How open is the Open Government Data published by Dutch provinces?

Brinke, Mark ten

2023

How open is the Open Government Data published by Dutch provinces?

MASTER THESIS – INTERNATIONAL AND EUROPEAN GOVERNANCE

MARK TEN BRINKE

THESIS SUPERVISOR: DR. M. M. YOUNG
09-06-2023
WORD COUNT: 9632

Table of contents

Abstract

Governments worldwide have been increasingly making Open Government Data (OGD) available. Arguments for OGD include improvements to transparency, accountability, policymaking, innovation, and economic growth while reducing corruption and the negative effects of New Public Management's outsourcing. However, research shows that not all published OGD are open according to the eight Sebastopol principles of open data. This research uses the ordinary citizen test, based on the Sebastopol principles, to determine the extent of openness of OGD published by Dutch provinces. This research uses the metadata of all published datasets on the Dutch national data portal. The findings show that around 30% of the assessed datasets qualify as open. The most common barrier to the openness of OGD published by Dutch provinces is that datasets are only available in proprietary formats. The OGD of Dutch provinces not qualifying as open can have implications for the perceived benefits of publishing OGD in the Netherlands.

Introduction

Open Government Data (OGD) are " a philosophy- and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all" (Organisation for Economic Co-operation and Development, n.d.). Stimulating governments to make their data freely available is not a new movement. Since 2009, governments have started initiatives to publish OGD through data portals (Gigler et al., 2011). An important event in starting the OGD movement was the publication of eight principles of OGD by thirty open government activists in 2007, this was later followed by the signing of the Open Data Charter by leaders of the G8 in 2013 (Wang & Shepherd, 2020).

The OGD initiatives can help satisfy the three cornerstone principles of an open government: transparency, participation, and collaboration (White House, 2009). Hence, opening government data is not an objective of its own, but it can stimulate the involvement of citizens to challenge and assist governments (Wang & Shepherd, 2020). Involving citizens in governance by publishing OGD can yield comprehensive benefits for governments, including increased transparency and accountability, economic growth, strengthening citizen participation, and improved provision of public goods and services (Gigler et al., 2011). Furthermore, publishing OGD can improve policy-making, stimulate innovation (Janssen et al., 2012), and improve government decisions (Puron-Cid, 2012).

Although publishing OGD has many benefits, the majority of available OGD do not qualify as open. The Global Open Data Index measures the openness of government data for 94 countries and shows that only 11% of dataset entries in the index are open (n.d.-a). Additionally, research into the openness of OGD in the United Kingdom (UK) found that only one out of ten datasets available on the UK's data portal is open according to the eight principles of OGD (Wang & Shepherd, 2020). The low amount of OGD qualifying as open data might prove problematic since the benefits of publishing OGD are based on the notion that data are actually open.

This research evaluates the openness of OGD published by Dutch provinces. There are three crucial dimensions in assessing the value of OGD to open government, these being; the extent of openness of published OGD, the contribution of OGD to open government based on the information contained in datasets, and the response of a government to citizens' opinions (Wang & Shepherd, 2020). The first dimension evaluates whether published OGD complies with the definition of open data. The second dimension refers to the value of topics covered by OGD to realise open government. Some OGD topics allow the government to create an illusion of open government without actually opening the government further. The last dimension relates to the willingness of governments to listen to feedback on OGD from society (Wang & Shepherd). This research focuses on the first dimension by answering the research question: To what extent do the OGD published by Dutch provinces qualify as open data?

To answer the research question, the metadata of all datasets published by Dutch provinces on the national data portal will be assessed based on open data principles. This approach will provide insights into the openness of Dutch provinces' OGD and show the most common barriers to openness. This research will proceed with a background chapter to present an overview of the theoretical motivations for publishing OGD, definitional and measurement frameworks regarding OGD, previous research on OGD, and the landscape of OGD in the Netherlands. The methodology chapter explains the research methods used, followed by the analysis. The discussion chapter reviews the main findings of this research, and the research will end with a conclusion.

Background

Theoretical motivations for Open Government Data

Transparency

There are different arguments for making government data freely available to the public. One of the main drivers for publishing Open Government Data (OGD) is creating more transparency (Bertot et al., 2010; McDermott, 2010). Transparency is defined as the "lack of hidden agendas and conditions, accompanied by the availability of full information required for collaboration, cooperation, and collective decisionmaking" (Farrell, 2016, p.445). More transparency can help to increase citizen participation, promote e-governance, and facilitate e-government (Halachmi & Greiling, 2013). Creating transparency can also be seen as a normative objective because better informing the public can improve the legitimacy of government decisions and trust between the government and the public (Wanna, 2018; Alessandro et al., 2021). Furthermore, improving transparency can help reduce corruption and prevent the abuse of authoritative or coercive powers that come with certain positions within the government (Bertot et al., 2010; Stamati et al., 2015).

Creating more transparency can also have negative consequences for governments. These include; disadvantages during negotiations with less transparent states, discouraging actors from delivering frank and fearless advice to the government, risk of simulated transparency and political spin, reduced government flexibility, and obligation to divulge confidential

information (Wanna, 2018). Although improving the government's transparency possibly comes with certain risks and unwanted consequences, the literature mainly focuses on the positive effects of creating transparency and disclosing information to the public (Wanna, 2018; Matheus & Janssen, 2020).

Improving a government's transparency, specifically by publishing digital data in the form of OGD, has other positive implications. First, those involved in government processes usually have more information than outsiders, creating an information asymmetry between the government and the public (Matheus & Janssen, 2020). Publishing OGD enables the public to gain insight into the performance of their governments (McDermott, 2010). Therefore, the transparency created by publishing OGD helps to overcome this information asymmetry. Additionally, opening government data, thereby providing transparency into governments' actions, potentially improves accountability (Helbig et al., 2010).

The benefit of transparency mentioned most often in the literature is improved accountability (Matheus & Janssen, 2020). Accountability is "the obligation to present an account of and answer for the execution of a set of responsibilities" (Lewis et al., 2014, p.401). The government's accountability to the public is a crucial factor in democracies to ensure appropriate behaviour and improve the performance of governments (Schillemans, 2016). Furthermore, accountability has become increasingly crucial in regulating, coordinating, and controlling government agencies (Schillemans, 2016). Although more accountability is considered positive, it can also have adverse effects. Accountability can cost public managers too much of their time, lead to a fixation on short-term success, and lead to behaviour where hitting the target is more important than acting in the public interest (Schillemans, 2016). Although the literature mainly endorses the positive relationship between transparency and accountability, there is still discussion about whether more transparency actually increases accountability (Park & Gil-Garcia, 2017).

Digital-Era Governance

Another important argument for governments to start publishing OGD concerns digital-era governance (DEG). New Public Management (NPM) was the dominant government management system for multiple decades (Dunleavy et al., 2006). NPM focused on realising managerial change by importing concepts from the corporate world into the government. The main themes in NPM ideology have focussed on splitting up public sector hierarchies, creating competition among providers, and introducing performance incentives for staff (Dunleavy et al., 2006). The popularity of NPM has decreased after key parts of the ideology led to policy disasters. These failures created an opportunity for a shift towards DEG, which is now the main framework for promoting government efficiency (Dunleavy et al., 2006; Margetts & Dunleavy, 2013).

The three main themes of DEG are reintegrating components of the government that NPM separated, simplifying and changing the relationships between agencies and clients, and implementing profound digital changes within the government (Dunleavy et al., 2006). One component of the 'implementing profound digital changes' theme relevant to OGD is radical

disintermediation. This component of DEG grants citizens, businesses, and other civil society actors direct access to state systems (Dunleavy et al., 2006). Disintermediation allows the public to decide what they want to receive from the government, thereby creating a form of 'do-it-yourself' government (Dunleavy et al., 2006; Margetts & Dunleavy, 2013). NPM advocated outsourcing contracts to the private sector, this outsourcing of public goods and services to private contractors often led to cost-cutting at the expense of quality (Dunleavy et al., 2006). Citizens may care about this decline in public goods and services quality but lack the evidence to act against it (Dunleavy et al., 2006).

Publishing OGD can satisfy the radical disintermediation component of DEG by allowing the public to gain direct access to government data. This also provides the public with evidence to counter the adverse effects of NPM's outsourcing on the quality of public goods and services. Making this data directly available to everyone has more positive implications for society since it improves the ability of policymakers to devise effective policies, stimulates innovation, increases economic growth (Janssen et al., 2012), and improves governments' decisions (Puron-Cid et al., 2012). Altogether, publishing OGD helps governments shift towards DEG through radical disintermediation, which counteracts the negative consequences of outsourcing and has positive effects on society as a whole.

The two theoretical motivations for publishing OGD that are most salient for this research are improving transparency and satisfying the radical disintermediation component of DEG. Increasing transparency and radical disintermediation through publishing OGD has positive effects on society by offering the public insight into government actions, but their actual effects are different. Transparency can improve trust in government, reduce corruption, prevent the abuse of authoritative or coercive powers, help overcome the information asymmetry between the government and the public, and increase accountability. Whereas radical disintermediation can counteract the negative effects of outsourcing on the quality of public goods and services, helps policymakers to devise effective policies, stimulates innovation, increases economic growth, and improves governments' decisions. The positive effects of OGD are abundant but only occur when published government data qualify as open data. Therefore, this research tries to determine to what extent the data published by Dutch provinces meet the requirements of OGD.

Defining Open Government Data

Open data are defined as data that can be freely used, modified, and shared by anyone for any purpose (Open Knowledge Foundation, n.d.-a). However, there are several specific requirements for data to qualify as open. First, data must be published in the public domain or issued under an open license. Data also must be provided in their entirety at no more than a reasonable one-time production cost and preferably be downloadable without charge. Furthermore, data have to be provided in an open format which entails that there are no restrictions on its use and that they are immediately readable by a computer where separate data elements can be easily accessed and modified. The last condition for data to be considered open is aimed at the associated use license. The license must allow free use,

redistribution, modification, separation, compilation and propagation of the data whilst also being non-discriminatory (Open Knowledge Foundation, n.d.-b).

