Open health data: Mapping the ecosystem

Roel Heijlen[1,2] (ID) and Joep Crompvoets[1]

Abstract

Background: Governments around the world own multiple datasets related to the policy domain of health. Datasets range from vaccination rates to the availability of health care practitioners in a region to the outcomes of certain surgeries. Health is believed to be a promising subject in the case of open government data policies. However, the specific properties of health data such as its sensibilities regarding privacy, ethics, and ownership encompass particular conditions either enabling or preventing datasets to become freely and easily accessible for everyone.

Objective and methods: This paper aims to map the ecosystem of open health data. By analyzing the foundations of health data and the commonalities of open data ecosystems via literature analysis, the socio-technical environment in which health data managed by governments are opened up or potentially stay closed is created. After its theoretical development, the open health data ecosystem is tested via a case study concerning the Data for Better Health initiative from the government of Belgium.

Results: Creation and assessment of an open health data ecosystem consisting of stakeholders, interests, information policies, and data preparation activities.

Conclusions: The policy domain of health includes de-identification activities, bioethical assessments, and the specific role of data providers within its open data ecosystem. However, the concept of open data does not always fully apply to the topic of health. Such several health datasets may be findable via government portals but not directly accessible. Differentiation within types of health data and data user capacities are recommendable for future research.

Keywords

Health data, open data, government portals, ecosystem, case study

Submission date: 14 October 2019; Acceptance date: 10 September 2021

Introduction

The concept of open government data (OGD) has drawn much attention. It refers to government data that can be (re)used by anyone without barriers such as fees, legal objections, or technical difficulties.[1–4] Many governments around the globe have taken action to open up government data.[5,6] Actions comprise for instance programs to stimulate the opening of datasets by government agencies, the development of web portals making government datasets findable and accessible, and the organization of competitive events to encourage creative use of the data by citizens, entrepreneurs, and academics. These initiatives to make government data open are on the one hand motivated by aspirations to provide more transparency on government functioning and to enable citizen participation in policy making. On the other hand, governments are opening datasets from an economic point of view as the availability of open data fuels innovations. It permits companies and citizens to add value to these data which in turn can lead to the

[1]KU Leuven Public Governance Institute, Leuven, Belgium
[2]Sciensano, Brussels, Belgium

Corresponding author:
Roel Heijlen, KU Leuven Public Goverance Institute, Parkstraat 45 3000 Leuven, Belgium.
Email: roel.heijlen@kuleuven.be

development of (information-based) products and services.[3,5–7] Within academic literature, the concept of open (government) data has been widely described ranging from benefits and risks concerning the opening up of data to the mapping of stakeholders and interests.[8–10] At the same time, frameworks regarding governance or ecosystems have been elaborated to guideline open data policies.[11–13] In most research, the topic of OGD is broadly approached not differentiating between types of government data. Some types of data (e.g. geographical data) are already widely available on OGD portals, while other types are still difficult to find, to access and to (re)use. Not every type of data can be "opened up" easily. As reflected in the citation of McGrail et al.[14]: "There is a big difference between making available a century of weather pattern data and last year's detailed records of acute inpatient hospital use." This article focuses on health data which are managed by governmental actors. Several open data manifestos mention the topic health as a promising subject for open data policies. Although proponents belief that the (re)use of health data leads to societal and economic advantages, sensitivities related to health may hinder free and public access. Interested in knowing the particularities of these types of data, the article explores how existing open data ecosystems can be applied to health data. The metaphor of ecosystems will be used to analyze the socio-technical environment in which open health data are created and in which values from these data are derived. An ecosystem contains the interactions between people such as data producers and data users, infrastructure or processes related to data management, and formal or informal institutions (e.g. motivation, policy, legislation).[4,15] Such interactions influence the publication or not of datasets and the reaping of expected benefits.

Research aim & design

Several generic OGD ecosystems have already been developed. However, it is recognized that according to the policy domain *(e.g. tourism, environment, energy, transport, health, etc.)* government actors participate in substantially different ecosystems each with their own particular stakeholders and dynamics.[15] Our research aims to map and test an ecosystem regarding open health data.

