Data due to various reasons like confidentiality or privacy (Blank, 2019; van Knippenberg, 2020). Therefore, such organisations cannot demonstrate their dedication towards the ideals and objectives of Open Data in being, for example, transparent and supportive of innovations and economic growth. This (negative) image can be costly for public organisations, as they may not gain (or even lose) the public trust. Being unable to share their data according to the full requirements of Open Data and not being recognised when sharing their processed data are two sides of the same problem that public organisations, particularly those in the justice domain, face today. It is worthwhile to note that, among those domains considered in both of the EU Open Data maturity reports in 2019 and 2020, the domain of justice, legal system and public safety ranks among the lowest in popularity (Blank, 2019; van Knippenberg, 2020). Blank (2019) further remarks that there is no way to say whether such an unpopularity is due to having lack of audience interest or due to publishing not enough relevant and valuable data.

In order to acknowledge the efforts of those organisations that share their data in a not-fully-Open-Data-compliant way, while they basically push the frontiers of information sharing towards the ideals of Open Data, Bargh et al. (2017) coin the term of Semi-Open Data for those initiatives. More specifically, "Semi-Open Data paradigm includes those data sharing solutions that aim at Open Data Objectives (like transparency, compliance, innovation, decision support, cost reduction, participation, and collaboration) but do not fulfil all conditions of Open Data" (Bargh et al., 2017, p.10). We note that the term Semi-Open Data refers to a wide spectrum of the types of data sharing initiatives that fall between the two extremes of closed/confidential data and Open Data. A main step to make Semi-Open Data initiatives visible is to somehow indicate their positions on this spectrum between the two extreme points. To this end, we propose assessing the degree of adherence of these Semi-Open Data initiatives to the requirements and objective of Open Data. We use the term 'degree of openness' to refer to how much Semi-Open Data initiatives adhere to the requirements of Open Data.

In this contribution, we aim at devising a systematic method for assessing the degree of openness of a dataset shared by an organisation. The proposed method determines the degree of openness of Semi-Open Data initiatives based on the Open Data requirements,

which, in turn, are based on the Open Definition (Open Definition, 2019). The existing assessment methods either

a indicate whether a data sharing initiative fulfils all requirements of Open Data or not, thus making a binary decision on data openness like (GODI methodology, 2019), or

b consider the degree of openness of a data sharing initiative in combination with the degree of data usage, thus providing an integrated measure of data openness and data utility like (Khan and Foti, 2018; Dodds and Newman, 2015).

Unlike these existing methods, our proposed method aims at providing a more granular indication of openness with respect to the exiting binary assessment methods and at making the efforts of the public organisations behind those data opening initiatives visible within the Open Data landscape, regardless of how they are being used. As a consequence of the latter advantage, our method is not affected by how the data are utilised by the public or third parties, thus just aiming at transparency at the supply side (so-called potential transparency). Note that one can define a similar measure to make the demand side transparent (so-called achieved transparency). Considering such a demand figure next to the supply figure defined in this contribution (especially in a longitudinal manner), one can gain insight in the impact of transparency in data opening in the two-dimensional plane of data openness vs. data impact. This perspective is, however, out of our scope in this contribution. In other words, our method provides a measure of potential transparency and does not indicate a measure of achieved transparency. Making the supply side efforts visible not only encourages such organisations to continue opening (more of) their data, but also provides a (more) realistic view on the landscape of Open Data (Bargh et al., 2017).

The contributions of this paper are threefold. Firstly, we build on the results of (Bargh et al., 2017, 2016a) about Semi-Open Data by specifying how to assess the degree of data openness in Semi-Open Data initiatives. Our proposed method adopts a multi-dimensional measurement approach to quantify Semi-Open Data initiatives in terms of their adherence to the Open Data requirements. Secondly, we report on the results of applying the assessment model in practice to eight datasets made open already in the Netherlands (four datasets) and four other countries (one dataset per country). Thirdly, we raise the importance of a data openness measure that is not binary (i.e., the data being open in its generic sense or not) and not closely intertwined with data usage aspects.

Our study touches the justice domain in two ways:

1 most datasets in the justice domain are prune to high-impact privacy and misinterpretation implications, which makes Semi-Open Data relevant for these datasets

2 for our case studies we choose some datasets from the justice domain to illustrate the feasibility of applying the proposed method in practice.

The justice domain here denotes the whole justice branch of government, not just courts and judgements. Although our study is concerned with the justice domain as mentioned above, our results are generic in that the proposed method is applicable to all datasets from any domain. The research methodology used for this study is based on three processes. Firstly, the proposed assessment method of data openness draws on the authors' experience to share data in the justice domain for many years. Secondly, we use a number of case studies to validate the feasibility of, and to illustrate how to define and

fine-tune the proposed method. Thirdly, the study is grounded in the body of literature. This paper is a revised and expanded version of our previous work (Bargh et al., 2019a). Compared to (Bargh et al., 2019a), we have newly added Section 2 (on the theoretical and practical background of this study) and Section 3 (about the related work), and have substantially extended Section 6 (the discussion and future work).

This paper is organised as follows. In Section 2, we provide some background information about the key concepts used in this contribution. In Section 3, we review a number of studies with overlapping scope to ours. In Section 4, we present the method proposed for assessing the degree of data openness. In Sections 5, we report on applying the assessment method to eight open datasets from the justice domain. In Section 6, we discuss the potentials and limitations of the proposed method and present some directions for future research. Finally, in Section 7, we present our conclusions.

2 Background

This section provides some background information about Open Data objectives (Section 2.1) and the need for adopting a broader scope for the justice domain data than the scope of the data pertaining to the traditional Open Justice (Section 2.2) as the datasets used as case study in this contribution belong to the former (i.e., the broader justice domain data). Finally, we describe the motivation behind and the approach of this study in Sections 2.3 and 2.4, respectively.

2.1 Open data objectives and roadmap

Governments want to provide citizens with the data of government organisations, which are often collected with public funds. In this way, governments intend to improve government transparency, to increase government accountability and compliance, to support participatory governance by citizens, to foster fair competition, innovations and economic growth, and to enable citizens and businesses to make informed personal and business decisions. In contributing to transparency, Open Data fulfil a necessary condition for a well-functioning democratic state of law. Transparency serves the legitimacy of public administrations and the trust of civilians in governments (Meijer et al., 2013, 2014). Open data initiatives can even foster democratic processes when they are designed appropriately, i.e., via transforming raw data into meaningful information by public administrators and citizens collectively (Ruijer et al., 2017). For a detailed overview of the, so-called, open government objectives, the interested reader is referred to (ODRN, 2013; Verhulst et al., 2014; Elena and Mercado, 2019; EU Directive 2019/1024, 2019; OECD, 2018).

Open government via Open Data can be considered as the intermittent step in transforming the use of information and communication technology (ICT) within government organisations. This transformation began from e-administration, then led to e-government, and, relying on open government, intends to establish smart-governments. The intention within the vision of smart government is to enable open innovation in public agencies and maximise interoperability among public agencies along semantic, technical, organisational, governance, etc. dimensions (Jiménez et al., 2014). Recently there is a noticeable shift from quantity to quality. Whereas the focus in early years was

on publishing more and more data, nowadays governments tend to focus on ensuring the value of data via their reuse. This shift of focus often means improving the quality of data primarily. Improving data quality here stems from the idea that high quality data enhance data re-use, which, in turn, impact society and the economy more significantly (Blank, 2019). Fostering a culture of value creation and problem-solving is seen helpful for targeting and prioritising the Open Data efforts in order to release reusable data rather than simply providing more data. In this context, the OECD has coined the term 'Publish with a Purpose' (OECD, 2018). Beyond Open Data, governments are becoming more aware of the opportunities arising from data sharing in general. "EU28 + countries are preparing to be effective at data sharing with other governments and organisations in a secure way and in full respect to intellectual property and privacy" (Blank, 2019).

2.2 From open justice to open data for the justice domain

Justice is one of the three pillars of government in a constitutional democracy and, therefore, open government can be enabled through the opening of the justice domain data to the public. Openness has been a deeply rooted principle in our justice systems. Courts have promoted the idea of opening their proceedings to the public, long before transparency became a prominent aspect of governance (McLachlin, 2014). This, so-called, Open Justice principle has traditionally been limited to informing citizens about how justice is being rendered via having open courts and making judicial judgements public. In this way, "the litigants, the media, legal scholars and ultimately the general public may follow, scrutinise and criticise what is done in the name of justice" (McLachlin, 2014, p.2). Hereby the transparency of judgements ensures that citizens' rights are upheld through giving the oversight necessary for protecting the public interest.

With the advent of governments' Open Data initiatives, we argue that the scope of Open Justice is extended along, at least, two directions. Firstly, in the justice domain, the objective sought from being open is extended from the transparency principle, as sought in procedurally and common law, to also the other principles of open government, namely: accountability, collaboration and participation (Jiménez-Gómez, 2017). Jiménez-Gómez (2017) coins the term 'Open Judiciary' to refer to this extended view. Secondly, we observe that the scope of the data has expanded from the data of court proceedings and judgements to the data gathered also within the administration processes and procedures of the whole justice branch of government. Elena and van Schalkwyk (2017) name court ruling data, statistics on operations, and budget and administrative data as the *least* a judiciary should open. This set of data types, we conclude, is subject to expansion. As a consequence of the second extension direction, we note that the data within the justice domain are generally gathered by various independent organisations involved in countries' justice domain. Lampoltshammer et al. (2017) use the term *justice system* to refer to the (chain of) bodies in the apparatus of law, which are involved in creating data, from legislative texts to judicial decisions; and not just those being involved in courts.