Open Government Data are government-produced data that meet these definitional standards of openness. Publishing OGD makes governments more transparent and more democratic to the general public by making government data free to use by anyone (Kučera et al., 2013). Additionally, reusing open data in products and services can yield economic benefits for the government. These ideas on the benefits of publishing open data are also shared by the European Union (EU). The EU argues that publishing data can benefit citizens, businesses, and public administrations while deepening European cooperation and serving as an essential resource for economic growth, job creation, and societal progress (data.europa.eu, 2020).

National governments have also recognised the potential benefits of OGD. Since 2009 many governments, predominantly in Europe and North America, have initiated OGD strategies (Gigler et al., 2011). The EU endorses this growth in publishing OGD. This growth is reflected by the rapid growth in available datasets on the European Union Portal, which grew from over 400.000 in 2016 to over 1.500.000 in 2023. The EU expects this growth to continue (data.europa.eu, 2020; data.europa.eu, n.d.).

Frameworks of open data

Sebastopol principles

There are different starting points to assess whether data should be considered open data. The first framework was established in 2007 in Sebastopol when a working group formulated eight principles for Open Government Data, generally known as the Sebastopol principles.

The first principle is that data must be complete (1), all public data should be made available as long as it is not subject to privacy, security or privilege limitations. Open data should also be primary (2), data should be made available as it was collected at the source with the greatest amount of granularity and not a modified version of it. Furthermore, data must be timely (3), meaning that it is made available as soon as possible to protect its value. The fourth principle is that data must be accessible (4), which entails that data are obtainable by the broadest range of users as possible but also for the most comprehensive range of purposes. Data should also be machine-readable (5), meaning that data are structured to allow for automated processing. Sixth, data must be non-discriminatory (6) by being available to anyone without registration. Open data must be non-proprietary (7) in the sense that no entity has exclusive control over the data. Finally, data should be license-free (8), which means that data are not subject to copyright, patent, trademark or trade secret regulation (Opengovdata, n.d.).

The list of principles was expanded in 2010 to ten principles by adding permanence (9) and usage costs (10). Permanence entails that the published information remains available over time, whereas the usage costs principle means that data should be accessible free of cost

(Sunlight Foundation, n.d.). Only when data adhere to all these ten principles can it be considered open data according to the Sebastopol principles.

International Open Data Charter

In 2013 the G8 leaders signed the Open Data Charter. This charter contains five principles that the G8 countries will act on to support innovation, provide more accountability and utilise the economic potential of open data. The five principles mentioned in the charter are open data by default, quality and quantity, useable by all, releasing data for improved governance and releasing data for innovation (Cabinet Office, 2013). Many other nations also welcomed the adoption of this charter, but there was a general perception that the principles should be improved to support wider global adoption. This eventually led to the launch of the International Open Data Charter with six principles on open data in 2015 at the United Nations General Assembly (The Open Data Charter, 2022-a).

The first principle, that data should be open by default (1), remained the same as in the original charter entailing that governments must justify withholding rather than releasing data. Data must also be published quickly and comprehensively (2) to ensure that it is relevant when published. The third principle is that data must be accessible and useable (3) by ensuring it is machine-readable and free of charge. Furthermore, data should be comparable and interoperable (4) since the availability of more quality datasets increases the potential value that the public can get out of it. The last two principles are not necessarily directives for open data but are aimed more at the results of publishing open data. The fifth principle is improved governance and citizen engagement (5). The idea behind this principle is that open data can help improve transparency, thereby improving public services and holding governments accountable. The final principle is inclusive development and innovation, which entails that open data can stimulate inclusive economic development by improving the efficiency of specific policy areas (The Open Data Charter, 2020).

Since the launch of this International Open Data Charter, it has been adopted by 95 governments, both national and subnational. The Dutch government has not yet adopted the charter (The Open Data Charter, 2022-b).

This research will use the Sebastopol principles in the analysis because this framework provides more tools to analyse single datasets than the International Open Data Charter. The International Open Data Charter takes a broader approach by considering the impact of publishing OGD on governance, citizens, and the economy. In contrast, the Sebastopol principles are specific requirements for publishing open data. The broader effects of publishing OGD on society are not part of this research, making the Sebastopol principles a better fit. The Sebastopol principles used to determine the openness of datasets are; accessible, machine-readable, non-discriminatory, non-proprietary, and license-free.

Measuring the ‘openness’ of open data

Global Open Data Index

Several approaches exist to measure precisely how ‘open’ OGD are. The first is the Global Open Data Index framework, which tries to assess the quality of OGD from a citizens' perspective by using experts in the domain to review datasets. The index enables stakeholders to follow the government’s progression in releasing open data and allows the government to receive user feedback. The Global Open Data Index does not cover all types of government datasets. Instead, this index concerns budget, spending, procurement, election results, company register, land ownership, national maps, administrative boundaries, locations, national statistics, draft legislation, national law, air quality and water quality data. Data from the different categories are examined using questions that determine their openness based on the Open Definition and the Open Data Charter.

This index evaluates datasets by scoring them with a maximum score of 100. A maximum of 40 points is given for open license, machine-readable and open file formats. The other 60 points comprise timely publication, data availability and accessibility. The eventual score is based on the following questions: Are the data available online without the need to register or request access to the data? (15 points); are the data available free of charge? (15 points); are the data downloadable at once? (15 points); are the data up to date? (15 points); are the data openly licensed/in the public domain? (20 points) and are the data available in open and machine-readable formats? (20 points).

The eventual scores of datasets have different implications for the openness of data. Data are considered open when it scores 100 on the Global Open Data Index. A score of up to 80 indicates public data. These data meet most conditions of open data but are not available in a machine-readable format. A score of up to 85% can indicate access-controlled data, which means that users have to register before they can download data. A score of zero is considered a data gap. This entails that governments do not produce data on this category. The Global Open Data Index is useable to determine the openness of data but does not assess the quality of datasets (Global Open Data Index, n.d.-b).

5-star rating system

Tim Berners-Lee introduced a different approach to rating the openness of open data. He developed a rating system for Linked Open Data to help governments publish ‘good’ Linked Data (Berners-Lee, 2010). The idea behind Linked Data is that individuals can use available data to find other data that is related to it. According to Berners-Lee, Linked Open Data is Linked Data that are available under an open license and reusable without further costs. Not all Linked Data have to be open, but this rating system is particularly made to assess the openness of Linked Data when it claims to be open. For data to be rated higher in this system, it must also meet all the preceding requirements.

Linked Open Data receive a 1-star score when it is available on the internet under an open license. To reach a 2-star score, the data must also be available as machine-readable

structured data. A 3-star score is reserved for data with a non-proprietary format, and data worthy of a 4-star score must also adhere to the open standards of the World Wide Web Consortium. Finally, a 5-star score is given to data that comply with all the other requirements and are also linked to other people's data (Berners-Lee, 2010).

Open Data Barometer

The Open Data Barometer determines how governments publish and use open data for accountability, innovation and social impact. The last report of the Open Data Barometer looks at the 30 governments that have adopted the International Open Data Charter Principles or the G20 Anti-Corruption Open Data Principles. The score of the Open Data Barometer is based on three kinds of data, first a peer-reviewed expert survey and a detailed dataset survey, second a self-assessment survey by governments and finally secondary data to supplement the expert survey data. These data are computed to a score between 0 and 100 for every government in this research, where a higher score indicates a better performance regarding the publishment and use of open data (Open Data Barometer, n.d.). The Open Data Barometer scores are insightful concerning governments' performance in publishing and using open data. However, the framework is not useable to determine the openness of individual datasets.

Ordinary citizen test

A study by Wang and Shepherd on the extent of openness of OGD in the United Kingdom (UK) constructed its own measurement framework, the ordinary citizen test (2020). They argue that the existing frameworks are all flawed in important dimensions and are therefore not useful to determine the openness of datasets from a citizen perspective. To illustrate, the 5-star system focuses on advanced aspects of open data that are not understandable to everyone, and there is also no attention to the granularity and permanence of data, whereas the Global Open Data Index and the Open Data Barometer lack specific attention to granularity (Wang & Shepherd, 2020). The ordinary citizen test is mainly based on the eight Sebastopol principles except for the completeness principle since citizens cannot determine whether the OGD provided are indeed complete or that information is missing.

The first dimension of the ordinary citizen test is granularity (1), this dimension assesses data based on a granularity scale ranging from reported to granular data, where more granular data correlate with more user value. The next dimension is recency (2), data are considered to be open when it was published in the previous 30 months. The dimension published (3) assesses whether data are published since some advertised datasets are unavailable. The fourth dimension is operable (4), which determines if the links are working, data are downloadable, and opening errors are absent. Machine-readable (5) means that data are open when computers can read and process the format. The sixth dimension is non-discriminatory (6), which requires data to be available without requiring permission or filling out registration forms. The non-proprietary (7) dimension is aimed at the data format where open data require an open software format. The final two dimensions are unrestricted license (8) and cost-free (9), which consider data as open when they have an unrestricted license and when they are available free of cost respectively (Wang & Shepherd, 2020).

Statistical Performance Indicators

A final approach to measuring the openness of open data is the Statistical Performance Indicators framework established by the World Bank. This framework consists of five pillars and 22 dimensions that can assess the maturity of national statistical systems. It differs from the other frameworks introduced since it does not measure the quality of individual datasets but considers factors that provide the foundation for publishing high-quality data in countries (Open Data Watch, 2018).

The first pillar of this framework is Data Use (1) which examines the demand side of the statistical system. The second pillar is Data Services (2), this pillar identifies information on the release of data and online access. The Data Products (3) pillar looks into relevant data indicators related to the United Nations' Sustainable Development Goals. The fourth pillar is Data Sources (4) and focuses on censuses, surveys, admin data and geospatial data. The last pillar is Data Infrastructure (5) which identifies standards and methodology used by countries in the classification of data. Of the 22 dimensions established, only 14 are used by the World Bank since there is no suitable data to measure the other eight dimensions (World Bank, n.d.). Although this framework assesses relevant factors regarding the foundation for publishing open data in countries, it cannot determine how 'open' single datasets are.