In order to fulfill the mapping and the testing of an open health data ecosystem a systemic plan is applied. As mentioned in the introduction, a wide range of general OGD theoretical frameworks have already been elaborated. Keeping these general frameworks in mind, this paper seeks to distinguish itself from the generic approach to government data by focusing on one particular type of data, moreover health data. Before the actual mapping of an open government health data ecosystem can start, the meaning of health data requires clarification. Therefore, this research plan begins with a definition of health data to delineate the type of data involved and to situate its meaning within the concept of OGD. Besides a theoretical

exercise of defining, the second step of the plan consists of an observation of some existing published open government health data sets and publication tools. This observation contributes to awareness on current practices that can be integrated in the mapping of the ecosystem. As OGD is often considered as a positive trend towards transparency and value creation, the third step explores the expected benefits related to opening up health data. The authors assume that comprehension regarding the expected benefits will help to understand some interests at stake related to health data. Although, there may be multiple benefits the particularities of health data may have an impact on whether or not certain data sets are published. Therefore, a fourth step consists of the decomposition of the characteristics of health data. By doing so, it examines what might distinguish this type of data from other government data. While the first four steps focus on health data knowledge gathering, the fifth step provides an overview of the elements of an open government ecosystem. This fifth step will enable the mapping of the accumulated knowledge about health data to its own customized ecosystem.

In sum, some preliminary questions are addressed to facilitate the creation of the open health data ecosystem:

- RQ1: How to define health data in the context of OGD?
- RQ2a: Which kind of government datasets are involved?
- RQ2b: Which kind of publication tools are used?
- RQ3: What are the expected benefits of opening up health data?
- RQ4: How to characterize health data?
- RQ5: Which are the key aspects of an open data ecosystem?

These questions will be answered via

1. analysis of academic and grey literature in the field of e-government, data governance, health law & health ethics, medical informatics, etc. *(=RQ1, RQ3, RQ4 & RQ5)*
2. observations of government practices by means of document analysis consisting of government information material *(=RQ2)*

Regarding the literature analysis, the authors started with some key literature on the general topic of OGD that is frequently cited in academic journals or books linked to e-government, information management, and/or public policy. Several of these articles have bundled insights on the concept, expectations, constraints, practices, and participants of OGD. Next, a search of the term "open health data" in the LIMO database of the authors' research institutions was executed. Considering the limited search results, the selection was added with:

(a) Recent reports from supranational organizations (e.g. Organisation for Economic Co-operation and Development, World Health Organisation, G8, European Union) or private think thanks that describe evolutions towards more innovative data use. Such evolutions include inter alia the presence and exploitation of (more) open data and big data. References in these reports also led to the inclusion of some relevant papers on current promises and challenges concerning health data management as the health(care) sector is considered to profit from data driven innovation.

(b) Literature related to societal debate on the potential impact of privacy or other (restrictive) regulations on the possibility whether or not to perform health(care) research. Such publications were chosen as they included arguments for data protection related to the personal sphere as well as arguments for data reuse to gain relevant knowledge for health(care) improvements. Both types of arguments were valuable to better understand the peculiarities of health data and aspects with regard to making data open or keeping it closed.

Supported by the answers to the preliminary questions, the open health ecosystem will be mapped and afterward tested via a case study concerning the introduction of an open health data policy in Belgium. In other words, the preliminary questions help to answer the overarching research question: "How to map an open health data ecosystem." A summary of the different research steps and methods included are presented in Figure 1.

Outline of the article

Following the introduction, the second part of the article starts with defining health data. After analyzing existing definitions from academic and gray literature a pragmatic working definition will be proposed. The third part provides a view on the state of the art regarding open health data. It explores which kind of health related datasets are already opened up within existing government practices and via which tools. Fourthly, the article offers an outline of the expected benefits. How can actors involved profit from access to these data? In the fifth part, the special nature of health data is investigated in order to express its specific characteristics and understand potential differences with other government data. Sixthly, the concept of open data ecosystems is introduced by providing information on the use of the ecosystem metaphor concerning open data policies and its key components. After the theoretical analyses aimed at grasping the foundations of health data & open data ecosystems in the former parts, the seventh part of the article is dedicated to the design of an open health data ecosystem. Components of a general OGD ecosystem are applied to the specific context of health data. Next, the designed open heath data ecosystem will be tried out via a use case concerning the introduction of a policy intended to open up health datasets. Finally, the article ends with conclusions and ideas for future research.