For example, the Open Justice initiative led by the California Department of Justice publishes such a wide range of data types from various sources within the criminal justice system (like trends in arrests, crimes, death in custody, hate crimes, homicide, juvenile court and probation). In the case of Dutch government's justice branch, as another example, the involved organisations agencies include the National Police, the Public Prosecution Service (PPS), the courts, the Central Fine Collection Agency (CFCA), the

agency of correctional organisations (i.e., prisons) and the Probation Service (PS). Note that the police are not considered as part of the justice domain in some countries. Without loss of generality, nevertheless, we mention it here and include police datasets in our case studies because the police are part of the justice domain in some other countries like The Netherlands. For many years the justice administration and procedural data are published by our organisation annually in reports (Kalidien et al., 2014). On the other hand, an internet site called Rechtspraak.nl publishes Netherlands' court proceedings and judgements for criminal, civil and administrative cases regularly.

Considering the variety of the organisations involved in the justice systems as mentioned above, we argue that opening the data pertained to traditional Open Justice (i.e., judicial and court data) provides a limited view on such justice systems. This argument becomes even stronger when one notes that in certain countries the organisations involved in crime detection, prosecution, trial and probation have overlapping tasks and functions nowadays, which cannot easily be separated. In the Netherlands, for example, the arrest of a suspect does not necessarily lead to further prosecution. In some cases, the police may decide to handle the case by dismissing it, proposing a transaction, or imposing a punishment order. Moreover, the public prosecutor decides which cases to be prosecuted or dealt with by courts. For example, the public prosecutor may dismiss a case, propose a transaction, impose a punishment order, or decide to send the case to court (van der Leij, 2016).

Therefore, we argue that opening data in the justice domain (i.e., the justice branch of government) may cover the whole justice system of a country and is not necessarily limited to just traditional judicial data. One may call this 'Open Justice in its extended sense' or alternatively as 'Open Data in the justice domain'. Consequently, Open Data in the justice domain not only contribute to realising the traditional vision of Open Justice but also help realising the progressive vision of smart justice, as sketched in Netten et al. (2018) and the references therein.

2.3 Motivation

Some Dutch organisations have been unable to open their datasets as expected (Bargh et al., 2016a). In Open Data trend report (2015) and Conradie and Choenni (2014) a number of reasons behind having a low number of opened datasets are mentioned. One of these reasons is the restrictive character of Open Data requirements. This restrictive character is relevant for our study as it underscores the fact that data opening might be measured not according to the current measurement regime, i.e., the requirements of Open Data. This is the case where the data to be opened have inconsistent, imprecise, uncertain, missing, and incomplete data objects (thus, having low quality), have private or business sensitive information (conceivably when combined with other datasets or background information), or have proprietary and unstandardised format and semantics.

These issues and deficiencies often exist for the data in the justice domain (van den Braak et al., 2013). Opening such data according to Open Data requirements (i.e., as they are, to the public, etc.) may lead to various problems such as privacy and business sensitive information disclosures, misinterpretation and misleading outcomes, and no or low economic growth. Consequently, organisations could not open such data according to the Open Data criteria, despite their willingness to share their data with some modification and edition, in a limited scope (e.g., in being discoverable online but not downloadable), or in a PDF (Portable Document Format) format. Through investing in

time, efforts and resources, organisations can eliminate data sensitivity, improve data quality, harmonise data format, and create appropriate metadata for the data. These operations, however, not only inflict extra costs on organisations but also result in opening only processed data. Both aspects (i.e., being costly and being modified) quite often violate some basic requirements of Open Data.

In conclusion, currently one cannot consider opening of processed and not-for-free datasets as Open Data, while they do serve the same purposes of Open Data to some degrees. As an example of serving the Open Data purposes, one can consider the case of entrepreneurs purchasing the processed data to make innovative services and products. This may lead to economic growth. Independent domain experts and the public can learn about public organisations by using the (high quality) processed data and can examine whether the public organisations adhere to their missions or to existing laws and regulations. As such, not-fully-compliant Open Data initiatives do also aspire and drive individuals, governments, and businesses for improving their existing or devising new policies, services, products, and processes.

Motivating our proposed method, we reason that every attempt to achieve part of these Open Data objectives must be recognised, acknowledged, and encouraged. As we noted in the forgoing, this recognition is particularly important in the justice domain where it is often infeasible to meet all requirements of Open Data.

2.4 Approach

Our approach distinguishes itself from similar works by decomposing the complex assessment process of Open Data and focusing only into one of its core components that is crucial for opening imperfect and sensitive data like the justice domain data at this point in time. This core component of Open Data assessment that is considered in our approach is the degree of data openness in sharing such imperfect and sensitive data, i.e., determining how far a Semi-Open Data initiative is from its ideal case of an Open Data initiative, as required by Open Data requirements. To this end, we will adopt the general requirements of Open Data as the indicators of our assessment method in order to make the proposed method intuitive and pragmatic. Being intuitive stems from the fact that these Open Data requirements have already been embedded in the definition of Open Data. Being pragmatic drives us to define a limited number of indicators and indicator levels in order to make the method feasible and affordable in practice. Therefore, we look for an assessment method that is actionable and that does not require too much effort (like making surveys and conducting interviews) or does not inflict ambiguity when determining its parameters.

Although our assessment method can be applicable to any domain, we focus on the justice domain as a meta-case study and show how to fine-tune the assessment method to eight datasets selected from this domain as our feasibility case studies.

3 Related work

In this section, we start with giving an overview of the related work on the concept of Semi-Open Data, a concept which has motivated us to devise the assessment method. Subsequently in Section 3.2, we provide a comparative review of the existing work on assessing Open Data initiatives with respect to our proposed method.

3.1 On semi-openness

In Bargh et al. (2016a, 2017) the concept of semi-openness for shared datasets are introduced. One can find some tracks of semi-openness in literature for characterising the data dissemination process. For example, a fact-finding committee of The Royal Netherlands Academy of Arts and Sciences considered a number of recent fraudulent practices in Dutch universities (van den Hoogen, 2013). The committee advised universities to be aware of the importance of careful and proper handling of research data. For example, the scientific research outputs such as the publications and the underlying data should be open to validate research results, prevent fraudulent research, and facilitate data reuse. To open scientific data for transparency, accountability, and innovation purposes, the School of Business and Economics of Maastricht University considered a form of Semi-Open Data for their research data. The "school … is thinking about a Semi-Open Data policy: raw, self-generated data are stored at the faculty level with restricted access and a careful description of the generation process, and made available for 10–15 years after generation (in case of problems). Manipulated data should be stored at a central university level, also well documented and in principle available for colleagues to use" (van den Hoogen, 2013). In Nosek et al. (2015) there are similar requirements proposed for publishing the information relevant for reproducibility of experiments, which are already adopted by some journals and conferences (Alberts, 2015). Similarly, we consider data published with restricted access, with enough metadata, and within limited usage intervals as examples of Semi-Open Data.

The Open Data in Developing Countries (ODDC) (ODRN, 2013) considers the way in which Open Data is defined as important since the definition determines which datasets should be classified as Open Data. The ODDC gives a broader definition of Open Data by considering also the impact and context of Open Data, namely "Open Data = Open definition + (impact/context)". For example, data granularity and timeliness can be context-dependent in the sense that the data are used for improving accountability or for creating economic impacts. Assumingly, one can deliver different versions of the dataset for these purposes. We think that making the Open Data definition dependent on its impact/context is inconvenient (as also the idea is not supported by others, to the best of our knowledge). We, nevertheless, acknowledge the existence of this dependency and argue that the proposed Semi-Open Data vision includes those aspects inherently.

3.2 Similar open data assessment methods

There are many perspectives for evaluating (the infrastructures of) Open Data initiatives and for the proposed metrics (Charalabidis et al., 2018). Proposing a framework for comparing Open Data policies, Zuiderwijk and Janssen (2014) conclude that current policies are inward looking in the sense that, when opening data, organisations do not focus on the impact of Open Data. Some other authors have also advocated for a comprehensive model to assess the success of Open Data initiatives. For example, Donker and Loenen (2017) propose a framework to assess the success of Open Data holistically, based on Open Data characteristics of data supply, data governance and data usage. Veljković et al. (2014) propose a similar holistic measurement method through defining an 'e-government openness index' based on whether high-value data categories are opened or not, on the amount of data openness (noting that this aspect comes close to our method), and on the amount of transparency. Using this so-called "e-government

openness index", the authors provide a maturity measure to indicate the government's readiness to change. There are also other Open Data assessment methods – like Open Data Barometer (ODB) developed by the World Wide Web Foundation (ODB, 2018), Open Government Partnership (OGP) Tagging Framework developed by Independent Reporting Mechanism (IRM) body (Khan and Foti, 2018), and the Maturity framework developed by the United Kingdom Open Data Institute (Dodds and Newman, 2015) – that measure a combination of the characteristics of Open Data considered in Donker and Loenen (2017). Another method in this category is Capgemini Consulting Open Data Benchmark (Capgemini, 2015), which aims at assessing the Open Data Maturity of each European country, based on two key indicators: Open Data Readiness (encompassing the usage and governance aspects) and Portal Maturity (encompassing the access network aspects) in terms of, for example, supporting regular data updates, enhanced search capabilities, user participation capabilities, and user feedback mechanisms. Sandoval-Almazan and Gil-Garcia (2016, 2018) propose comprehensive assessment frameworks to evaluate the websites of open government and e-justice/Open Justice, respectively, from the viewpoint of practitioners, information availability, and/or participatory mechanisms of the websites.