Previous research on Open Government Data

Most research on OGD develops and tests models for open government and OGD quality, describes potential benefits and barriers to publishing OGD, or evaluates the OGD portals from governments. To mitigate the harms of corruption and violations of fundamental human rights, modern democratic societies must monitor their government to improve transparency (Khurshid et al., 2020). This has led to a societal demand for access to public data (Khurshid et al., 2020). Although there is a societal demand for access to public data and publishing OGD can improve a government's transparency, efficiency, and effectiveness, over 90% of governments are not considered open (Hossain et al., 2021).

The literature identifies several barriers to the publishing of OGD. First, insufficient technical infrastructure and a lack of skilled professionals hinder the ability of governments to publish OGD (Hossain et al., 2021). Furthermore, politicians acting defensively (Grimmelikhuijsen, 2012), a lack of awareness of the nature and value of OGD (Zhao & Fan, 2018), and privacy and security considerations (Janssen & van den Hoven, 2015) can all be potential barriers to governments publishing OGD. Finally, existing legislation and ambiguous policies on the subject of OGD can also prevent governments from opening their data (Janssen et al., 2012).

The success of OGD initiatives depends on multiple factors. One of these is the system used to make OGD publicly accessible, often called 'portals'. Many governments use OGD portals to publish their datasets, but the usability of these portals varies significantly between countries (Nikiforova & McBride, 2021). Constructing these portals is relatively easy for governments and can be a showcase for public transparency at a low initial cost (Abella et al., 2022). The problem is that the data behind these portals are often poorly structured,

inconsistent in quality and quantity, and difficult to reuse (Abella et al., 2022). In most cases, the technical aspects of OGD portals have been implemented well, but the actual data published on these portals lack usability, usefulness, and quality (Nikiforova & McBride, 2021). Although setting up a data portal to make OGD publicly accessible is relatively easy, the data available through these portals are often inadequate.

Another factor that influences the success of OGD initiatives is the quality of OGD themselves. Publishing data without some form of quality control, usually absent in OGD, can threaten both the reusability of datasets and public participation but also increases the costs of obtaining and interpreting the data (Vetrò et al., 2016). Additionally, the low quality of both data and metadata of OGD published by governments threatens the value users can gain from these data (Sadiq & Indulska, 2017). Furthermore, publishing low-quality OGD can result in inaccurate reports on a government's performance (Lee & Kwak, 2012), damage its reputation, and lead its users to wrong conclusions (Kubler et al., 2018). A study on the quality of OGD from the UK, the USA, and South Korea shows that over 50% of OGD published by the UK and the USA are not machine-readable, making it harder for users to work with the data (Yi, 2019). On top of that, incomplete OGD are present in all three countries, which also decreases its quality (Yi, 2019). Although governments are opening more data, the quality of these data is often insufficient, jeopardising its user value.

The final factor impacting OGD initiatives is the openness of data. The benefits of publishing OGD are improved transparency and accountability, increased trust in government, more effective policymaking, stimulated innovation, and economic growth. These benefits are based on the fact that published OGD are actually open. Among scholars, there is no consensus on which approach to measuring openness should be used, but most frameworks are either based on the 5-star rating system from Berners-Lee or the Sebastopol principles (Attard et al., 2015). A study on the open data portals of Australia's largest cities uses machine readability to determine the openness of datasets (Chatfield & Reddick, 2017). Since this research looks at many aspects of open data portals and machine readability is the only indicator that concerns the openness of data, it is not that relevant regarding the openness of OGD. Another study tried to set up a framework to assess the quality of OGD, including the 5-star rating system from Berners-Lee and the timely (3) Sebastopol principle (Vetrò et al., 2016). Although this comes closer to measuring the openness of datasets, only a small part of the whole framework is aimed at determining the openness of OGD.

An article that comes close to what this research tries to accomplish is a study on the openness of OGD in the United Kingdom by Wang and Shepherd (2020). They examined a sample of 400 datasets derived from the dataset portal of the United Kingdom's government. Wang and Shepherd's findings show that approximately only one out of ten datasets published by the government is open, which is remarkable because the United Kingdom is generally seen as a global leader in an OGD context, according to their study. Over half of the examined datasets contain aggregated data instead of granular data, which is the main barrier to the openness of OGD in the UK, the second most common barrier is aged data (Wang & Shepherd, 2020).

The study also found that none of the previously identified measurement frameworks are adopted by other open data researchers. Furthermore, they argue that the 5-star rating system is flawed when used to determine the openness of data since there is no attention to dimensions like granularity and timeliness. The authors claim that the Global Open Data Index and the Open Data Barometer are more suitable for determining the openness of data since they are rooted in the Sebastopol principles. However, there is still insufficient attention to granularity within these two frameworks. Since no measurement framework is available that assesses all the aspects of OGD, Wang and Shepherd established and used the ordinary citizen test to determine the extent of openness (2020).

Open Government Data in the Netherlands

The global shift towards publishing more OGD is also visible in the Netherlands. The Dutch government launched its data portal, data.overheid.nl, where data from over 150 different government organisations are made publicly available (data.overheid.nl, n.d.-a). Furthermore, the national government has formulated different laws concerning OGD. The Dutch access to information law, called *wet open overheid* (WOO), was introduced in 2022 as a successor to the *wet openbaarheid van bestuur* (WOB) and arranges that information is published by the government both actively and when requested, provided that it does not harm certain specified interests (Digitale Overheid, 2023). The law on the reuse of government information, wet hergebruik van overheidsinformatie (Who), was introduced in 2015 and arranges that citizens and companies can file a request to reuse government information (Rijksprogramma Voor Duurzaam Digitale Informatiehuishouding, 2020). Where the WOO regulates the legitimacy and controllability of the government, the WHO is focused on the economic forces behind sharing data (data.overheid.nl, 2020).

The national government also formulated specific requirements for the metadata accompanying OGD. According to a guide on publishing data drafted by data.overheid.nl and the *Kennis- en Exploitatiecentrum Officiële Overheidspublicaties* (KOOP), metadata describe the data behind datasets by providing information about the structure, meaning, and validity of data (z.d.). All published OGD in the Netherlands must contain metadata compliant with the DCAT-AP-DONL 1.1 standard (KOOP, z.d.-a). Following this standard, 47 metadata values can be provided by the publisher. Out of these 47 values, 11 are required before the dataset can be published on the national data portal and providing the other 36 is optional (KOOP, z.d.-b). The 11 required values are aimed at describing the structure of data, describing in what way data are available and how they can be used, and improving the findability of data (data.overheid.nl & KOOP, z.d.).

Publishing OGD is not just reserved for the Dutch national government and its ministries. Lower levels of government, like the provinces and municipalities, also make open data available. The national government has initiated this by organising meetings to formulate a list of high-value datasets together with the provinces and municipalities. The idea behind these meetings is to help these lower levels of government to start opening their data and help

them to prioritise what data to open first. These meetings led to a list of high-value datasets for the provinces in 2019 and the municipalities in 2016 (data.overheid.nl, n.d-b; data.overheid.nl, n.d.-c). A dataset is considered high value when it contributes to transparency, the execution of the legislative task, cost reduction, servicing the target audience, and if there is a potential to reuse the data (data.overheid.nl, n.d.-b). These high-value datasets are gathered per subject and made available through the national government's data portal. Some provinces have their own data portals where open data are published, but the same datasets are also available through the national government's data portal.

Methodology

Data collection

A quantitative research method is used to determine the extent of openness of Open Government Data (OGD) published by the Dutch provinces. This research will analyse every dataset published by Dutch provincial governments on the national data portal before April 25, 2023, to determine whether these files meet the accessible, machine-readable, non-discriminatory, non-proprietary, and license-free Sebastopol principles. Data for this project were collected using the Application Programming Interface (API) protocol of data.overheid.nl via code written in Python 3.7. The data collected are "metadata" – data used to describe the characteristics of a dataset. For example, the format of a dataset (e.g., CSV) constitutes a piece of metadata about the dataset. The different metadata and the information they contain are described on the CKAN extension DCAT-AP-DONL website (ckanext-dcatdonl, z.d.). In total, metadata for 7.423 datasets published by provinces were collected. Analysis of all data was done using SPSS 27. The code used to access the API and the SPSS syntax used to perform the analysis can be found in the appendix.

Four metadata were used to determine whether each dataset meets the definitional standard of open data. These include: dataset_status, access_rights, license_title, and resources. All variables are computed into dummy variables that indicate whether the dataset qualifies as open in that specific research dimension. The resources variable contains information about the format of datasets which was first extracted and computed into a dummy variable later. Descriptive statistics show what percentage of datasets adhere to the specific research dimensions for both provincial governments as a group and per province.

Ordinary citizen test

This research is loosely based on the ordinary citizen test introduced by Wang and Shepherd to determine whether datasets qualify as open (2020). The five-star rating system from Berners-Lee has too much attention to linked data, and the Global Open Data Index uses data unavailable to this research. Furthermore, the Open Data Barometer and Statistical Performance Indicators offer insights into the landscape of everything surrounding OGD within a country but cannot determine the openness of datasets. The ability to determine the extent of openness of single datasets from an ordinary citizen's perspective while also considering the Sebastopol principles and showing what principles are insufficiently satisfied when data do not qualify as open, makes the ordinary citizen test the best fit for this research.

Table 1
Research dimensions per Sebastopol principle

Sebastopol principles		Research dimensions	Variable
Complete	(1)	N/a	
Primary	(2)	N/a	
Timely	(3)	N/a	
Accessible	(4)	Available on data portal	Dataset_status
Machine-readable	(5)	Machine-readable	Resources (format)
Non-discriminatory	(6)	Available without restrictions	Access_rights
Non-proprietary	(7)	Non-proprietary	Resources (format)
License-free	(8)	Public license	License_title

The dimensions of the ordinary citizen test and corresponding Sebastopol principles provided by Wang and Shepherd are displayed in Table 1, as well as the variables from the metadata used to determine if a dataset adheres to these dimensions. The completeness (1), primary (2) and timely (3) principles will not be part of the analysis. Completeness is omitted because, from a citizen's perspective, it is impossible to determine whether the government published all available data. The primary principle will not be part of the analysis because it is not possible to assess the granularity of datasets based on their metadata. The timely principle is left out because it is not possible to determine whether the data were published as soon as possible. Certain literature sees the timely principle as the recency of data (Wang & Shepherd, 2020), the continuous release of data (Lourenço, 2015), and the presence of updates (Vetró et al., 2016). This research will not adopt these definitions of the timely principle because older data are not less open than newer data according to the Sebastopol principles.