Health data: what?

Since this articles focuses on data related to health the first step of our research consists of giving meaning to the term health data. While scrutinizing existing definitions of health data, one can notice a term consisting of several components (see Figure 2). A first component regards the level as health data can either refer to the health of an individual person or to the health of a population.[16] On an individual level health data are extremely personal as they are related to one single person.[17] By contrast, data on population level are aggregated data providing bundled information on a (inter)national, regional or local group of people. A second definition component, includes the type of health normally not differentiating between physical and mental
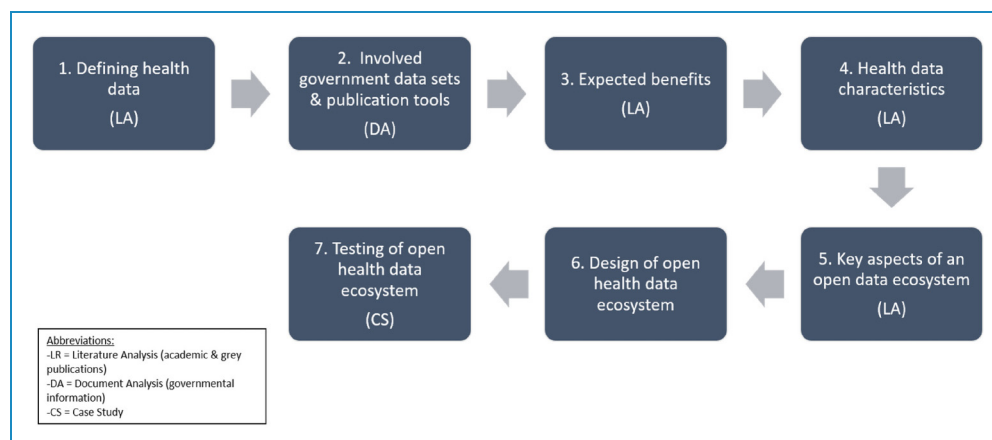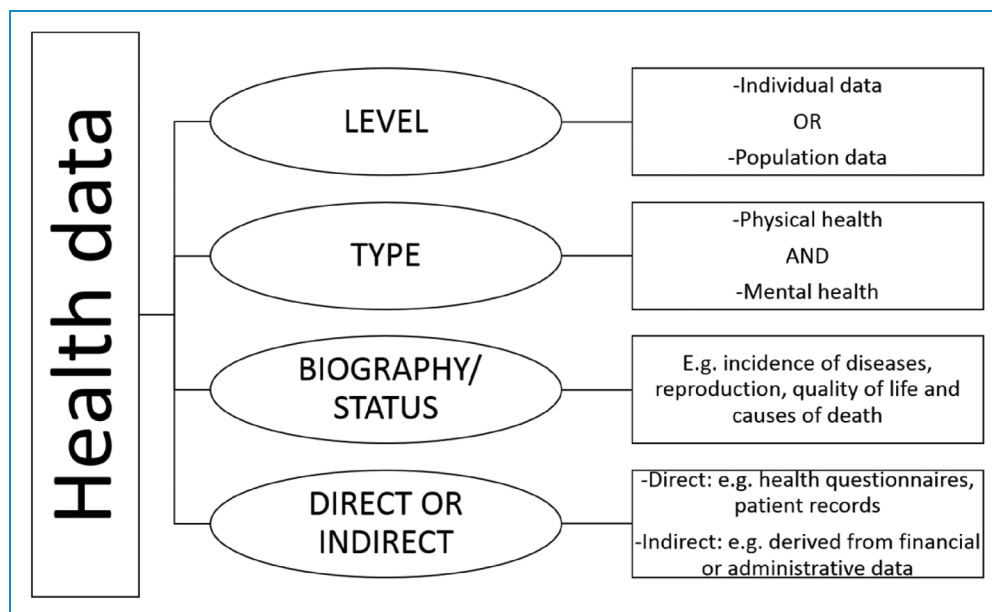

Figure 1. Study design.