Unlike the assessment methods mentioned above, our assessment method aims at assessing the degree of data openness per dataset so that organisations can add these data openness values up for all datasets that they have opened. In this way, every organisation can count the effective total number of the datasets it has opened. The other assessment methods (i.e., those mentioned above), however, do include the other aspects like governance and usage aspects, are limited to some specific aspects like the access network aspect, or do provide a normalised value for all organisations in a country. In Section 6 we discuss the differences of the proposed method for assessing data openness with the other data assessment methods in more depth after laying down the principles of the proposed method in Sections 4 and 5.

Book chapter (Charalabidis et al., 2018) provides a classification (i.e., a taxonomy) of the metrics to evaluate various aspects of the lifecycle of Open Data initiatives (like those for readiness, impact and value creation, performance, quality and post-adoption). The proposed taxonomy is developed based on various evaluation models of information systems (ISs). These evaluation models are either subjective, i.e., being based on collecting users' opinions about a system (e.g., usefulness), or objective, i.e., being based on predefined metrics and values. The example subjective evaluation models considered are IS evaluation models, technology acceptance models, IS success models, e-services evaluation models, maturity assessment models, organisation/data readiness assessment models, post adoption assessment models, and impact assessment models. These evaluation models could be either qualitative or quantitative. Considering the scoping of Open Data metrics discussed in Charalabidis et al. (2018), our assessment method is quantitative and objective. In other words, being quantitative enables our method to capture the impacts of all datasets opened by the organisation, and being objective enables our method to capture the efforts put by the organisation to open its datasets according to Open Data ideals. These properties of the proposed assessment method conform the aspects that we intend to evaluate for Open Data initiatives within our organisation, as described in the introduction section. Having a conformance between the aspects to evaluate Open Data initiatives and the method chosen is indeed emphasised in Charalabidis et al. (2018). We further note that (Charalabidis et al., 2018) provides a comprehensive taxonomy of Open Data evaluation metrics that includes all subjective

and objective as well as quantitative and qualitative aspects. Out proposed method employs those elements of the taxonomy that are relevant for our evaluation purpose. As such, the relevant metrics and their values in the taxonomy of Charalabidis et al. (2018) can be used to extend our method (i.e., its indicators and their indicator values) in the future.

4 Proposed method for assessing data openness

In this section, we present the proposed method for assessing the degree of data openness.

4.1 Open data characteristics

Various governments and organisations have defined a number of requirements for achieving the objectives of Open Data. These requirements determine the characteristics of Open Data. Bargh et al. (2016a) investigate these requirements from four sources: Canadian government (Open Definition, 2019; Wonderlich, 2010), US government (Burwell et al., 2013), Dutch government (Open data trend report, 2015), and from South Africa's Open Data Research Network (ODRN, 2013). These sources agree on some of the Open Data requirements. For example, all of them require Open Datasets to be licence free and reusable. The sources, however, do not have the same viewpoint on the aggregation level of the data opened, for example, with respect to the 'as is' requirement of Open Data. Our method includes all these requirements in order to accommodate the viewpoints of all these sources. This inclusiveness can be extended readily, should one decide to comply with also the new requirements of other sources.

4.2 Indicators of openness

For comparing entities, one can use a benchmarking tool with a set of defined indicators (Rorissa et al., 2011). For measuring the degree of openness, Veljković et al. (2014) use eight characteristics of Open Data from Open Government Working Group (Open Government Working Group, 2007) as indicators. These indicators are: Being complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary and licence free. Similarly, our method relies on defining a number of indicators of data openness, but our proposal bases them on the identified requirements of Open Data, as summarised in Bargh et al. (2016a). Our indicators are quite similar to the generally accepted requirements of Open Data in order to make them

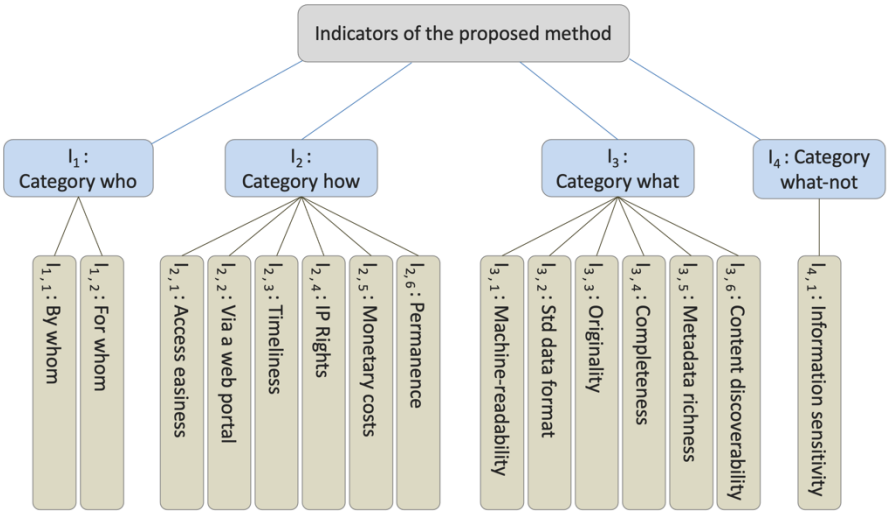
a meaningful – as they are backed up and promoted by some prominent sponsors and proponents of Open Data ideals

b actionable – as we investigate and show the feasibility of their implementation in practice in this paper by applying them to a range of open/shared datasets from the justice domain (see the following sections).

Further, we are careful not to define too many indicators as the curse of high dimensionality may result in meaningless distance measures (Bishop, 2006).

We structure our indicators at two hierarchical levels, as indicated in Figure 1. To this end, we are inspired by our experience in Open Data research as well as our daily practice of opening justice domain datasets.

Figure 1 An illustration of the two-level hierarchical structure of the proposed model in categories and indicators (see online version for colours)



The proposed high-level and low-level indicators are:

Category who (denoted by I_1) which captures the target group for data opening and the entity in charge of data opening. The low-level indicators of category who are:

- For whom indicator (denoted by $I_{1,1}$) that determines who may receive the data. Ideally, the data should be available for everybody without any discrimination, without any need for identification, or without any need for usage justification. This access is restricted to the extent that is permitted by law and regulations (i.e., subject to privacy, confidentiality, security, or other valid restrictions).
- By whom (or primacy) indicator (denoted by $I_{1,2}$) that determines the source primacy of the data. Ideally the data should be published by its primary sources in order to, among others, enable a proper control of data collection and storage.

Category how (denoted by I_2) that captures how the data are made available and how the data may be used. The low-level indicators of category how are:

- Access easiness indicator (denoted by $I_{2,1}$) that determines how easy the process of data access is. Ideally the data should be accessible without any technological (e.g., requiring browser-oriented technologies) and procedural restrictions.
- Via a web-portal indicator (denoted by $I_{2,2}$) that determines how to download the data. Ideally the data should be accessible via an application programming interface (API) for automatic data processing and be accessible via a web-portal (e.g., in the Netherlands via www.opendata.nl).

- Timeliness indicator (denoted by $I_{2,3}$) that determines how fresh the data are published. Ideally the data should be opened timely (i.e., as quickly as possible).
- Intellectual property rights (IPR) indicator (denoted by $I_{2,4}$) that defines the conditions under which the data should be used. Ideally the data should be licence-free and reusable for data recipients in order to enable (commercial) innovations and (commercial) reuse of the data.
- Monetary costs indicator (denoted by $I_{2,5}$) that determines how much one should pay in order to access the data. Ideally the data should be accessed free of charge.
- Permanence indicator (denoted by $I_{2,6}$) that determines how sustainably the data are made available. Ideally the data should remain online with appropriate version tracking, archiving, and history logging over time.

Category what (denoted by I_3) that captures the content and format of the data to be opened as well as what the accompanying data should be. The low-level indicators of category what are:

- Machine-readability indicator (denoted by $I_{3,1}$) that defines how easy the data can be used by automated processes. Ideally machine-readable datasets are based on common file formats, like comma-separated values (CSV) and extensible markup language (XML), which are suitable for machine processing.
- Standardised data format indicator (denoted by $I_{3,2}$) that determines how well defined the schemata of the data is. Ideally the data format should be standardised well for efficient data storage and processing.
- Originality indicator (denoted by $I_{3,3}$) that determines how much processing is done on the data. Ideally, data should be made *open as is* as much as possible.
- Completeness indicator (denoted by $I_{3,4}$) that determines how complete the data bulk should be opened. Ideally, no part of the data should be removed.
- Metadata richness indicator (denoted by $I_{3,5}$) that determines how self-contained the opened data are. With rich metadata the data recipients obtain sufficient information to understand, for example, the strengths, weaknesses, analytical limitations, and security and privacy requirements of the data. Through this understanding data recipients can interpret and process the data appropriately. For example, data processors can treat the data correctly if they are provided with additional descriptions of the purpose of data collection, the population of interest, the characteristics of data samples, and the method of data collection.
- Content discoverability (denoted by $I_{3,6}$) indicator that captures how easy it is to locate the data objects. Ideally, the data objects should be provided with (persistent) unified resource identifiers (URIs) to enable locating data objects.