Furthermore, the principles of permanence and usage costs were later added to the eight original Sebastopol principles but are not part of the ordinary citizen test. This research will also disregard these principles because it is impossible to determine whether data will remain available in the future, and OGD published by Dutch government organisations are always available free of cost (data.overheid.nl, z.d.-e).

To determine whether data are accessible (4), the variable dataset status will be used, which captures the availability of the dataset on the national data portal. The machine-readable (5) principle is satisfied when datasets are available in formats that can be processed by a computer. The different formats present in the resources variable and whether the format is machine-readable are displayed in Table 2. Datasets are available in multiple formats, data are considered machine-readable when at least one of the available formats is machine-readable. Within an HTML format, data enclosed in a data tag are machine-readable, but the file as a whole is not. Consequently, this research regards the HTML format as not machine-readable. Furthermore, some datasets are available in a ZIP format, but the files contained

within the ZIP are unknown. In total, there are 3.011 ZIP formats present in the data. Therefore, the ZIP format will not be part of the analysis. Non-discriminatory (6) is defined as not requiring access permissions or filling out forms before datasets can be accessed. If datasets adhere to this principle will be determined by reviewing the file's associated access rights.

The non-proprietary (7) principle entails that data are published in an open format which does not require specific software to work with the dataset. The formats that are considered to be non-proprietary are displayed in Table 2. Similarly as with the machine-readable principle, datasets are non-proprietary if one of the available formats is considered non-proprietary and the ZIP format will not be part of the analysis because the files contained within are unknown. License-free (8) means that the license accompanying a dataset should be unrestricted. The licenses present in the metadata are CC-0, CC-BY, no open license, unknown license, and public domain. The CC-0, CC-BY, and public domain licenses are considered open licenses and datasets with these licenses are therefore considered open data in this dimension (data.overheid.nl, z.d.-d). No open license means that the dataset is not public and is consequently considered a closed license and no open data. An unknown license is given to a dataset when the owner does not know what license applies to the dataset. Unknown licenses will therefore be disregarded in the analysis and are present in 772 cases.

Table 2
Data formats

Format	Machine-readable	Non-proprietary
GDB	Yes	No
HTML	No*	Yes
JSON	Yes	Yes
PNG	No	Yes
TIFF	No	Yes
WFS	Yes	No
WMS	Yes	No
XLS	Yes	No
ATOM	Yes	Yes
GML	Yes	Yes
JPEG	No	Yes
ZIP**	N/a	N/a

*Although HTML data enclosed in a data tag is machine-readable, the rest of the file is not
**Files contained in the ZIP file are unknown

The five dimensions that are used to determine whether the datasets qualify as open data are also included in other measurement frameworks. The machine-readable, non-proprietary, and public license dimensions are all part of both the Global Open Data Index and the 5-star rating system. Furthermore, the dimension capturing if a dataset is available on the data portal

is only part of the 5-star rating system. In contrast, the available without restrictions dimension is only included in the Global Open Data Index. The fact that all five research dimensions are also part of other measurement frameworks regarding the openness of data shows that they are adequate indicators to determine the openness of OGD published by Dutch provinces.

Analysis

Distribution of datasets

Table 3 shows how the 7.423 published datasets are distributed over the different provinces along with the number of inhabitants according to the Dutch Central Statistical Office as of January 1, 2023 (CBS Statline, 2023). It stands out that the province of Drenthe is accountable for over 26% of all datasets published by provinces, even though it is one of the smaller provinces. Zuid-Holland also ranks high on the number of datasets published, which is expected because it is the province with the most inhabitants. Noord-Holland and Noord-Brabant, numbers two and three when looking at the number of inhabitants respectively, rank surprisingly low on the number of datasets published.

Table 3

Datasets per province

Province	Number of datasets	Percentage of total	Inhabitants	Files per 100.000 inhabitants
Utrecht	768	10.4%	1.387.657	55.3
Groningen	547	7.4%	596.163	91.8
Zeeland	699	9.4%	391.142	178.7
Drenthe	1.986	26.8%	502.120	395.5
Friesland	226	3.0%	659.612	34.3
Overijssel	389	5.2%	1.184.551	32.8
Z.-Holland	1.345	18.1%	3.804.737	35.4
N.-Holland	332	4.5%	2.956.223	11.2
N.-Brabant	552	7.4%	2.626.368	21.0
Limburg	67	0.9%	1.128.334	5.9
Flevoland	50	0.7%	444.580	11.2
Gelderland	462	6.2%	2.133.751	21.7

When looking at the number of files published per 100.000 inhabitants, the provinces Drenthe, Zeeland, and Groningen are the three provinces that rank the highest. These three provinces are part of the four provinces with the lowest number of inhabitants, which might indicate that provinces with fewer inhabitants relatively publish more datasets. Additionally, the high amount of datasets per 100.000 inhabitants of Zeeland and Groningen might also be caused by not archiving older versions of datasets. Noord-Holland, Limburg, and Flevoland are the provinces that publish the least datasets per 100.000 inhabitants. There is no apparent reason for these relatively low numbers, and all of these provinces advertise publishing Open

Government Data (OGD) the same way other provinces do. These differences between provinces might exist because some provinces delegate responsibilities regarding publishing OGD to the cities, but there is no available information to substantiate this claim.

Findings ordinary citizen test

The results of scoring all datasets on the research dimensions of the ordinary citizen test are displayed in Table 4. The most common barrier to the openness of Dutch provinces' OGD is a proprietary format. Almost half of the datasets in this research are only available in proprietary formats, which undermines the non-proprietary Sebastopol principle. When looking at the research of Wang and Shepherd, the assessed datasets in the United Kingdom (UK) are only available in proprietary formats in almost 20% of cases (2020). This shows that the Dutch provinces perform considerably worse regarding the non-proprietary principle than the UK.

Table 4
Barriers to the openness of Dutch provinces' OGD

Research dimension	Barrier	Number of datasets	Percentage
Availability on data portal	Not available	0	0%
Machine-readable	Format not machine-readable	1.331	17.9%
Available without restrictions	Restricted access	1.470	19.8%
Non-proprietary	Proprietary format	3.614	48.7%
Public license	No public license	698	10.5%

The availability on the data portal research dimension shows that every dataset in this sample is currently available, which means that all Dutch provinces' OGD adheres to the availability Sebastopol principle. Furthermore, almost 18% of datasets are not available on the national data portal in at least one machine-readable format. Although not all datasets are machine-readable, over four out of five datasets are, which means that the Sebastopol principle is mainly satisfied. Most datasets not adhering to the machine-readable research dimension are available in a ZIP format. Since the content of ZIP files is unknown and therefore disregarded in the analysis, the actual number of datasets available in a machine-readable format could be even higher.

When looking at the availability without restrictions dimension, access to almost 20% of datasets is restricted. This means that the dataset is not directly available through the data portal. The data portal does not contain information on why access to these datasets is restricted. In the non-proprietary dimension, almost half of the datasets are only available in proprietary formats. This large amount of datasets available in proprietary formats is problematic because there is no guarantee that these data will still be usable in the future. In the last research dimension, public license, a little over 10% of datasets does not have a

public license which means that the information in these datasets is not freely usable and may not be distributed.

When comparing the results of the ordinary citizen test of the Dutch provinces with those of the UK in the study by Wang and Shepherd, the Dutch provinces perform worse in all dimensions except availability (2020). The number of datasets that adhere to all research dimensions and are therefore considered open data is 2.237, representing 30.1% of all datasets. This means that over two out of three datasets published by Dutch provinces are not open. Although Dutch provinces only perform better regarding the availability principle, the percentage of open datasets is remarkably higher than in the study of Wang and Shepherd, where approximately one out of ten datasets was open (2020). This difference can be explained by Wang and Shepherd including the completeness, primary, and timely principles in their analysis. When disregarding the non-proprietary format dimension, 4.923 datasets are considered open, which accounts for 66.4% of datasets. This suggests that the Dutch provinces should mainly focus on making datasets available in non-proprietary formats to improve the openness of their OGD.

Table 5 shows the percentage of datasets that do not satisfy the research dimensions and the percentage of datasets that adhere to all research dimensions per province. When looking at these numbers, a few things stand out. When looking at the machine-readable dimension, it is noticeable that all provinces score relatively well except for Drenthe and Noord-Brabant. The availability without restrictions dimension shows considerable differences between the provinces. Groningen, Friesland, Limburg, and Flevoland have zero datasets that contain access restrictions, whereas Utrecht and Noord-Brabant have access restrictions in over half of their datasets. There is no known reason for these differences in access rights between the provinces. Regarding the non-proprietary dimension, most provinces perform similarly, where around half of the datasets are available in a non-proprietary format. The provinces standing out are Overijssel and Noord-Brabant. For Overijssel, 25% of its datasets are only available in a proprietary format which is remarkably better than the other provinces. On the other hand, Noord-Brabant has 88% of its datasets only available in a proprietary format.

The final research dimension, public license, shows that only Zeeland, Drenthe, Overijssel, Zuid-Holland, and Noord-Brabant have datasets that do not contain a public license. In the case of Noord-Brabant, over half of its datasets have no open license, which is notably more than the other provinces. There is no information on why Noord-Brabant performs worse in this dimension than every other province does. When looking at the performance of the provinces regarding the openness of their OGD, most provinces perform similarly. The only province standing out is Noord-Brabant, they perform the worst in all dimensions. This entails that the OGD published by Noord-Brabant are considered the least open according to the Sebastopol principles.