Figure 2. Definition components of health data.

health.[17–19] This component stipulates the broad and modern interpretation concerning the concept of health. As a consequence, health data ranges from information on minor body ailments to information on emotional well-being.[20] Thirdly, definitions of health data incorporate a component related to a health biography or a health status. This means that health data comprise topics providing a historic and/or current view on the health of an individual or a population.[17–19] General topics include for instance the incidence of diseases, reproduction, aspects regarding quality of life, and causes of death.[16] In other words, health data encompass details concerning the presence or absence of health related problems for a person or population. Finally, a fourth definition component clarifies that information regarding the health biography or health status of an individual or a population can be either directly or indirectly derived. This component touches upon the sources of health data. Information on the health of an individual or a population might for instance be directly obtained via health questionnaires filled in by citizens or by examining the content of patient records maintained by health care providers. These patient records harness detailed information on patient demographics, clinical risk factors, diagnoses, immunizations, medications, and medical devices, medical test results, and care plans.[21] However, the information can also be indirectly induced when analyzing for instance financial data related to public or private health insurance and administrative data from the health care system such as hospital check-ins.[17–19,22] Medical cost payments and registrations regarding the use of health care services can unintentionally reveal if a person or a population suffers from a certain health problem.

In the context of open data, the topic health can also refer to health care data. Health data & health care data are closely related to each other. Health care data, a term the authors did not found frequently within their literature analysis, are defined as "that information used to provide, manage, pay and/or report on the services used across the entire health care system."[23] Health data are required to manage health care services and consequently could be considered as a part of health care data. A government administration will rely for instance on information concerning the health status of individuals and populations to foresee a proper level of qualitative health-related services. Nevertheless, health care data might be interpreted more broadly than a patient or a population their health status as these data could also entail information on the availability of health care providers in a region and/or the adherence to quality standards, policy guidelines, and regulations within the health care system. As the theme "health" is often not defined when referred to as a promising topic for open data policies, a wide meaning is applied in this primary exploration regarding an open health ecosystem.

Hence, this article choses a pragmatic approach to deal with the topic health in the context of open data. It aims to explore the opening up of all types of datasets linked to the topics health or health care which are managed by government agencies. Therefore, within this article, health data points out to any government data concerning information on the health of a population and/or the management of a health care system. Open data are by default anonymized data resulting in the fact that the data can only be at the population level. However, it is important to remember that in the case of health data the data often originate

from personal data. As will be discussed later, risks regarding the (re)identification of individuals cannot always be excluded. The working definition on open health data in Table 1 gives on the one hand meaning to the term health, on the other hand, it links this term to the characteristics of open data more specifically an unfettered accessibility.

Government data related to health: involved datasets and publication tools

In addition to the theoretical reflection on the definition of health data, the authors examine which datasets are already being published in practice and via which means. Several international initiatives which aim to stimulate the opening up of government data have considered health as a valuable subject for open data policies. For instance:

1. Health is one of the themes of the Open Government Partnership, a voluntary partnership between more than 70 governments and civil society representatives taking actions to improve governmental transparency, accountability, and responsiveness. The Open Government Partnership beliefs that data on the health and medical history of citizens, is a vital resource to improve health systems and patient care. They promote open health data for instance to empower citizens to have more choices and to take control of their own medical care.[24]
2. The leaders of the G8 countries (United Kingdom, France, Germany, Italy, Japan, United States, Canada, and Russia) included health in their Open Data Charter as one of the fourteen high-value areas in which data should be made available to provide transparency and to encourage innovative reuse of data. The charter includes two examples of health datasets to

open, namely prescription data and health care performance data.[25]
3. Human health & safety is one of the themes within the INSPIRE Directive, a legal framework from the European Union intended to make spatial data infrastructures more interoperable and stimulate the intergovernmental and public sharing of environmental spatial information.[26,27] Although the theme originally focused on the link between human health and the environment, member states of the European Union can choose freely to accommodate health data more broadly. Theme components within the INSPIRE Implementers Roadmap include for instance data on the prevalence of diseases, data on the availability of health care/health services, and health determinant measurement data.