Category what-not (denoted by I_4) that captures whether the data may not be shared at all from the viewpoint of law, regulations and ethics. We observe that (privacy) sensitive information falls out of the scope of Open Data in all Open Data requirements lists that we have investigated. For example, the exceptions of Open Data in the Netherlands include those datasets that contain privacy sensitive data, national security data, and

business sensitive data. Sensitive information include privacy sensitive and policy sensitive data (Zuiderwijk et al., 2012).

- Information sensitivity (denoted by $I_{4,1}$) indicator for a dataset indicates whether the dataset can be shared with others (i.e., either be opened to the public or be shared with a group) or not. Specifically, $I_{4,1} = 0$ means that the dataset cannot be shared/opened at all, and $I_{4,1} = 1$ means that it can be shared with the public or a specific group.

Note that for sensitive data that can be shared, for example, only with a specific group, our model assumes $I_{4,1} = 1$ and specifies the scope of data recipients by indicator $I_{1,1}$. In this way, our model accommodates non-binary data sensitivity measures such as the one mentioned in Zuiderwijk et al. (2012) with three possible levels: open access, closed access and restricted access.

4.3 Indicator levels and values

The method proposed here for assessing the degree of data openness is based on defining some meaningful and actionable levels for each indicator defined in Section 4.2. According to this method, every indicator can eventually be characterised by a number of ordinal levels based on some agreed upon guidelines, principles and standards. For example, indicator 'for whom' defines the scope of data recipients, i.e., for which groups the data are opened. The 'for whom' indicator can subsequently be specified in the range between 'share with no one' corresponding to confidential data and 'share with the public' corresponding to the common setting of Open Data. The example intermittent levels of the 'for whom' indicator could be: 'share the data within a specific group', 'share the data within a department of an organisation', 'share the data within an organisation/ministry', and 'share the data among a federation of organisations'. As another example, for the 'standard data format' indicator, which relates to data interoperability and link-ability, one can think of the following levels: 'without any specific data format', 'with a data format of acceptable convertibility' (applicable within data space environments, see van Dijk et al., 2013 for the definition of data space), 'with a data format of high/precise convertibility' (applicable within data warehouse environments, see Franklin et al., 2005), and 'with a standardised data format' (applicable within a database management system).

Defining multiple ordinal levels for the indicators of our method should be done based on the consensus of domain experts and stakeholders, given the set of (already opened) Semi-Open Datasets in a domain. Defining these levels per indicator is out of the scope of this contribution, as creating the consensus among all stakeholders in a domain is generally a time-consuming process. This process could resemble those processes for product standardisations, which are guided by also subjective and business interests.

For every ordinal level of an indicator, as defined above, one must subsequently assign a numeric value between 0 and 1. These assigned values should be meaningful in indicating some meaningful distances to the ideal level for that indicator. Note that the ideal level for an indicator, which is assigned a value 1, is defined by Open Data requirements for that indicator (e.g., the ideal level of indicator 'for whom' is 'share with the public'). Ratio values 0 and 1 mean maximum and minimum distances to the ideal value, respectively. Like assigning indicator levels, assigning the values per indicator

should be based on the consensus of domain experts and stakeholders, given the set of (already opened) Semi-Open datasets in a domain. A comprehensive and advanced method for assigning these values is out of the scope of this contribution.

For the purpose of illustrating the approach (i.e., as a proof of concept), we will define these levels for and based on those datasets that we have considered as our case studies. Our intention here is not to be comprehensive but to show the feasibility of the proposed method and our approach for defining its parameters. With a widespread adoption of the proposed assessment method for Semi-Open Data in a certain domain, the stakeholders within a concerned domain could agree upon a comprehensive, actionable and meaningful data model for quantifying every indicator, considering all partially opened datasets as well as the preferences of the corresponding stakeholders.

4.4 Proposed measurement method

Bargh et al. (2016a) propose adopting a multi-dimensional multi-level measurement method for measuring the degree of openness for those initiatives that push the data sharing frontiers towards the Open Data ideals. This proposal aims at replacing those previous methods that make a binary decision on whether a data sharing initiative fulfils all conditions of Open Data or not. In this section, we elaborate on the measurement method in more details as follows.

Let $I_{m,n}$ denote the Open Data indicator representing the mth high-level and the nth low-level indicator as defined in Section 2.2. We assume that indicator $I_{m,n}$ takes a number of ordinal levels from $\{i_{m,n}\}$, which are defined based on some agreed upon criteria (see Section 4.3). Further, we assume that every defined ordinal level $i_{m,n}$ is mapped to a value between 0 and 1, represented by value $|i_{m,n}|$ or $|I_{m,n}|$ alternatively.

We define the degree of openness (DO) for a dataset D, for which indicator $I_{4,1}$ has an indicator level value of 1 (i.e., the information is not sensitive), based on:

$$DO = \left(\sum_{m<4} \sum_{n} | i_{m,n} | \right) \Big/ (\| i \| - 1) \tag{1}$$

where $\|I\|$ is the total number of indicators and $|i_{m,n}|$ is the assigned level-value to indicator $I_{m,n}$ of dataset D. One can also define the Degree of Openness per each high-level indicator as

$$DO_m = \left(\sum_{n} | i_{m,n} | \right) \Big/ \| i_m \|, \tag{2}$$

where $m < 4$ and $\|I_m\|$ is the total number of indicators within category m.

Note that the last indicator (i.e., $I_{4,1}$ representing information sensitivity) has a veto role for opening data (go vs. no-go). If a dataset contains sensitive data, then it cannot be opened due to legal or policy reasons. One can also assume similar roles for any of the other indicators. For example, according to Burwell et al. (2013) processed, also derived or aggregate data may be considered as Open Data only if the data dissemination includes the corresponding primary data. Further, some indicators may affect each other, depending on how they are defined. For example, if data suppression is the processing carried out on the data, then data completeness is affected adversely. Such data

processing means that indicators $I_{3,3}$ and $I_{3,4}$ affect each other. Studying elaborated models that capture these restrictions and impacts is left for future studies.

One can assign weights to the indicators in relation (1), based on the relative importance of corresponding Open Data requirement as:

$$DO^{w} = \left(\sum_{m<4} \sum_{n} w_{m,n} \cdot | i_{m,n} | \right) \Big/ (\| i \| - 1), \tag{3}$$

where $\Sigma_{m<4}\ \Sigma_{n}\ w_{m,n} = 1$. For our case studies we assume all indicators are equivalent. Thus, we used relation (1) instead of relation (3).

5 Case studies

In order to illustrate the feasibility of the proposed method and how its parameters can be finetuned, we apply the proposed method to eight datasets from the justice domain to assess their openness. The intention here is to evaluate the proposed method and it is by no means to compare the performances of the organisations/countries responsible in releasing these datasets.

5.1 Datasets

The analysed datasets are denoted by DS_1, …, DS_8, which are arbitrarily chosen from different sources from the justice domain. Note that these datasets are not representative for the corresponding organisations or countries and they are of different types, formats and sizes. Further, as mentioned before, the police are not considered as part of the justice domain in some countries. Without loss of generality, nevertheless, we include police datasets in our case studies because the police are part of the justice domain in some countries such as The Netherlands.

In the following, we briefly provide more information about these datasets and the organisations that have shared (or partially opened) them. Note that the first four datasets are from the Netherlands and the other four datasets are from India, the USA, Australia and Canada.

Dataset DS_1 (Data set 1, 2019) presents the crime statistics at the national level in The Netherlands. This dataset is based on the microdata from a large number of organisations involved in the Dutch criminal justice system. The dataset is published by Research and Documentation Centre of the Dutch Ministry of Justice and Security (abbreviated as WODC in Dutch) and Statistics Netherlands (abbreviated as CBS in Dutch). It has been published annually since 1985 on topics related to crime and law enforcement in the Netherlands (Meijer et al., 2020) and to the local police and city councils (Smit and Dijk, 2014). The data are presented in 36 tables with different number of attributes and records. The total number of the attributes is about 550.

Dataset DS_2 (Data set 2, 2019) contains an elaborated selection of verdicts from different courts in the Netherlands (e.g., the verdicts of the High Court and Courts of Appeal, as well as the full bench of the court and mediatised verdicts). The Council for the Judiciary of the Netherlands publishes the database. The verdicts are accessible via a web-portal and are stored in an XML format. Each record contains nine attributes. Seven attributes present structured data and two attributes provide semi-structured data. The

structured attributes are: Authority, verdict date, published date, case number, formal relation, law branch, and extraordinary distinguishing values. The semi-structured attributes are a summary of the verdict and the verdict itself.