The percentage of open datasets per province shows that less than 5% of datasets published by Utrecht and Noord-Brabant adhere to all the research dimensions, which is remarkably lower than for other provinces. Noord-Brabant performed the worst regarding the individual

research dimension, therefore, it is unsurprising that most of their datasets are not open. On the other hand, Utrecht does not score worse than most other provinces in the individual research dimensions. Groningen and Limburg are the two provinces with the most open datasets available. Over half of their published OGD adheres to all the research dimensions.

Table 5
Barriers to the openness of OGD and percentage of open datasets per province

Provinces	Machine-readable	No restrictions	Non-proprietary	Public license	Open datasets
Utrecht	0%	50%	47%	0%	3.9%
Groningen	0%	0%	49%	0%	51.4%
Zeeland	2%	37%	41%	23%	39.3%
Drenthe	39%	1%	47%	1%	14.1%
Friesland	1%	0%	53%	0%	46.0%
Overijssel	9%	38%	25%	27%	49.4%
Z.-Holland	3%	24%	43%	18%	48.8%
N.-Holland	0%	4%	57%	0%	39.2%
N.-Brabant	84%	62%	88%	53%	4.9%
Limburg	0%	0%	49%	0%	50.7%
Flevoland	6%	0%	52%	0%	42.0%
Gelderland	1%	2%	53%	0%	44.4%

Discussion

Assessment frameworks for Open Government Data

This research tries to answer whether the Open Government Data (OGD) published by Dutch provinces indeed qualify as open data and what differences exist between the provinces. This is done using an altered version of the ordinary citizen test based on the Sebastopol principles as introduced by Wang and Shepherd to assess the OGD in the United Kingdom (UK) (2020). Other frameworks to assess open data are available, and the five research dimensions used in this research are also included in the Global Open Data Index and the five-star rating system.

The ordinary citizen test uses the upload date of datasets to determine whether the data adhere to the timely Sebastopol principle, where datasets published less than 30 months ago are considered open (Wang & Shepherd, 2020). This research did not adopt Wang and Shepherd's definition of the timely Sebastopol principle because the assumption that datasets published 31 months ago are less open than those published 29 months ago is not based on any definition of open data. The completeness and primary Sebastopol principles are also not part of this research because there is no information available in the metadata about these principles. Aggregate data were the main barrier to OGD in the UK which begs the question of whether this is also the case for the OGD of Dutch provinces (Wang & Shepherd, 2020).

Both the Global Open Data Index and the Open Data Barometer frameworks are useful for comparing the performance of countries in the openness of their OGD. However, the problem

with these frameworks is that they cannot determine the extent of openness of OGD. Roughly the same goes for the Statistical Performance Indicators, this framework can compare the foundation of countries for publishing OGD but cannot assess the openness of OGD. Furthermore, the five-star rating system is made specifically to assess linked data. Since most governments do not claim to publish linked data and the five-star rating system does not include multiple Sebastopol principles, it does not help assess the extent of openness of OGD.

Although multiple frameworks regarding OGD exist, most frameworks lack certain components for determining the openness of OGD. Additionally, the term open data is often used without being defined or the definitions are vague, which allows governments to frame OGD in a way that fits their agenda (Barry & Bannister, 2014; Cole, 2012). Wang and Shepherd tried to bridge this gap by constructing the ordinary citizen test. However, certain parts of this framework are incompatible with using metadata or are not based on any existing definition of open data. Until a measurement framework is established that covers all aspects of open data, the Sebastopol principles seem to be the best reference point for assessing the openness of OGD.

Interpretation of results

The analysis shows that around 30% of datasets published by Dutch provinces comply with all the research dimensions of the ordinary citizen test. This indicates that the majority of OGD published by provinces do not qualify as open data. In the UK, the dimension that is most often absent in OGD is granularity (Wang & Shepherd, 2020). Therefore, when the granularity dimension would have been part of this research, it is expected that the number of datasets qualifying as open would be even lower. The benefits affiliated with publishing OGD are: increased transparency and accountability, economic growth, strengthened citizen participation, improved provision of public goods and services (Gigler et al., 2011), improved policy-making, more innovation (Janssen et al., 2012), and improved government decisions (Puron-Cid, 2012). Since most OGD published by Dutch provinces do not qualify as open data, it is expected that the provinces will also not be able to gain the benefits associated with publishing OGD.

The research dimension most frequently violated in the OGD of Dutch provinces is that of a non-proprietary format. Almost half of the datasets assessed in this research are only available on the national data portal in proprietary formats. The problem with making OGD available exclusively in proprietary formats is that they can only be opened with specific software. This endangers the usability of these datasets in the years to come, making the data less open. Furthermore, the dimension capturing the availability of datasets on the data portal was satisfied in all cases. It is unknown if all datasets are actually available, but the metadata suggests they are. In the UK, 13.3% of assessed datasets were unavailable through the data portal, which suggests that the Dutch provinces perform noticeably better in this regard (Wang & Shepherd, 2020).

When looking at the individual performance of the provinces, multiple things are noteworthy. First, Drenthe and Zeeland publish far more datasets in relation to their number of inhabitants than any other province. The Green Land, an organisation that tries to humanise data, mentions that Drenthe publishes historical geo-data, whereas most provinces archive a dataset when a newer version is available (z.d.). This explains Drenthe's high number of datasets and might also be the reason that Zeeland published so many datasets in relation to its population. Second, most provinces score average on the non-proprietary format research dimension, except for Overijssel and Noord-Brabant. Only 25% of datasets published by Overijssel are only available in proprietary formats, for Noord-Brabant this is 88%. Additionally, Noord-Brabant and Drenthe are the two provinces that perform considerably worse than other provinces regarding the machine-readable dimension, 84% and 39% of datasets respectively are not available in a machine-readable format. The data show that many datasets are available in a ZIP format. Since the files contained within this ZIP format are unknown, it might be that more datasets are available in machine-readable and non-proprietary formats than the analysis suggests.

The analysis also shows that Utrecht and Noord-Brabant have the lowest percentage of datasets adhering to all research dimensions. Less than 5% of their datasets qualify as open for both provinces. Since most datasets of Dutch provinces are not open and there are substantial differences between provinces in the openness of OGD, the initiative by the national government to help lower levels of government to start opening their data has not been particularly successful. The presence of a national standard concerning the publishing of OGD could be helpful for provinces to reach a higher and more equal level of openness in published OGD.

Conclusion

The literature identifies many benefits related to publishing Open Government Data (OGD). Nevertheless, most governments are not considered to be open. This research has used the metadata of all provincial datasets published on the national data portal to determine the extent of openness of Dutch provinces' OGD. The findings show that approximately only 3 out of 10 datasets qualify as open. This means that most data presented as OGD cannot actually be considered OGD. The Dutch government has stimulated the provinces to start opening their data, but this research shows that they still have a long way to go. The most common barrier to OGD is that data are only available in proprietary formats. The findings also show that the two provinces performing the worst regarding the openness of OGD are Noord-Brabant and Utrecht, whereas Groningen and Limburg perform best. If the government wants to continue its efforts to help provinces open their data, it should prioritise helping provinces publish data in non-proprietary formats. Furthermore, the provinces of Noord-Brabant and Utrecht should compare their OGD strategies with those of Groningen and Limburg to see what can be improved.

Additionally, this research has described the flaws of the ordinary citizen test as well as the flaws of other measurement frameworks regarding open data. Although the ordinary citizen

test comes closer than previous frameworks in assessing the extent of openness of OGD, there is still not a single framework that accommodates all aspects of open data. Since the Sebastopol principles were an adequate basis for assessing OGD and the ordinary citizen test proved more useful than other measurement frameworks, future research on the subject could build on the ordinary citizen test to construct an all-encompassing framework for assessing OGD.

The research methods used are loosely based on the ordinary citizen test, which was constructed to operationalise the eight Sebastopol principles. This research only used five Sebastopol to determine the extent of Dutch provinces' OGD, omitting the completeness, primary, and timely principles. Further research into this subject could focus on the OGD published by Dutch government agencies or municipalities to create a more complete image of the state of OGD in the Netherlands. It could also help to determine whether the lack of openness of Dutch provinces' OGD is a nationwide problem or contained to the provinces.

References

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2022). Criteria for the identification of ineffective open data portals: pretender open data portals. *El Profesional de La Informacion, 31*(1).

Alessandro, M., Cardinale Lagomarsino, B., Scartascini, C., Streb, J., & Torrealday, J. (2021). Transparency and Trust in Government. Evidence from a Survey Experiment. *World Development, 138*, 105223.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), (pp. 399–418.)

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity, 19*(1,2), 129–152.

Berners-Lee, T. (2010). *Linked Data.* https://www.w3.org/DesignIssues/LinkedData.html

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*(3), 264–271.

Cabinet Office. (2013, June 18). *Open Data Charter.* https://www.gov.uk/government/publications/open-data-charter

CBS Statline. (2023, April 20). Regionale kerncijfers Nederland. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072NED/table?fromstatweb

Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly, 34*(2), 231–243.

ckanext-dcatdonl. (z.d.) *CKAN Schema.* https://ckanextdcatdonl.readthedocs.io/en/latest/schema.html

Cole, R. (2012). Some observations on the practice of "open data" as opposed to its promise. *The Journal of Community Informatics, 8*(2).