Following these initiatives that promote open data policies, the authors expect to encounter some practical examples. Consequently, the creation of the health data ecosystem can also take into account existing examples. Secondarily, these examples provide an opportunity to assess the proposed working definition for open health data. Observing open data in practice, one can notice that already several governments publish few health datasets on their national, regional, or urban open data portals.[19,28] In these cases, health is one of many topics among for instance culture, tourism, or education on which open data can be found. As shown in Table 1, the health datasets on general open data platforms as the ones from New Zealand or the Canadian city of Surry range from statistics on smokers to the use of mental health services. Besides the inclusion of the health topic on general platforms, some limited examples exist of portals exclusively dedicated to the topic of health.[29] These health data platforms are intended to serve as a central location where existing health datasets are findable and depending on the dataset potentially accessible. Such portals like the ones in the United States,

Table 1. Overview of definitions.

	Definition	Source
Health data	Any data related to physical or mental health conditions, reproductive outcomes, causes of death, and quality of life for an individual or population which could be directly or indirectly derived	Chassang, 2017[22]; European Group on Ethics in Science and New Technologies, 1999[18]; OECD, 2015[46]; Segen, 2002[16]; Van Veen, 2018[17]
Health care data	Information used to provide, manage, pay and/or report on the services used across the entire health care system	Lin & Chan, 2000[23]
Open health data	Government data concerning information on the health of a population and/or the management of a health care system that is freely, legally, and technically open for (re)use	Authors of this paper

the State of New York, and Scotland are managed by governmental agencies responsible for health policies in their country or region (see Table 2). By means of these portals, the governments involved try to stimulate the reuse of health data by citizens, entrepreneurs, and researchers in order to enhance health outcomes for everybody.

Overviewing examples of datasets in Table 2, a simplified categorization of available open health data can be deduced. In current practice, open health data consists of datasets related to:

- The availability of health care practitioners (e.g. doctors, pharmacies, dentists) & health care services within a geographic area

- The quality or performance of health care: for instance measures regarding patient satisfaction, health care-related infections, and hospital readmissions
- Epidemiology: the prevalence of health-related problems such as cancer, diabetes, suicide, or influenza
- Health determinants, providing information on how factors as sex, ethnicity, health insurance status, or environment might influence the health status of citizens[30]
- Reimbursement schemes, health costs, and management of resources: prescription data, number of specific surgeries, medical products…[31]
- General health and population statistics: e.g. statistics about births, smokers, vaccination rates…

These derived categories may overlap with each other. For instance, depending on government goals certain

Table 2. Examples of portals with accessible and/or findable health datasets.

General portals			
Portal	Data.govt.nz (=the open data portal of New Zealand)	Surrey's open data site (=the open data portal of the city of Surry)	Datos.gob.es (=open data portal of Spain)
Examples of health datasets	Maternity clinical indicators, discharges from publicly and private funded hospitals, suicide facts, use of mental health services, certified health care providers	Availability of Doctors, current Smokers, low birth weight babies	User satisfaction concerning centers for addiction treatment, lists of pharmacies, death rates cancer, dentist visits according to socioeconomic groups
Source	Government Information Services (Department of Internal Affairs) and Stats NZ, www.data.govt.nz	City of Surrey departments and the GIS section, https://data.surrey.ca/	Ministry of Economy and Business, the Ministry of Territorial Policy and Civil Service et al. https://datos.gob.es
Portals related to health			
Portal	National Health System (NHS) Scotland open data (=the open health data portal from Scotland)	HealthData.gov (=the open health data portal from the United States)	Healthdata NY (=the open health data portal form the State New York)
Examples of health datasets	Prescriptions in the community, Body Mass Index statistics, cancer waiting times, alcohol-related hospital statistics, psychiatric inpatient activity, infant feeding, births in Scottish Hospitals	Surveys on patient experiences, heart attack payments, hospital readmissions, adult depression, cancer incidence, vaccination rates, health care-associated infections	Hospital inpatient dischargers, nursing home costs, cardiac surgery by hospitals, diabetes related indicators, vaccination statistics, long-term care performance data, influenza laboratory-confirmed cases
Source	NHS National Services Scotland, www.opendata.nhs.scot/	U.S. Department of Health and Human Services Office of the Chief Technology Officer, HealthData.gov	New York State Department of Health, https://healthdata.ny.gov/