Dataset DS_3 (Data set 3, 2019) provides data about vandalism and public order problems leading to public space degradation in Amsterdam police jurisdiction and four police districts. The dataset compares these regional problems with those in the whole Netherlands. The dataset pertains to 2016 and is published by the Department for Research, Information and Statistics of the municipality of Amsterdam (abbreviated as OIS in Dutch). The source of the data is the Safety Monitor (abbreviated as VM in Dutch) of the CBS. This dataset contains five attributes: clutter on the street, public property destroyed, graffiti against walls and buildings, dog excrement on the streets and other depletions.

Dataset DS_4 (Data set 4, 2019) contains information about recent (attempted) burglary in the Netherlands. This dataset is published by the Dutch Police. The data can be accessed only via the web-portal of the Dutch Police by entering a postal code of a house. The results are aggregated at least on 4 digits postal code level (out of maximum 6 digits). The results are plotted on a map and may be grouped on higher aggregation level by choosing between five predefined ranges, varying from 500 meters to maximum 25 km.

Dataset DS_5 (Data set 5, 2019) is published by the Ministry of Home Affairs, Department of States, and National Crime Records Bureau of India. It contains data on crimes against scheduled castes – scheduled castes (and scheduled tribes) are various officially designated groups of historically disadvantaged people in India – consisting of three attributes: States/Union Territories (UTs), crime head and one attribute called 2014. Every record in this dataset represents how often a certain type of crime has occurred in a specific state in 2014. The crime data are accessible via a web-platform and may be downloaded in CSV format.

Dataset DS_6 (Data set 6, 2019) presents crime statistics in US at a national level by its volume and rate per 100,000 inhabitants for the period of 1997–2016. This dataset, which is published by the FBI's Criminal Justice Information Services (CJIS), is built up from the data of a large number of organisations involved in the US criminal justice system. The data are presented on an online platform and can be downloaded in an XLS file format with different number of attributes and records. The dataset is structured and has more than 20 attributes.

Dataset DS_7 (Data set 7, 2019) presents crime statistics at the national level in Australia. The dataset, published by the Australian Bureau of Statistics (ABS), contains data about the victims of selected offences, recorded by national police agencies within various states and territories in 2016. Conforming to Australia's National Crime Recording Standard (NCRS), the dataset is presented in five tables, each with a different number of attributes and records.

Dataset DS_8 (Data set 8, 2019) presents crime statistics in British Columbia in Canada. The dataset, published by Statistics Canada, contains 14 attributes and is only accessible by downloading a file which can be done in several formats (like XLS, CSV or SCSV format).

We assume that, as these datasets are already made open, they do not have privacy issues. In other words, the corresponding data sensitivity indicator value $|I_{4,1}| = 1$.

5.2 Illustration of the approach and the results

In order to provide a simple, meaningful, and pragmatic way of assessing the degree of data openness, we defined the indicators similarly to the generally accepted requirements of Open Data as they are promoted by prominent sponsors and proponents of Open Data ideals. For enhancing the feasibility and meaningfulness of the proposed method, we further propose to define the indicator levels and to assign their values in a sense making way, based on, among others, the characteristics of a wide range of partially opened/shared datasets. We will illustrate our approach for defining the indicator levels and their values in the following by applying the approach to datasets DS_1, …, DS_8, and summarise the results in Table 1.

Table 1 A summary of the scoring results for the 8 datasets studied

High and low level indicators		*From the Netherlands*				*From other countries*			
		DS_1	DS_2	DS_3	DS_4	DS_5	DS_6	DS_7	DS_8
Who	For whom	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	By whom	0.50	1.00	1.00	1.00	1.00	0.50	0.50	1.00
How	Ease of access	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Via a web portal	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
	Timeliness	0.50	0.50	0.50	1.00	0.50	0.50	0.50	0.50
	IPR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Monetary costs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Permanence	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
what	Mach. readable	0.25	1.00	0.25	1.00	1.00	0.25	0.25	1.00
	Standard format	0.5	1.00	0.50	1.00	1.00	0.50	0.50	1.00
	Originality	0.75	1.00	0.50	1.00	1.00	0.75	0.75	0.75
	Completeness	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
	Metadata	1.00	1.00	1.00	1.00	0.75	1.00	1.00	0.75
	Discoverability	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
What n.	Info sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Degree of openness (DO)		0.73	0.95	0.75	0.98	0.86	0.80	0.73	0.84

As the datasets are accessible online, i.e., they are ‘for the public’, we set Indicator ‘for whom’ to the value of one (denoted by $|I_{1,1}| = 1.0$) for all of the datasets without attempting to specify any other ordinal levels (and the corresponding ratio values) for indicator $I_{1,1}$. Note that, from now on, we shall similarly assign the maximum value 1.0 for those indicators that assume their ideal level possible. For datasets DS_1, …, DS_8, therefore, we have $|I_{2,1}| = |I_{2,4}| = |I_{2,5}| = |I_{2,6}| = |I_{3,6}| = |I_{4,1}| = 1.0$.

Open Data initiatives normally use a number of attributes as metadata to describe datasets. For determining the value of metadata indicator, we required that the metadata should provide the following information about the data: What the data are about as well as when, where, by whom, how and why the data are collected. We note that in this way we adopt a more specific definition for metadata, compared to the definitions used in database community. We identify the following eight metadata-items in the datasets we

studied: title, description, language, theme, keywords, licence, publisher, and references. To score the corresponding indicator $I_{3,5}$ we gave values 0.00 or 1.00 if all of these items are missing or present, respectively, and deduce 0.125 per missing metadata item. For example, dataset DS_5 and DS_8 have two metadata items missing, therefore their $|I_{3,5}| = 0.75$.

Only datasets DS_2, DS_4 and DS_6 can be accessible from a web-portal directly (thus their indicator $|I_{2,2}| = 1.00$). The other datasets should be downloaded before one can access their content (thus the corresponding indicator $|I_{2,2}| = 0.00$).

For the standard format indicator $I_{3,2}$, we assign value 1.0 to those data formats that are interoperable with any other format (such as XML or CSV). Standard formats that are not interoperable with other standards (such as the XLS format) obtain value 0.50. For all the other cases, the indicator value 0.00 is assigned. Datasets DS_2, DS_4, DS_5 and DS_8 are in XML or CSV format (so their $|I_{3,2}| = 1.0$) and the rest of the datasets are in XLS format (so their $|I_{3,2}| = 0.5$).

For the timeliness indicator $I_{2,3}$, it appeared difficult to verify whether the investigated datasets are published in a reasonable timeframe, noting that most of these datasets are published in a rather long timeframe. Therefore, we assign $|I_{2,3}| = 0.50$ for all these datasets except for dataset DS_4 that we give a perfect score because it has a much better time frame than the other datasets.

For the originality indicator $I_{3,3}$, we assigned a value of 1.00 for those datasets that contain only plain numbers or strings. For those datasets with edited data in a way that we can somehow trace back to the original data values, we gave a value of 0.75. For those datasets with edited data items that we cannot trace back to the original data values, we gave a rating of 0.5 or lower. For example, if there is not enough information available to compute from a given percentage the absolute original number, the indicator value 0.5 is assigned.

As we searched for Open Datasets, the found and investigated dataset appeared to be incomplete in some way. For example, we could see all crime statistics, but we could not find all police reports about those statistics. Therefore, we concluded that all of the investigated datasets are in fact incomplete and assigned a value of 0.75 for their completeness indicator $I_{3,4}$.

Regarding the machine-readable indicator $I_{3,1}$, the question is how clean, for example, XLS files should be in order to be considered as machine-readable. We determined our rating as follows. If the XLS file can be opened with another format than XLS, for example if one can open an XLS file as a CSV file, then we consider it as clean (thus machine readable $|I_{3,1}| = 1.00$). If the file is not a CSV file, i.e., cannot directly be opened with any tool, but there is some pre-processing needed to open the file and make it machine-readable, then $|I_{3,1}| = 0.50$ or 0.25 depending how much pre-processing is needed, for example, ranging from removing a mark-up layer to adapting binary codes. For datasets DS_1, DS_3, DS_6 and DS_7 there was a significant amount of pre-processing necessary in order to make them machine-readable (like adapting binary codes). Therefore, for these datasets we assigned $|I_{3,1}| = 0.25$. For the other datasets in XML and CSV files, the readability is the highest, i.e., $|I_{3,1}| = 1.00$.

To score by whom (i.e., primacy) indicator, we assign $|I_{1,2}| = 0.50$ if a dataset is the product of the data collected from several organisations and assign $|I_{1,2}| = 1.0$ if a dataset is published by an organisation that also produces the dataset.

6 Discussion

In this section, we discuss the potentials and limitations of this study and elaborate on some future research directions.