Data.europa.eu. (2020, January 22). *The benefits and value of open data*. https://data.europa.eu/en/publications/datastories/benefits-and-value-open-data

Data.europa.eu. (n.d.). *Home.* https://data.europa.eu/en

data.overheid.nl & KOOP. (z.d.). *Handreiking data.overheid.nl 2.0.* https://data.overheid.nl/sites/default/files/uploaded_files/Handreiking%20data.overheid.nl%202.0.pdf

data.overheid.nl. (2020, February 12). *Handreiking 2.0.* https://data.overheid.nl/sites/default/files/uploaded_files/Handreiking%20data.overheid.nl%202.0.pdf

data.overheid.nl. (n.d.-a). *About data.overheid.nl.* https://data.overheid.nl/en/ondersteuning/algemeen/over-dataoverheidnl

data.overheid.nl. (n.d.-b). *Provinciale high value datalijst*. https://data.overheid.nl/community/maatschappij/high-value/provincies

data.overheid.nl. (n.d.-c). *Gemeentelijke high value datalijst.* https://data.overheid.nl/community/maatschappij/high-value/gemeenten

data.overheid.nl. (z.d.-d). *Licentie keuze.* https://data.overheid.nl/ondersteuning/data-publiceren/licentie-keuze

data.overheid.nl. (z.d.-e). *Wat is open data.* https://data.overheid.nl/en/ondersteuning/open-data/wat-is-open-data

Digitale Overheid. (2023, January 31). *Wet open overheid Wetgeving - Digitale Overheid*. https://www.digitaleoverheid.nl/wet-open-overheid/

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New Public Management Is Dead—Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory, 16*(3), 467–494.

Farrell, M. (2016). Transparency. *Journal of Library Administration, 56*(4), 444–452.

Gigler, S., Custer, S., & Rahemtulla, H. (2011). *Realizing the vision of open government data*. Paper presented at the Open Development Technology Alliance, Washington, District of Columbia, USA.

Global Open Data Index. (n.d.-a). *Index.* http://index.okfn.org/index.html

Global Open Data Index. (n.d.-b). *Methodology*. http://index.okfn.org/methodology/

Grimmelikhuijsen, S. (2012). A good man but a bad wizard. About the limits and future of transparency of democratic governments. *Information Polity, 17*(3/4), 293–302.

Halachmi, A., & Greiling, D. (2013). Transparency, E-Government, and Accountability: Some Issues and Considerations. *Public Performance & Management Review, 36*(4), 572–584.

Helbig, N., Styrin, E., Canestraro, D., & Pardo, T. (2010). *Information and transparency: Learning from recovery act reporting experiences.* Paper presented at the Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities, Puebla, Mexico.

Hossain, M. A., Rahman, S., Quaddus, M., Hooi, E., & Olanrewaju, A.-S. (2021). Factors Affecting Performance of Open Government Data Initiatives: A Multi-Method Approach Using Sem and FSQCA. *Journal of Organizational Computing and Electronic Commerce, 31*(4), 300–319.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly, 32*, 363–368.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29*(4), 258–268.

Khurshid, M. M., Zakaria, N. H., Rashid, A., Ahmad, M. N., Arfeen, M. I., & Faisal Shehzad, H. M. (2020). Modeling of open government data for public sector organizations using the potential theories and determinants—a systematic review. *Informatics, 7*(24), 1–16.

KOOP. (z.d.-a). *DCAT-AP-DONL 1.1*. https://dcat-ap-donl.readthedocs.io/en/latest/

KOOP. (z.d.-b). CKANEXT-DCATDONL - CKAN Schema - Dataset. https://ckanext-dcatdonl.readthedocs.io/en/latest/schema-dataset.html

Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly, 35*(1), 13–29.

Kučera, J., Chlapek, D., & Nečaský, M. (2013). Open Government Data Catalogs: Current Approaches and Quality Perspective. In Kő, A., Leitner, C., Leitold, H. & Prosser, A. (Eds.) *Technology-Enabled Innovation for Democracy, Government and Governance* (pp. 152–166). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. *Government Information Quarterly, 29*(4), 492–503.

Lewis, J. M., O'Flynn, J., & Sullivan, H. (2014). Accountability: To Whom, in Relation to What, and Why? *Australian Journal of Public Administration, 73*(4), 401–407.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly, 32*(3), pp. 323–332.

Margetts, H., & Dunleavy, P. (2013). The Second Wave of Digital-Era Governance: A Quasi-Paradigm for Government on the Web. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*(1987): 20120382.

Matheus, R., & Janssen, M. (2020). A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory. *Public Performance & Management Review*, *43*(3), 503–534.

McDermott, P. (2010). Building open government. *Government Information Quarterly, 27*(4), 401–413.

Nikiforova, A., & McBride, K. (2021). Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics, 58*, 101539.

Open Data Barometer. (n.d.). *Methodology.* https://opendatabarometer.org/leadersedition/methodology/

Open Data Watch. (2018, October 21). *Indexes of Data Quality and Openness*. https://opendatawatch.com/blog/indexes-of-data-quality-and-openness/

Open Knowledge Foundation. (n.d.-a). *Open Definition - Defining Open in Open Data, Open Content and Open Knowledge*. http://opendefinition.org/

Open Knowledge Foundation. (n.d.-b). *Open Definition 2.1 - Defining Open in Open Data, Open Content and Open Knowledge*. https://opendefinition.org/od/2.1/en/

Opengovdata. (n.d.). *The 8 Principles of Open Government Data*. https://opengovdata.org/

Organisation for Economic Co-operation and Development. (z.d.). *Open Government Data.* https://www.oecd.org/gov/digital-government/open-government-data.htm

Park, S., & Gil-Garcia, J. R. (2017). *Understanding Transparency and Accountability in Open Government Ecosystems: The Case of Health Data Visualizations in a State Government.* Paper presented at the Proceedings of the 18th Annual International Conference on Digital Government Research, Staten Island, New York, USA.

Puron-Cid, G., Gil-Garcia, J. R., & Luna-Reyes, L. F. (2012). IT-enabled policy analysis: New technologies, sophisticated analysis and open data for better government decisions. Paper presented at the Proceedings of the 13th Annual International Conference on Digital Government Research, College Park, Maryland, USA.

Rijksprogramma Voor Duurzaam Digitale Informatiehuishouding. (2020, April 6). *Wet hergebruik van overheidsinformatie (Who)*. https://www.informatiehuishouding.nl/onderwerpen/wet-hergebruik-overheidsinformatie

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management, 37*(3), 150–154.

Schillemans, T. (2016). Calibrating Public Sector Accountability: Translating experimental findings to public sector accountability. *Public Management Review, 18*(9), 1400–1420.

Stamati, T., Papadopoulos, T., & Anagnostopoulos, D. (2015). Social media for openness and accountability in the public sector: Cases in the Greek context. *Government Information Quarterly, 32*(1), 12–29.

Sunlight Foundation. (n.d.). *Ten Principles For Opening Up Government Information.* https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

The Green Land. (z.d.). *Provinciale open data: alleen maar geo!?.* https://thegreenland.eu/2018/08/provinciale-open-data-alleen-maar-geo/

The Open Data Charter. (2020, November 5). *Principles.* https://opendatacharter.net/principles/

The Open Data Charter. (2022-a, July 21). *Our history*. https://opendatacharter.net/our-history/

The Open Data Charter. (2022-b, December 14). *Government adopters.* https://opendatacharter.net/government-adopters

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly, 33*(2), 325–337.

Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data – A critique of open government datasets in the UK. *Government Information Quarterly, 37*(1), 101405.

Wanna, J. (2018). Opening government: Transparency and engagement in the information age. In Wanna, J. & Vincent, S. (Eds.), *Opening government: Transparency and engagement in the information age* (pp. 3-24). ANU Press.

White House (2009, December 8). *Open Government Directive*. https://obamawhitehouse.archives.gov/open/documents/opengovernment-directive

World Bank. (n.d.). *Statistical Performance Indicators - Framework*. https://www.worldbank.org/en/programs/statistical-performance-indicators/Framework

Yi, M. (2019). Exploring the quality of government open data : comparison study of the UK, the USA and Korea. *Electronic Library, 37*(1), 35–48)

Zhao, Y., & Fan, B. (2018). Exploring open government data capacity of government agency: Based on the resource-based theory. *Government Information Quarterly, 35*(1), 1–12.

Appendix

Python code used to access the Application Programming Interface

```
import urllib
import json
import pprint
import os
import pandas as pd
from functools import reduce
from tqdm.notebook import tqdm
import sys

response = urllib.request.urlopen('https://data.overheid.nl/data/api/3/action/package_list')
response.code

def get_response(url):
  response = urllib.request.urlopen(url)
  response_dict = json.loads(response.read())
  if response.code != 200 or not response_dict["success"]:
```

```
    return None
  return response_dict["result"]

if not os.path.isfile("data.csv"):
  dataset_dicts = []
  for dataset in tqdm(all_datasets):
    qurl = query_url.format(dataset)
    dataset_dict = get_response(qurl)
    dataset_dicts.append(dataset_dict)
  df = pd.json_normalize(dataset_dicts, sep='_')
  df.to_csv('data.csv', encoding='utf-8', index=False)
dataset_dicts = pd.read_csv('data.csv')

for d in dataset_dicts:
  print(d)
```