epidemiology data, data on health determinants, or population statistics are also useful to monitor health care performance. This can be the case when governments strive to diminish the prevalence of certain diseases, the health improvement of vulnerable population groups, or the spread of antitobacco measures. The observed categories are in line with our working definition of open health data as they comprise both information on the health status of a population as management aspects of a health care system.

Expected benefits of open health data

As mentioned earlier, health is a theme that is strongly promoted to include in open data policies of governments. How to explain this enthusiasm of open data advocates concerning health? Gaining insights into the expected benefits will help to incorporate certain interests within the ecosystem. Within literature regarding innovative data use, the expected benefits of open health data can be summarized as the ability for governments to make better-informed decisions for policy making, the development of health(care) related products by the private sector, more knowledge acquisition on the health care system by citizens/citizen groups and the use of data for science.[29,32,33]

Drawing upon available data regarding aspects as evolutions within population health or the outcomes of treatments, policy makers are able to make justified decisions concerning health care like tackling the underuse or overuse of specific medical interventions. By making relevant data available for several policy levels and entities, open data policies have the power to overcome potential reluctance on intergovernmental data sharing which is widely observed within e-government literature.[34] As the competence for health (care) policies, is often divided between distinct government levels or agencies open data is a helpful tool to obtain more quickly and easily access to data.

Data is a crucial resource for industry to determine which domains are worth investing, to discover opportunities and create innovative products ranging from medicines to medical measure equipment or health apps.[35] Current barriers to data access for industry include for instance rigid legal frameworks, restrictive requirements regarding partnerships with governmental data owners or distrust caused by isolated examples of data misuse.[36] Therefore, having access to more health data via open data policies will support certain business activities. An open dataset can be used for a single application, or it can be linked with other datasets potentially generating big data. Employing big data analytics for single or combined datasets proves inter alia promising for the pharmaceutical industries to improve their stagnant R&D and customize innovations for instance in the field of personalized medicine.[37,38]

Transparency concerning the supply of health care services and their quality provide citizens and nongovernmental organizations on the one hand the opportunity to match their health care demands. For instance: the use of open data to answer questions as "Where to find suited care providers for his/her health problems" and "Are there any waiting times?" On the other hand, transparency makes deficiencies within the health care system visible, valuable information that can be used to claim improvement actions from responsible governments.[29]

Lastly, open health data facilitates scientific health projects. The slogan "Data saves lives" is often used to stress the importance of data for medical knowledge and to promote conditions that facilitate access to health data for science.[39] Without data, the causes of several diseases such as smoking and lung cancer would have never been discovered.[31] In this context, open health data avoids certain administrative, legal and financial hurdles complicating scientific health research. Certain datasets of government agencies become suddenly accessible without facing complex and long request procedures.

The characteristics of health data

Despite the observation of limited open health data examples and the high expectations regarding benefits, several barriers and challenges exist to make (more) health datasets publicly accessible. To comprehend the tension between expected benefits and obstacles concerning opening up health data, the distinctiveness of this type of data should be investigated. Which are the characteristics of health data potentially impeding or limiting its openness?

Two first characteristic of health data is that the data are *extremely personal* and *subject to the risk of misuse*. It reveals intimate information about the physical and mental status of persons, information which in the wrong hands could bring harm to the person involved. Health information could lead for instance to situations of stigmatization *(e.g. HIV or mental problems)* or the denial of certain services or chances *(e.g. insurances or jobs)*. Therefore, health data is generally considered as a special category of personal data *requiring severe privacy and data protection measures*.[22,40] Although, open data normally applies to anonymized data potential risks to (re-) identify persons should always be taken into account before making datasets open. (Re-)identification of groups and individuals within anonymized, for instance by using advanced data mining techniques, cannot always be excluded.[14,41,42] As such, not all data should be opened by governments.[9,43] Cautiousness certainly applies when data is as sensitive as health data.