6.1 Comparison with similar measures

As mentioned in Section 3.2, many scholars and organisations have advocated for adopting a holistic and comprehensive model to assess the success of Open Data initiatives (Zuiderwijk and Janssen, 2014; Donker and Loenen, 2017; Veljković et al., 2014; ODB, 2018; Khan and Foti, 2018; Dodds and Newman, 2015; Capgemini, 2015; Sandoval-Almazan and Gil-Garcia, 2016, 2018). Such holistic open data assessment frameworks are sensible for ex-post evaluations, where Open Data initiatives have already been rolled out for a long enough time so that the data could be absorbed and consumed by data recipients. However, for newly opened datasets or for those Open Data initiatives that are being rolled out in new domains (like those initiatives in the justice domain), we believe it becomes too restrictive to also measure the data usage and impact aspects. Only when a good number of datasets are opened and have been subjected to data recipient examinations for a substantial period of time, it makes sense to take into account also the data usage aspects and even further to examine the relation between the degree of data openness (i.e., the aspect that we want to assess in this contribution) and the degree of data usage. This separation of concerns and the feasibility of the assessment method are instrumental in emerging Open Data domains (like the justice domain). Assessing new Open Data initiatives asks for a specific method (like ours) rather than a comprehensive method, as the latter requires, among others, a long observation period for appropriately assessing Open Data initiatives. Furthermore, in our daily practice within justice domain organisations we witness an urge to share (or open up) the justice domain data at this stage, without being concerned about or occupied with the holistic aspects of Open Data. Our method, therefore, aims at providing such a meaningful, to the point, and thus actionable method that assesses just the degree of data openness per dataset. Using our data openness assessment method, these organisations can determine their total degree of openness (i.e., for all datasets of an organisation). In this way, these organisations can show how much they are compliant with Open Data objectives.

There are also Open Data assessment methods that come close to our proposed method. The Global Open Data Index (GODI) developed by Open Knowledge International (GODI, 2018) is one of them, which uses the Open Definition (Open Definition, 2018) to set the principles along which the openness of data and the data content are assessed. Seeking a meaningful and actionable indicator, similarly to our method, the GODI tracks the openness state of government data in a given place by using a non-probability sampling technique known as ‘snowball sampling’ (GODI methodology, 2018). The indicators of the GODI are similar to ours, but they are measured at two levels per indicator, i.e., having a binary outcome per indicator. Our model allows more fine-grained levels per indicator. Furthermore, the GODI is used for another purpose than ours. The GODI intends to indicate if a published dataset adheres to Open Data requirements or not, i.e., to make a binary decision in being Open Data or not. A GODI score of less than one indicates that the corresponding dataset does not adhere to Open Data principles, thus GODI highlights the aspects upon which organisations may

improve their Open Data publications (GODI methodology, 2018). Specifically, the GODI does not show a linear increase of openness in the sense that, for example, how far an 80% openness is from the ideal/desired Open Data case. We aim at assessing the degree of openness for a dataset to indicate such a distance to the ideal Open Data case. For our purpose, the distance measure is sought through defining reasonable and meaningful indicators, the levels per indicator, and assigning appropriate values to those levels as illustrated in Section 5.2 by our case studies.

Another measure that comes close to our method is the 'data openness indicator' which is part of the 'e-government openness index' proposed in Veljković et al. (2014). We categorised this component indicator as a comprehensive model in the beginning of this section, comprising eight characteristics of Open Data from Open Government Working Group (2007), namely, being: complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary and licence free (Open Government Working Group, 2007). This component indicator is similar to GODI in the sense that

a it does not aim at indicating a distance to the ideal case of Open Data (although its three out of eight Open Data characteristics assume non-binary values, see Table 1 in Veljković et al., 2014)

b it is applied to a sampled subset of the opened datasets as the objective is to indicate the average degree of openness for a government (while ours is to indicate the total absolute vale of opened datasets, each of which is measured with a realistic distance to the ideal case of Open Data).

6.2 Added value

To clarify the relevancy and importance of the concept of Semi-Open Data, in general, and the degree of data openness, in particular, it is useful to consider them within the context of the maturity of e-government development. There are many studies proposing multiple-stage maturity models for developing e-government. A noticeable maturity model is by Lee and Kwak (2012), which consists of five levels: initial conditions, data transparency, open participation, open collaboration, and ubiquitous engagement. The two first levels are of particular interest for our paper. Level 1 is concerned with the initial conditions, focusing primarily on cataloguing and broadcasting information to the public with no or few metrics to assess public engagement. Level 2 is about data transparency, focusing on increasing transparency of government processes and performance by

a publishing relevant, high-value, and high-impact data online and sharing them with the public

b establishing data management functions as well as improving and assuring data quality in terms of accuracy, consistency, and timeliness.

We notice that most public organisations in the Netherlands, particularly in the justice domain, are in a transition state, moving from Level 1 to Level 2. At this state, lots of effort are put into improving the quality of high-value and high-impact data and share them to the public. As there is a no effective assessment metric to capture these efforts, moreover as the aforementioned binary methods capture only fully Open Data complaint efforts, we sense the need to devise a method that

a captures how far a data opening initiative lies from an ideal Data Open initiative in order to have a sense of distance

b provides the total accumulative distances of all the datasets shared by an organisation, thus to indicate the organisation's efforts at the current transition state.

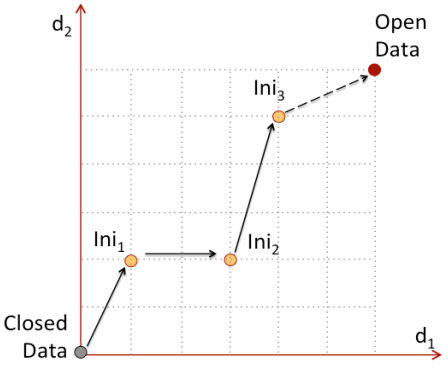
We defined a number of indicators of data openness, requiring them to be meaningful and actionable and further, we are careful not to define too many indicators as the curse of high dimensionality may result in meaningless distance measures (Bishop, 2006). Defining these levels per indicator and their values, which we proposed to be done in a close collaboration among all relevant stakeholders in a given domain, are considered out of the scope of this contribution. Creating consensus among all stakeholders is generally a complex and time-consuming process, like the process we witness in standardisations of products that are guided by, among others, both scientific facts and (business) interests. Consequently, we did not attempt to evaluate the proposed method based on a comprehensive evaluation like analytical, experimental, and testing (Hevner et al., 2004) in this contribution. We, however, evaluated the designed method based on the two types mentioned in Hevner et al. (2004), namely: Observational via 'Case Study' and descriptive via 'Informed Argument'.

We introduced an information sensitivity indicator that indicates whether a dataset can be shared with others (i.e., either be opened to the public or be shared with a group) or not. In the former case, i.e., the dataset can be shared with the public or a specific group, with or without transformation (e.g., via data minimisation to mitigate privacy risks). In such cases, the scope of data recipients and the amount of data transformation can be specified by indicator $I_{1,1}$ and indicator $I_{3,3}$, respectively. In this way, our model accommodates the data protection requirements imposed by privacy laws and regulations such as the General Data Protection Regulation (GDPR). The offered flexibility allows taking justice domain datasets into account within the landscape of Open Data initiatives as they are often shared with specific groups and/or as they undergo data minimisation transformations due to being highly privacy sensitive (Bargh et al., 2019b). Opening or sharing of such transformed datasets, therefore, are counted as Semi-Open Data initiatives as positioned in Bargh et al. (2016a).

6.3 Potentials

The proposed assessment method to measure the degree of data openness enables learning about and considering not fully Open Data initiatives. Further, such a measurement can make it possible to compare organisations in how far their Open Data initiatives have advanced, relative to one another and relative to the average (in a specific sector/domain). Longitudinal comparison of Semi-Open Data initiatives within a given context or organisation can also become more fine-grained using the assessment method proposed. Having such views on the progress of Open Data in time (i.e., a longitudinal view), as illustrated in Figure 2, and space (i.e., among peer organisations) can help organisations to improve their Open Data policies and strategies, e.g., by identifying (and focusing on improving) those Open Data indicators that can yield maximum use for these datasets.

Figure 2 A two-dimensional illustration of the progress of data opening initiatives, from Bargh et al. (2016a) (see online version for colours)



In this view, one should also measure how much an already opened dataset is used in practice in ex post evaluations (i.e., when enough time is elapsed after the release of a dataset) by using the so-called holistic methods, similarly to Donker and Loenen (2017). By putting the two measurement dimensions of data openness and data usage next to each other, one can also deduce how much data openness leads to maximum data use, thereby making possible a cost-benefit trade-off. Making this trade-off is a topic for our ongoing research. In addition to these dimensions (i.e., data openness and data use), one can also consider other dimensions like data governance, monetary costs imposed on taxpayers, impact of the opened data for making those trade-offs.

This study showed that assessing the degree of data openness is feasible. Further, the assessment method is flexible and extensible in the sense that it can be enhanced as needed (for example, by improving upon the indicators, the structure of the indicators, and the levels and their values per indicator). Through adapting the weighing factors of the indicators appropriately, see relation (3), the assessment method can also be tuned to various contextual situations and viewpoints. Fine-tuning of the assessment method and its weighing factors is subject of our future research. Nevertheless, the way that we carried out our case studies shows a way for constructing and fine-tuning the parameters of the method incrementally in a sense making way (i.e., based on consensus among parties/stakeholders involved in a given domain and based on iterative application of the method to a sufficient number of datasets in that domain). When these parameters are well defined and agreed upon by the experts/stakeholders in a domain, we foresee the possibility of applying some of them automatically to datasets.

6.4 Limitations

In this contribution we took the first step towards defining the degrees of openness for Semi-Data initiatives for and based on eight datasets from justice domain. This is not a one-shut and strictly top-down operation. It requires going through iterations, by reflecting on the results obtained and learning from the feedback given by practitioners, domain experts, and peers. Building upon this initial work, we should improve our data model (i.e., the appropriate levels and their values per indicator) by investigating the characteristics of more datasets.