SPSS syntax

```
RECODE publisher ('http://standaarden.overheid.nl/owms/terms/Groningen_(provincie)'=2)
    ('http://standaarden.overheid.nl/owms/terms/Utrecht_(provincie)'=1)
    ('http://standaarden.overheid.nl/owms/terms/Zeeland_(provincie)'=3)
    ('http://standaarden.overheid.nl/owms/terms/Drenthe'=4)
    ('http://standaarden.overheid.nl/owms/terms/Fryslan'=5)
    ('http://standaarden.overheid.nl/owms/terms/Overijssel'=6)
    ('http://standaarden.overheid.nl/owms/terms/Zuid-Holland'=7)
    ('http://standaarden.overheid.nl/owms/terms/Noord-Holland'=8)
    ('http://standaarden.overheid.nl/owms/terms/Noord-Brabant'=9)
    ('http://standaarden.overheid.nl/owms/terms/Limburg'=10)
    ('http://standaarden.overheid.nl/owms/terms/Flevoland'=11)
    ('http://standaarden.overheid.nl/owms/terms/Gelderland'=12) (ELSE=99) INTO
Publisher_numeric.
EXECUTE.

SELECT IF (ANY(Publisher_numeric,1,2,3,4,5,6,7,8,9,10,11,12)).
EXECUTE.

RECODE access_rights ('http://publications.europa.eu/resource/authority/access-
right/PUBLIC'=1)
    ('http://publications.europa.eu/resource/authority/access-right/RESTRICTED'=0) INTO
    Non_discriminatory.
EXECUTE.

RECODE license_title ('CC-0 (1.0)'=1) ('CC-BY (4.0)'=1) ('Geen open licentie'=0)
('Onbekende '+  'licentie'=SYSMIS) ('Publiek domein'=1) INTO License_free.
```

```
EXECUTE.

RECODE format1 format2 format3 format4 format5 format6 format7 format8 format9
format10 format11 format12 format13 format14 format15 format16 format17 format18
format19 format20 format21 format22 format23 format24 format25 format26 format27
format28 format29 format30 format31 format32 format33 format34 format35 format36
('http://publications.europa.eu/resource/authority/file-type/GDB'=1)
('http://publications.europa.eu/resource/authority/file-type/HTML'=2)
('http://publications.europa.eu/resource/authority/file-type/JSON'=3)
('http://publications.europa.eu/resource/authority/file-type/PNG'=4)
('http://publications.europa.eu/resource/authority/file-type/TIFF'=5)
('http://publications.europa.eu/resource/authority/file-type/WFS_SRVC'=6)
('http://publications.europa.eu/resource/authority/file-type/WMS_SRVC'=7)
('http://publications.europa.eu/resource/authority/file-type/XLS'=8)
('http://publications.europa.eu/resource/authority/file-type/ZIP'=9)
('https://data.overheid.nl/format/unknown'=SYSMIS)
('http://publications.europa.eu/resource/authority/file-type/ATOM'=10)
('http://publications.europa.eu/resource/authority/file-type/GML'=11)
('http://publications.europa.eu/resource/authority/file-type/JPEG'=12) INTO
Format1_numeric Format2_numeric Format3_numeric Format4_numeric Format5_numeric
Format6_numeric Format7_numeric Format8_numeric Format9_numeric Format10_numeric
Format11_numeric Format12_numeric Format13_numeric Format14_numeric
Format15_numeric Format16_numeric Format17_numeric Format18_numeric
Format19_numeric Format20_numeric Format21_numeric Format22_numeric
Format23_numeric Format24_numeric Format25_numeric Format26_numeric
Format27_numeric Format28_numeric Format29_numeric Format30_numeric
Format31_numeric Format32_numeric Format33_numeric Format34_numeric
Format35_numeric Format36_numeric.
EXECUTE.

compute M = 0.
IF (Format1_numeric = 1 or Format1_numeric = 3 or Format1_numeric = 6 or
Format1_numeric = 7 or Format1_numeric = 8 or Format1_numeric = 10 or
Format1_numeric = 11) M = 1.
IF (Format2_numeric = 1 or Format2_numeric = 3 or Format2_numeric = 6 or
Format2_numeric = 7 or Format2_numeric = 8 or Format2_numeric = 10 or
Format2_numeric = 11) M = 1.
IF (Format3_numeric = 1 or Format3_numeric = 3 or Format3_numeric = 6 or
Format3_numeric = 7 or Format3_numeric = 8 or Format3_numeric = 10 or
Format3_numeric = 11) M = 1.
IF (Format4_numeric = 1 or Format4_numeric = 3 or Format4_numeric = 6 or
Format4_numeric = 7 or Format4_numeric = 8 or Format4_numeric = 10 or
Format4_numeric = 11) M = 1.
```

IF (Format5_numeric = 1 or Format5_numeric = 3 or Format5_numeric = 6 or Format5_numeric = 7 or Format5_numeric = 8 or Format5_numeric = 10 or Format5_numeric = 11) M = 1.
IF (Format6_numeric = 1 or Format6_numeric = 3 or Format6_numeric = 6 or Format6_numeric = 7 or Format6_numeric = 8 or Format6_numeric = 10 or Format6_numeric = 11) M = 1.
IF (Format7_numeric = 1 or Format7_numeric = 3 or Format7_numeric = 6 or Format7_numeric = 7 or Format7_numeric = 8 or Format7_numeric = 10 or Format7_numeric = 11) M = 1.
IF (Format8_numeric = 1 or Format8_numeric = 3 or Format8_numeric = 6 or Format8_numeric = 7 or Format8_numeric = 8 or Format8_numeric = 10 or Format8_numeric = 11) M = 1.
IF (Format9_numeric = 1 or Format9_numeric = 3 or Format9_numeric = 6 or Format9_numeric = 7 or Format9_numeric = 8 or Format9_numeric = 10 or Format9_numeric = 11) M = 1.
IF (Format10_numeric = 1 or Format10_numeric = 3 or Format10_numeric = 6 or Format10_numeric = 7 or Format10_numeric = 8 or Format10_numeric = 10 or Format10_numeric = 11) M = 1.
IF (Format11_numeric = 1 or Format11_numeric = 3 or Format11_numeric = 6 or Format11_numeric = 7 or Format11_numeric = 8 or Format11_numeric = 10 or Format11_numeric = 11) M = 1.
IF (Format12_numeric = 1 or Format12_numeric = 3 or Format12_numeric = 6 or Format12_numeric = 7 or Format12_numeric = 8 or Format12_numeric = 10 or Format12_numeric = 11) M = 1.
IF (Format13_numeric = 1 or Format13_numeric = 3 or Format13_numeric = 6 or Format13_numeric = 7 or Format13_numeric = 8 or Format13_numeric = 10 or Format13_numeric = 11) M = 1.
IF (Format14_numeric = 1 or Format14_numeric = 3 or Format14_numeric = 6 or Format14_numeric = 7 or Format14_numeric = 8 or Format14_numeric = 10 or Format14_numeric = 11) M = 1.
IF (Format15_numeric = 1 or Format15_numeric = 3 or Format15_numeric = 6 or Format15_numeric = 7 or Format15_numeric = 8 or Format15_numeric = 10 or Format15_numeric = 11) M = 1.
IF (Format16_numeric = 1 or Format16_numeric = 3 or Format16_numeric = 6 or Format16_numeric = 7 or Format16_numeric = 8 or Format16_numeric = 10 or Format16_numeric = 11) M = 1.
IF (Format17_numeric = 1 or Format17_numeric = 3 or Format17_numeric = 6 or Format17_numeric = 7 or Format17_numeric = 8 or Format17_numeric = 10 or Format17_numeric = 11) M = 1.
IF (Format18_numeric = 1 or Format18_numeric = 3 or Format18_numeric = 6 or Format18_numeric = 7 or Format18_numeric = 8 or Format18_numeric = 10 or Format18_numeric = 11) M = 1.

IF (Format19_numeric = 1 or Format19_numeric = 3 or Format19_numeric = 6 or Format19_numeric = 7 or Format19_numeric = 8 or Format19_numeric = 10 or Format19_numeric = 11) M = 1.
IF (Format20_numeric = 1 or Format20_numeric = 3 or Format20_numeric = 6 or Format20_numeric = 7 or Format20_numeric = 8 or Format20_numeric = 10 or Format20_numeric = 11) M = 1.
IF (Format21_numeric = 1 or Format21_numeric = 3 or Format21_numeric = 6 or Format21_numeric = 7 or Format21_numeric = 8 or Format21_numeric = 10 or Format21_numeric = 11) M = 1.
IF (Format22_numeric = 1 or Format22_numeric = 3 or Format22_numeric = 6 or Format22_numeric = 7 or Format22_numeric = 8 or Format22_numeric = 10 or Format22_numeric = 11) M = 1.
IF (Format23_numeric = 1 or Format23_numeric = 3 or Format23_numeric = 6 or Format23_numeric = 7 or Format23_numeric = 8 or Format23_numeric = 10 or Format23_numeric = 11) M = 1.
IF (Format24_numeric = 1 or Format24_numeric = 3 or Format24_numeric = 6 or Format24_numeric = 7 or Format24_numeric = 8 or Format24_numeric = 10 or Format24_numeric = 11) M = 1.
IF (Format25_numeric = 1 or Format25_numeric = 3 or Format25_numeric = 6 or Format25_numeric = 7 or Format25_numeric = 8 or Format25_numeric = 10 or Format25_numeric = 11) M = 1.
IF (Format26_numeric = 1 or Format26_numeric = 3 or Format26_numeric = 6 or Format26_numeric = 7 or Format26_numeric = 8 or Format26_numeric = 10 or Format26_numeric = 11) M = 1.
IF (Format27_numeric = 1 or Format27_numeric = 3 or Format27_numeric = 6 or Format27_numeric = 7 or Format27_numeric = 8 or Format27_numeric = 10 or Format27_numeric = 11) M = 1.
IF (Format28_numeric = 1 or Format28_numeric = 3 or Format28_numeric = 6 or Format28_numeric = 7 or Format28_numeric = 8 or Format28_numeric = 10 or Format28_numeric = 11) M = 1.
IF (Format29_numeric = 1 or Format29_numeric = 3 or Format29_numeric = 6 or Format29_numeric = 7 or Format29_numeric = 8 or Format29_numeric = 10 or Format29_numeric = 11) M = 1.
IF (Format30_numeric = 1 or Format30_numeric = 3 or Format30_numeric = 6 or Format30_numeric = 7 or Format30_numeric = 8 or Format30_numeric = 10 or Format30_numeric = 11) M = 1.
IF (Format31_numeric = 1 or Format31_numeric = 3 or Format31_numeric = 6 or Format31_numeric = 7 or Format31_numeric = 8 or Format31_numeric = 10 or Format31_numeric = 11) M = 1.
IF (Format32_numeric = 1 or Format32_numeric = 3 or Format32_numeric = 6 or Format32_numeric = 7 or Format32_numeric = 8 or Format32_numeric = 10 or Format32_numeric = 11) M = 1.

```
IF (Format33_numeric = 1 or Format33_numeric = 3 or Format33_numeric = 6 or
Format33_numeric = 7 or Format33_numeric = 8 or Format33_numeric = 10 or
Format33_numeric = 11) M = 1.
IF (Format34_numeric = 1 or Format34_numeric = 3 or Format34_numeric = 6 or
Format34_numeric = 7 or Format34_numeric = 8 or Format34_numeric = 10 or
Format34_numeric = 11) M = 1.
IF (Format35_numeric = 1 or Format35_numeric = 3 or Format35_numeric = 6 or
Format35_numeric = 7 or Format35_numeric = 8 or Format35_numeric = 10 or
Format35_numeric = 11) M = 1.
IF (Format36_numeric = 1 or Format36_numeric = 3 or Format36_numeric = 6 or
Format36_numeric = 7 or Format36_numeric = 8 or Format36_numeric = 10 or
Format36_numeric = 11) M = 1.
execute.

compute Nonproprietary_dummy = 0.
IF (Format1_numeric = 2 or Format1_numeric = 3 or Format1_numeric = 4 or
Format1_numeric = 5 or Format1_numeric = 10 or Format1_numeric = 11 or
Format1_numeric = 12) Nonproprietary_dummy = 1.
IF (Format2_numeric = 2 or Format2_numeric = 3 or Format2_numeric = 4 or
Format2_numeric = 5 or Format2_numeric = 10 or Format2_numeric = 11 or
Format2_numeric = 12) Nonproprietary_dummy = 1.
IF (Format3_numeric = 2 or Format3_numeric = 3 or Format3_numeric = 4 or
Format3_numeric = 5 or Format3_numeric = 10 or Format3_numeric = 11 or
Format3_numeric = 12) Nonproprietary_dummy = 1.
IF (Format4_numeric = 2 or Format4_numeric = 3 or Format4_numeric = 4 or
Format4_numeric = 5 or Format4_numeric = 10 or Format4_numeric = 11 or
Format4_numeric = 12) Nonproprietary_dummy = 1.
IF (Format5_numeric = 2 or Format5_numeric = 3 or Format5_numeric = 4 or
Format5_numeric = 5 or Format5_numeric = 10 or Format5_numeric = 11 or
Format5_numeric = 12) Nonproprietary_dummy = 1.
IF (Format6_numeric = 2 or Format6_numeric = 3 or Format6_numeric = 4 or
Format6_numeric = 5 or Format6_numeric = 10 or Format6_numeric = 11 or
Format6_numeric = 12) Nonproprietary_dummy = 1.
IF (Format7_numeric = 2 or Format7_numeric = 3 or Format7_numeric = 4 or
Format7_numeric = 5 or Format7_numeric = 10 or Format7_numeric = 11 or
Format7_numeric = 12) Nonproprietary_dummy = 1.
IF (Format8_numeric = 2 or Format8_numeric = 3 or Format8_numeric = 4 or
Format8_numeric = 5 or Format8_numeric = 10 or Format8_numeric = 11 or
Format8_numeric = 12) Nonproprietary_dummy = 1.
IF (Format9_numeric = 2 or Format9_numeric = 3 or Format9_numeric = 4 or
Format9_numeric = 5 or Format9_numeric = 10 or Format9_numeric = 11 or
Format9_numeric = 12) Nonproprietary_dummy = 1.
```