Related to the characteristics of extreme personal and subject to the risk of misuse concerns the characteristic that *health data raises ethical questions*. Shocked by medical research malpractices during the Nazi regime,

several bioethical frameworks emerged in the second half of the last century.[14] Within bioethical debates and principles, topics as informed patient consent concerning the use of their health data, exceptional conditions which allow the use without informed consent, and legitimate purposes for data use are thoroughly addressed.[17,44] The ethical dimension is for instance present in the preevaluation of medical science projects, where the collection and/or use of health data is commonly dependent on approval by an ethical committee.[14] Subsequently, ethical aspects impact decisions to open health data or not. To which degree do patients need to know and agree that their anonymized data will be accessible, will the data be reused in the benefit of the public good, etcetera?

Health data are data with a *high commercial value*. McKinsey Global Institute claims for example that *medical clinical information providers, which aggregate data and perform the analyses necessary to improve health care efficiency could compete in market more than 10 billion dollar by 2020.*[45] Data driven innovation could support savings within the health care *(for instance by detecting fraud within health insurance or overuse of care)* but at the same time it might also generate costs as it boosts the development of advanced and expensive therapies or medical equipment.[37,46,47] For many actors, health data are a key resource for big business. This raises questions as who profits from opening up health data and are these benefits less or more equally shared.

Health data are *surrounded by a broad field of actors with specific institutions*. Primary sources of health data are inter alia electronic medical records, diagnostic results, hospital bills, prescription claims, and health survey questionnaires.[48] These sources often feed the health datasets used by governments to gain knowledge on public health and the health care system. Collecting data from the primary sources depends on negotiations between government and data providers (e.g. hospitals, pharmacies, nursing homes), legal obligations, and/or patient consent. The power of the stakeholders such as doctor associations or patient organizations and cultural views concerning public health research may impact the amount and availability health datasets in a country.[49] The volume and variety of actors involved results in *discussions about ownership*[33]: for instance does data extracted from an electronic medical record belongs to the patient or to his caregiver? Or should there be as some propose a kind of solidarity regarding the sharing of anonymized or pseudonymized health data as it could lead to new medical insights and health care improvements?[17,22]

Ecosystems and OGD

Having analyzed and outlined some foundations of health data *(namely definition, dataset examples, user benefits & characteristics)* in the previous parts of this article, the next parts of this paper will focus on the concept of open data ecosystem. Before applying ecosystem to the subject health some background information on open data ecosystems is presented.

Within open data literature, the metaphor of an ecosystem is often used to map the flow of actors, activities, and (technological) tools constituting an open data environment.[15] The metaphor, derived from nature, stipulates how interrelationships between people, infrastructure, and institutions are crucial for either great or limited success of open data settings. Van Loenen et al.[4] define an open data ecosystem as "a cyclical, sustainable, demand-driven environment oriented around agents that are mutually interdependent in the creation and delivery of value from open data."

Researchers have inter alia applied the ecosystem metaphor to support the design and assessment of OGD programs,[11] to compare similarities and differences of national open data government policies[12] and to detect essential factors enabling innovation with open data.[50] As illustrated in Table 3, examples of open data ecosystems find in literature consist of different elements. Considering the elements of these existing open data ecosystems, they show commonalities which will be discussed next. First of all, the ecosystems are composed around stakeholders and their interests concerning open data. Traditionally, stakeholders consist of the governmental organizations producing the data, the actors that use the data to create added value and eventually the consumers of products and services created with open data. Each actor has their own perspectives and possibilities to engage in open data, different interests have to be appeased. User needs are for instance measured or gathered via specific studies, public consultations, expert groups, or online platforms to submit ideas.[13] The second commonality of open data ecosystem models, is the presence or lack of information policies. It refers to legislation, practices, and policies influencing the opening up of government data.[51] Thirdly, data preparation activities are a shared part of open data ecosystems. Prior to publicizing any dataset for reuse, preparations are required to enable data reuse. These include inter alia assessments of the data quality, provision of metadata, and choosing appropriate data formats.[52] Fourthly, each open data ecosystem mentions the infrastructural elements. Portals to find and/or access government datasets and tools regarding data analysis or visualizations are typical examples. Lastly, there are the drivers giving impulse to become involved in open data. Global trends, culture within government administrations, demands from stakeholders or ambitions on the level of economy, and transparency may for instance motivate policy makers to open up data. Dynamic relationships between these common elements impact the outcomes of open data initiatives over time. For example, feedback loops from open data users, foreign hypes regarding