Although the datasets/cases considered in this paper were (close to) Open Data in the justice domain, they could not be marked as Open Data in its traditional sense if one considers the scores the datasets received in Table 1. According to (Open data trend report, 2015), for example, a dataset was considered as Open Data if the dataset met all requirements of the definition of Open Data fully. Therefore, considering the scores in Table 1, the efforts that are spent to share the datasets that are investigated as case studies in this paper, could not be recognised within the scope of the current Open Data initiatives. While, according to the Semi-Open Data assessment method, one can easily see that these datasets are actually quite close to the ideal definition of Open Data.

In this work, we defined high-level indicators based on their logical similarities (who, how, what and what-not). It is for future research to validate the hierarchical structure proposed or to search for a more sensible and meaningful structure among the indicators so that a more insightful measure of semi-openness can be delivered. Further, some of the proposed indicators may affect each other, depending on how they are defined. For example, if data suppression is the processing carried out on the data, then data completeness is affected adversely. Such data processing means that the corresponding indicators may affect each other. Studying elaborated models that capture these restrictions and impacts is left for future studies.

Due to the fact that we considered only a few partially opened datasets, some of the indicators were ideally satisfied for all datasets. Therefore, in this study we did not (attempt to) identify the possible fine-grained levels for those attributes whose score was 1.00 for all the datasets considered (i.e., $I_{1,1}$, $I_{2,1}$, $I_{2,4}$, $I_{2,5}$, $I_{2,6}$, $I_{3,6}$ and $I_{4,1}$). For future research one should try to assign the appropriate levels and their values for these indicators as well. The need for improvement also holds for those levels that were defined and for their values that were assigned in this contribution. Eventually, after having a sufficient number of case studies, the Open Data community should reach consensus on (and standardise) the indicators, the indicator structure, and the indicator levels and values. This standardised data model can provide a uniform view on the openness of data opening initiatives in a given domain.

Assessing those datasets that are opened by other organisations (for example datasets DS_5, …, DS_8) is not sufficiently feasible if there is no information about how the (raw) data are collected, processed and opened. This implies that the information needed to specify the degree of openness should be provided as, for example, metadata in future (Semi) Open Data releases if they want their datasets to be considered as Semi-Open Data initiative fairly. The proposed assessment method, however, can be used optimally for those organisations who know the ins and outs of how a dataset is collected, processed, and (is going to be) opened/shared (for example, in case of the organisations that directly share their datasets with others).

7 Conclusion

In this contribution, we provided a method to assess the degree of openness for Semi-Open Data initiatives. This assessment model enables us to recognise and reward the extent of organisations' efforts to meet the Open Data requirements in a fine-grained way. This study showed that the proposed way of assessing the degree of data openness is feasible. Relying on the measurement results, it appeared that the Semi-Open Data concept exposes below the tip of the iceberg, i.e., provides a means to account for those

data opening initiatives that are close to Open Data. Without having the Semi-Open Data concept and without calculating the degree of openness of these initiatives it was impossible to recognise and acknowledge the (extent of the) existence of such data opening efforts.

We applied the method to a number of datasets from the justice domain and showed that the model is able to expose and quantify partially Open Justice datasets. This ability, in our opinion, is crucial for the justice domain as the opening of justice domain datasets in its ideal sense is often not possible due to, among others, information/privacy sensitivity, misinterpretation and misleading challenges. The case studies also showed that the proposed measurement method appears to be flexible and extensible. In particular, we proposed (and illustrated by means of our case studies) how to construct and fine-tune the parameters of the method incrementally in a sense making way, i.e., based on consensus among parties/stakeholders involved and based on iterative application of the method to a sufficient number of datasets in a given domain (in our case the justice domain).

Further, we argued that the proposed assessment method can be used to compare the data opening performances of organisations with each other and with themselves longitudinally. As such, the measurement method can guide organisations for opening their data and can enable them to move faster towards the ideals of Open Data. It is for our future research to carry out more case studies in order to fine-tune the assessment method and its weighing factors.

References

Alberts, B., Cicerone, R.J., Fienberg, S.E., Kamb, A., McNutt, M., Nerem, R.M., Schekman, R., Shiffrin, R., Stodden, V., Suresh, S., Zuber, M.T., Kline Pope, B. and Hall Jamieson, K. (2015) 'Self-correction in science at work', *Science*, Vol. 348, No. 6242, pp.1420–1422.

Bargh, M.S., Choenni, S. and Meijer, R. (2015) 'Privacy and information sharing in a judicial setting: a wicked problem', *Proc. of the 16th Annual International Conference on Digital Government Research (dg.o)*, ACM, Phoenix, AZ, USA.

Bargh, M.S., Choenni, S. and Meijer, R. (2016a) 'Meeting open data halfway: on semi-open data paradigm', *Proc. of the 9th International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, March, Montevideo Uruguay, pp 199–206.

Bargh, M.S., Choenni, S. and Meijer, R. (2016b) 'On design and deployment of two privacy-preserving procedures for judicial-data dissemination', *Government Information Quarterly Journal (GIQ)*, Elsevier, Vol. 33, No. 3, July, pp.481–493.

Bargh, M.S., Choenni, S. and Meijer, R.F. (2017) 'Integrating semi-open data in a criminal judicial setting', in Jiménez-Gómez, C.E. and Gascó-Hernández, M. (Eds.): *Achieving Open Justice Through Citizen Participation and Transparency*, pp.137–156.

Bargh, M.S., Choenni, S., Meijer, R.F. and Choenni, S. (2019a) 'A method for assessing the openness of semi-open data initiatives: applied to the justice domain', *Proc. of the 8th International Conference on e-Democracy (e-Democracy*), 12–13 December, Athens, Greece.

Bargh, M.S., Meijer, R., Vink, M., Schirm, W., van den Braak, S. and Choenni, S. (2019b) 'Opening privacy sensitive microdata sets in light of GDPR: the case of opening criminal justice domain microdata', *Proc. of the 20th Annual International Conference on Digital Government Research* (*dg.o*), 18–20 June, Dubai, UAE.

Bishop, C.M. (2006) 'Pattern recognition and machine learning', in Jordan, M., Kleinberg, J. and Scholkopf, B. (Eds.): *Information Science and Statistics*, Springer-Verlag, Berlin, Heidelberg.

Blank, M. (2019) *Open Data Maturity 2019*, Technical Report by the European Commission, Directorate-General of Communications Networks, Content and Technology, doi: 102830/073835, December, https://www.Europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf (Accessed 17 January, 2021)

Burwell, S.M., van Roekel, S., Part, T. and Mancini, D.J. (2013) *Open Data Policy – Managing Information as an Asset*, White House Memo on Open Data Policy, 9 May, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf (Accessed 17 June, 2018).

Capgemini (2015) *Open Data Maturity in Europe 2015. Insights Into the European State of Play*. Report from Capgemini Consulting, http://www.Europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n1_-_final.pdf (Accessed 5 February, 2018).

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T. and Ferro, E. (2018) 'Open data evaluation models: theory and practice', *The World of Open Data. Public Administration and Information Technology*, Vol. 28, Springer, Cham, https://doi.org/10.1007/978.3.319.90850.2_8 (Accessed 17 January, 2021).

Conradie, P. and Choenni, S. (2014) 'On the barriers for local government releasing open data', *Government Information Quarterly*, Vol. 31, pp.S10–S17.

Data set 1 (2019) From Website: https://www.wodc.nl/binaries/C%26R2015_tabellen-h03_tcm28-131083.xls (Accessed 15 August, 2019).

Data set 2 (2019) From Website: http://data.rechtspraak.nl/uitspraken/(Accessed 15 August, 2019)

Data set 3 (2019) From Website: http://www.ois.amsterdam.nl/download/321-fysieke-verloedering-naar-veiligheidsregios-2016.xlsx (Accessed 15 August, 2019).

Data set 4 (2019) From Website: https://www.politie.nl/mijn-buurt/misdaad-inkaart?geoquery=2274& distance=5.0 (Accessed 15 June, 2019).

Data set 5 (2019) From Website: https://data.gov.in/sites/default/files/datafile/crcCASc_2014_1.csv (Accessed 15 August, 2019).

Data set 6 (2019) From Website: Www.ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/tables/table-1 (Accessed 15 June, 2019).

Data set 7 (2019) From Website: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4510.02016?OpenDocument (Accessed 15 August, 2019).

Data set 8 (2019) From Website: http://www5.statcan.gc.ca/access_acces/alternative_alternatif?l=eng & keng=5.94& kfra=5.94& teng=Download%20file%20from%20CANSIM & tfra=Fichier%20extrait%20de%20CANSIM & loc=http://www5.statcan.gc.ca/cansim/results/cansim-2520081-eng-8979932552853622326.csv (Accessed 15 August, 2019).

Dodds, L. and Newman, A. (2015) *A Guide to the Open Data Maturity Model: Assessing your Open Data Publishing and Use*, Technical report from Open Data Institute (ODI), website http://theodi.org/guides/maturity-model (Accessed 5 February, 2018).

Donker, F.W. and van Loenen, B. (2017) 'How to assess the success of the open data ecosystem?', *International Journal of Digital Earth*, Vol. 10, No. 3, pp.284–306, http://doi.org/10.1080/17538947.2016.1224938 (Accessed 17 January, 2021).