IF (Format10_numeric = 2 or Format10_numeric = 3 or Format10_numeric = 4 or Format10_numeric = 5 or Format10_numeric = 10 or Format10_numeric = 11 or Format10_numeric = 12) Nonproprietary_dummy = 1.
IF (Format11_numeric = 2 or Format11_numeric = 3 or Format11_numeric = 4 or Format11_numeric = 5 or Format11_numeric = 10 or Format11_numeric = 11 or Format11_numeric = 12) Nonproprietary_dummy = 1.
IF (Format12_numeric = 2 or Format12_numeric = 3 or Format12_numeric = 4 or Format12_numeric = 5 or Format12_numeric = 10 or Format12_numeric = 11 or Format12_numeric = 12) Nonproprietary_dummy = 1.
IF (Format13_numeric = 2 or Format13_numeric = 3 or Format13_numeric = 4 or Format13_numeric = 5 or Format13_numeric = 10 or Format13_numeric = 11 or Format13_numeric = 12) Nonproprietary_dummy = 1.
IF (Format14_numeric = 2 or Format14_numeric = 3 or Format14_numeric = 4 or Format14_numeric = 5 or Format14_numeric = 10 or Format14_numeric = 11 or Format14_numeric = 12) Nonproprietary_dummy = 1.
IF (Format15_numeric = 2 or Format15_numeric = 3 or Format15_numeric = 4 or Format15_numeric = 5 or Format15_numeric = 10 or Format15_numeric = 11 or Format15_numeric = 12) Nonproprietary_dummy = 1.
IF (Format16_numeric = 2 or Format16_numeric = 3 or Format16_numeric = 4 or Format16_numeric = 5 or Format16_numeric = 10 or Format16_numeric = 11 or Format16_numeric = 12) Nonproprietary_dummy = 1.
IF (Format17_numeric = 2 or Format17_numeric = 3 or Format17_numeric = 4 or Format17_numeric = 5 or Format17_numeric = 10 or Format17_numeric = 11 or Format17_numeric = 12) Nonproprietary_dummy = 1.
IF (Format18_numeric = 2 or Format18_numeric = 3 or Format18_numeric = 4 or Format18_numeric = 5 or Format18_numeric = 10 or Format18_numeric = 11 or Format18_numeric = 12) Nonproprietary_dummy = 1.
IF (Format19_numeric = 2 or Format19_numeric = 3 or Format19_numeric = 4 or Format19_numeric = 5 or Format19_numeric = 10 or Format19_numeric = 11 or Format19_numeric = 12) Nonproprietary_dummy = 1.
IF (Format20_numeric = 2 or Format20_numeric = 3 or Format20_numeric = 4 or Format20_numeric = 5 or Format20_numeric = 10 or Format20_numeric = 11 or Format20_numeric = 12) Nonproprietary_dummy = 1.
IF (Format21_numeric = 2 or Format21_numeric = 3 or Format21_numeric = 4 or Format21_numeric = 5 or Format21_numeric = 10 or Format21_numeric = 11 or Format21_numeric = 12) Nonproprietary_dummy = 1.
IF (Format22_numeric = 2 or Format22_numeric = 3 or Format22_numeric = 4 or Format22_numeric = 5 or Format22_numeric = 10 or Format22_numeric = 11 or Format22_numeric = 12) Nonproprietary_dummy = 1.
IF (Format23_numeric = 2 or Format23_numeric = 3 or Format23_numeric = 4 or Format23_numeric = 5 or Format23_numeric = 10 or Format23_numeric = 11 or Format23_numeric = 12) Nonproprietary_dummy = 1.

```
IF (Format24_numeric = 2 or Format24_numeric = 3 or Format24_numeric = 4 or
Format24_numeric = 5 or Format24_numeric = 10 or Format24_numeric = 11 or
Format24_numeric = 12) Nonproprietary_dummy = 1.
IF (Format25_numeric = 2 or Format25_numeric = 3 or Format25_numeric = 4 or
Format25_numeric = 5 or Format25_numeric = 10 or Format25_numeric = 11 or
Format25_numeric = 12) Nonproprietary_dummy = 1.
IF (Format26_numeric = 2 or Format26_numeric = 3 or Format26_numeric = 4 or
Format26_numeric = 5 or Format26_numeric = 10 or Format26_numeric = 11 or
Format26_numeric = 12) Nonproprietary_dummy = 1.
IF (Format27_numeric = 2 or Format27_numeric = 3 or Format27_numeric = 4 or
Format27_numeric = 5 or Format27_numeric = 10 or Format27_numeric = 11 or
Format27_numeric = 12) Nonproprietary_dummy = 1.
IF (Format28_numeric = 2 or Format28_numeric = 3 or Format28_numeric = 4 or
Format28_numeric = 5 or Format28_numeric = 10 or Format28_numeric = 11 or
Format28_numeric = 12) Nonproprietary_dummy = 1.
IF (Format29_numeric = 2 or Format29_numeric = 3 or Format29_numeric = 4 or
Format29_numeric = 5 or Format29_numeric = 10 or Format29_numeric = 11 or
Format29_numeric = 12) Nonproprietary_dummy = 1.
IF (Format30_numeric = 2 or Format30_numeric = 3 or Format30_numeric = 4 or
Format30_numeric = 5 or Format30_numeric = 10 or Format30_numeric = 11 or
Format30_numeric = 12) Nonproprietary_dummy = 1.
IF (Format31_numeric = 2 or Format31_numeric = 3 or Format31_numeric = 4 or
Format31_numeric = 5 or Format31_numeric = 10 or Format31_numeric = 11 or
Format31_numeric = 12) Nonproprietary_dummy = 1.
IF (Format32_numeric = 2 or Format32_numeric = 3 or Format32_numeric = 4 or
Format32_numeric = 5 or Format32_numeric = 10 or Format32_numeric = 11 or
Format32_numeric = 12) Nonproprietary_dummy = 1.
IF (Format33_numeric = 2 or Format33_numeric = 3 or Format33_numeric = 4 or
Format33_numeric = 5 or Format33_numeric = 10 or Format33_numeric = 11 or
Format33_numeric = 12) Nonproprietary_dummy = 1.
IF (Format34_numeric = 2 or Format34_numeric = 3 or Format34_numeric = 4 or
Format34_numeric = 5 or Format34_numeric = 10 or Format34_numeric = 11 or
Format34_numeric = 12) Nonproprietary_dummy = 1.
IF (Format35_numeric = 2 or Format35_numeric = 3 or Format35_numeric = 4 or
Format35_numeric = 5 or Format35_numeric = 10 or Format35_numeric = 11 or
Format35_numeric = 12) Nonproprietary_dummy = 1.
IF (Format36_numeric = 2 or Format36_numeric = 3 or Format36_numeric = 4 or
Format36_numeric = 5 or Format36_numeric = 10 or Format36_numeric = 11 or
Format36_numeric = 12) Nonproprietary_dummy = 1.
execute.

compute open = 0.
IF (Non_discriminatory = 1 AND License_free = 1 AND Machineprocessable_dummy = 1
AND Nonproprietary_dummy = 1) open = 1.
```

```
execute.

compute open_without_proprietary = 0.
IF (Non_discriminatory = 1 AND License_free = 1 AND Machineprocessable_dummy = 1)
open_without_proprietary = 1.
execute.
```