Elena, S. and Mercado, G. (2019) *Open Justice: An Innovation-Driven Agenda for Inclusive Societies*. Ediciones SAIJ, May, Ciudad Autónoma, http://www.bibliotecadigital.gob.ar/items/show/2569 (Accessed 17 January, 2021).

Elena, S. and van Schalkwyk, F. (2017) 'Open data for open justice in seven Latin American countries', *Achieving Open Justice Through Citizen Participation and Transparency*, p.210.

EU Directive 2019/1024 (2019). Directive (EU) *2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the re-use of Public Sector Information*, PE/28/2019/REV/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG, Accessed 17 Jan 2021 (Accessed 17 January, 2021).

Franklin, M., Halevy, A. and Maier, D. (2005) 'From databases to data spaces: a new abstraction for information management', *ACM SIGMOD Record*, Vol. 34, No. 4, pp.27–33.

GODI (2018) *The Global Open Data Index, Tracking the State of Open Government Data*, website http://index.okfn.org/ (Accessed 5 February, 2018).

GODI methodology (2019) https://index.okfn.org/methodology/ (Accessed 5 February, 2018).

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004) 'Design science in information systems research', *MIS Quarterly*, Vol. 28, No. 1, pp.75–105.

Jiménez, C.E., Solanas, A. and Falcone, F. (2014) 'e-government interoperability: linking open and smart government', *Computer*, Vol. 47, No. 10, pp.22–24.

Jiménez-Gómez, C.E. (2017) 'Open judiciary worldwide: best practices and lessons learnt', *Achieving Open Justice Through Citizen Participation and Transparency*, p.1.

Kalidien, S., van Witzenburg, R. and Choenni, R. (2014) 'Developing, implementing, and evaluating a web interface in the field of e-government', in Yannacopoulos, D., Manolitzas, P., Matsatsinis, N. and Grigoroudis, E. (Eds.): *Evaluating Websites and Web Services: Interdisciplinary Perspectives on User Satisfaction*, Chapter, Vol. 4, pp.60–70.

Khan, S. and Foti, J. (2018) *Independent Reporting Mechanism: Aligning Supply and Demand for Better Governance, Open Data in the Open Government Partnership*, Report http://www.opengovpartnership.org/sites/default/files/IRMReport-OpenData.pdf (Accessed 5 February, 2018).

Lampoltshammer, T.J., Guadamuz, A., Wass, C. and Heistracher, T. (2017) *Openlaws. Eu: Open Justice in Europe Through Open Access to Legal Information*, Achieving Open Justice through Citizen Participation and Transparency, p.173.

Lee, G. and Kwak, Y.H. (2012) 'An open government maturity model for social media-based public engagement', *Government Information Quarterly*, Vol. 29, No. 4, pp.492–503, https://doi.org/10.1016/j.giq.2012.06.001 (Accessed 17 January, 2021).

McLachlin, B.P.C. (2014) *Openness and the Rule of Law*, https://www.barcouncil.org.uk/media/270848/jan_8__2014_-_12_pt.__rule_of_law_-_annual_international_rule_of_law_lecture.pdf (Accessed 15 June, 2018).

Meijer, R.F., van den Braak, S.W. and Choenni, R. (2020) *Criminaliteit En Rechtshandhaving 2019*, WODC.

Meijer, R., Choenni, S., Alibaks, R.S. and Conradie, P. (2013) 'Bridging the contradictions of open data', *Proceedings of the 13th European Conference on eGovernment,* Como, Italy, pp.329–336.

Meijer, R., Conradie, P. and Choenni, S. (2014) 'Reconciling contradictions of open data regarding transparency, privacy, security and trust', *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9, No. 3, pp.32–44.

Netten, N., Bargh, M.S., van den Braak, S., Choenni, S. and Leeuw, F. (2018) 'Legal logistics – a framework to unify data centric services for smart and open justice', in Ae Chun, S., Mellouli, S. and Arens, Y. (Eds.): *International Journal of E-Planning Research, the Special Issue of Models and Strategies Toward Planning and Developing Smart Cities*, Vol. 7, No. 2.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S J., … and Yarkoni, T. (2015) 'Promoting an open research culture', *Science*, Vol. 348, No. 6242, pp.1422–1425.

ODB (2018) *Website of Open Data Barometer*, http://opendatabarometer.org/and its methodology http://opendatabarometer.org/4thed./methodology/ (Accessed 5 February, 2018).

ODRN (2013) *Open Data Research Network* (*ODRN*), http://www.opendataresearch.org/content/2013/566/early-insight-2-definitions-open-data (Accessed 15 June, 2018)

OECD (2018) *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/9789264305847-en (Accessed 17 January, 2021).

Open data trend report (2015) *Technical Report by The Netherlands Court of Audit*, The Hague (in Dutch: Algemene Rekenkamer, Trendrapport open data 2015, Den Haag).

Open data trend report (2016). *Technical Report by The Netherlands Court of Audit*, The Hague (in Dutch: Algemene Rekenkamer, Trendrapport open data 2016, Den Haag).

Open Definition (2019) *Website of Open Definition*, http://opendefinition.org (Accessed 5 February, 2018).

Open Government Working Group (2007) *Open Government Data Principles*, https://public.resource.org/8_principles.html (Accessed 12 October, 2018).

Rorissa, A., Demissie, D. and Pardo, T. (2011) 'Benchmarking e-government: a comparison of frameworks for computing e-government index and ranking', *Government Information Quarterly*, Vol. 28, No. 3, pp.354–362.

Ruijer, E., Grimmelikhuijsen, S. and Meijer, A. (2017) 'Open data for democracy: developing a theoretical framework for open data use', *Government Information Quarterly*, Elsevier, Vol. 34, No. 1, pp.45–52.

Sandoval-Almazan, R. and Gil-Garcia, J.R. (2016) 'Toward an integrative assessment of open government: proposing conceptual lenses and practical components', *Journal of Organizational Computing and Electronic Commerce*, Vol. 26, No. 1–2, http://www.tandfonline.com/doi/pdf/10.1080/10919392.2015.1125190 (Accessed 17 January, 2021).

Sandoval-Almazan, R. and Gil-Garcia, J.R. (2018) 'Understanding e-justice and open justice through the assessment of judicial websites: toward a conceptual framework', *Social Science Computer Review*, https://doi.org/10.1177/0894439318785957 (Accessed 17 January, 2021).

Smit, P.R. and Dijk, J.V. (2014) 'History of the dutch crime victimization survey(s)', in Bruinsma, G. and Weisburd, D. (Eds.): *Encyclopedia of Criminology and Criminal Justice.* Springer, New York, NY.

van den Braak, S., Choenni, S. and Verwer, S. (2013) 'Combining and analysing judicial databases', *Discrimination and Privacy in the Information Society*, Springer, Berlin, Heidelberg, pp.191–206.

van den Braak, S., Choenni, S., Meijer, R. and Zuiderwijk, A. (2012) 'Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector', *Proc. of the 13th Annual International Conference on Digital Government Research (dg.o)*, ACM, pp.135–144.

van den Hoogen, H. (2013) 'Taking care of research data: outcome of a seminar on managing data integrity in research', *Library Connect Newsletter, Issue: Research Data Management*, Vol. 1, No. 1, p.18 March.

Van der Leij, J.B.L. (2016) 'Het nederlandse strafrechtssysteem', in Kalidien, S.N. (Ed.: *Criminaliteit En Rechtshandhaving*, WODC Cahier 2017-12.

van Dijk, J., Choenni, S., Leertouwer, E., Spruit, M. and Brinkkemper, S. (2013) 'A data-space system for the criminal justice chain', *Proc. of On the Move to Meaningful Internet Systems (OTM) Conferences,* Springer, Berlin, Heidelberg, pp.755–763.

van Knippenberg, L. (2020) *Open Data Maturity 2020*, Technical Report by the European Commission, Directorate-General of Communications Networks, Content and Technology, doi: 102830/629187, 15 December, https://www.Europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf (Accessed 17 January, 2021).

Veljković, N., Bogdanović-Dinić, S. and Stoimenov, L. (2014) 'Benchmarking open government: an open data perspective', *Government Information Quarterly* (*GIQ*), Vol. 31, No. 2, pp.278–290, https://doi.org/10.1016/j.giq.2013.10.011 (Accessed 30 October, 2018)

Verhulst, S., Noveck Simone, B., Robyn, C., Brown, K. and Paz, C. (2014) *Released: The Open Data Era in Health and Social Care*, http://www.thegovlab.org/static/files/publications/nhs-full-report.pdf (Accessed 15 June, 2018).

Wonderlich, J. (2010) *Ten Principles for Opening Up Government Information*, Website of Sunlight Foundation, http://sunlightfoundation.com/policy/documents/ten-open-data-principles/ (Accessed 15 June, 2018).

Zuiderwijk, A. and Janssen, M. (2014) 'Open data policies, their implementation and impact: a framework for comparison', *Government Information Quarterly* (*GIQ*), Vol. 31, No. 1, pp.17–29.

Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, S., Charalabidis, Y. and Jeffery, K. (2012) 'Issues and guiding principles for opening governmental judicial research data', *International Conference on Electronic Government (EGOV*), September, Springer, Berlin, Heidelberg, pp.90–101.