Table 3. Examples of open data ecosystem elements.

Source	OGD ecosystem of Zuiderwijk et al. (2014)[50]	Ecosystem model of OGD programs of Dawes et al. (2016)[11]	OGD ecosystem comparison research of Styrin et al. (2017)[12]
Elements	*Tools and services:* OGD portals, programs, tools, and services *Data producer and contextual level:* OGD policy and legal frameworks, culture, technology, financing, organizations, resources, provision of data and metadata to the public by governmental organizations *Data user level (e.g. citizens and businesses):* activities, use of tools to analyze data	OGD producers (political leaders, administrative agencies), OGD users (transparency advocates, civic technology community), OGD beneficiaries (consumers of OGD products & services) *National & global trends:* e.g. demands for transparency, economic competition *Motivation for OGD development* *OGD Policies and Strategies:* e.g. legal, resources *Data publication* *Data use and products* *Feedback and communication mechanisms:* e.g. opinion polls, requests for new datasets, invitations for OGD hackatons *Benefits:* created value from data use *Community characteristics:* e.g. nature of the civil sector, climate for innovation	*Government policy and practices:* e.g. programs, legislation, monitoring of policy outcomes *Government data management:* the implementation by agencies and implementation processes regarding aspects as releasing and publishing data, data quality measures, data analysis *Stakeholders engagements:* how do developers, entrepreneurs, and citizens engage?

OGD: open government data.

innovative data use, or technological progress might influence open data policies in ways of organization, amount of published datasets, or the available budget.

Mapping the open health data ecosystem

By making use of the acquired knowledge on health data and inspired by existing literature on OGD ecosystems, the generic ecosystem components will be mapped to the matter of health data. What constitutes the socio-technical environment in which health data managed by governments are opened up or potentially stay closed? This environment and its dynamics will be mapped according to the commonalities of open data ecosystems: (a) stakeholders and their interests, (b) information policies, (c) data preparation activities, (d) infrastructure, and (e) drivers. The mapping of the open health data ecosystem is summarized in Figure 3. This section ends with a reflection on the mapping result in relation to the generic health data ecosystems.

Stakeholders and their interests

Regarding stakeholders, a first actor includes the governmental data producers which manage health datasets. Diverse government agencies or institutions can be responsible for aspects of health policy, a wide-ranging policy domain covering multiple aspects. Aspects include for instance the funding and the quality of the health care system, the utilization, and evaluation of pharmaceuticals and medical technologies in a country, the surveillance of (infectious) diseases, the prevention of certain health problems, and the encouraging of healthy lifestyles, the implementation of measures to diminish health inequities within a population, etc.[53] The organization of health policies differs between countries, which means the responsibilities of health policy can be either centralized or decentralized both on the level of state structure and the level of governmental departments. Therefore, cautiousness regarding generalization should be applied. Nevertheless, considering these sub-domains of health policy we expect that government agencies or government institutes dealing with public health insurance, the supply and audits of health care services, the regulation of pharmaceuticals and medical devices, disease control, health promotion, health technology assessments, public health research and general national statistics *(e.g. births, deaths)* represent the data producers in a health data ecosystem. They manage data which are either generated by the government agency themselves *(such as data collected during a quality audit of a health care service)* or data which are provided by citizens *(for instance in the case of national health surveys or patient experience questionnaires),* individual care providers *(e.g. doctors, pharmacists, nurses),* care organizations *(e.g. hospitals, nursing homes)* and the medical industry.
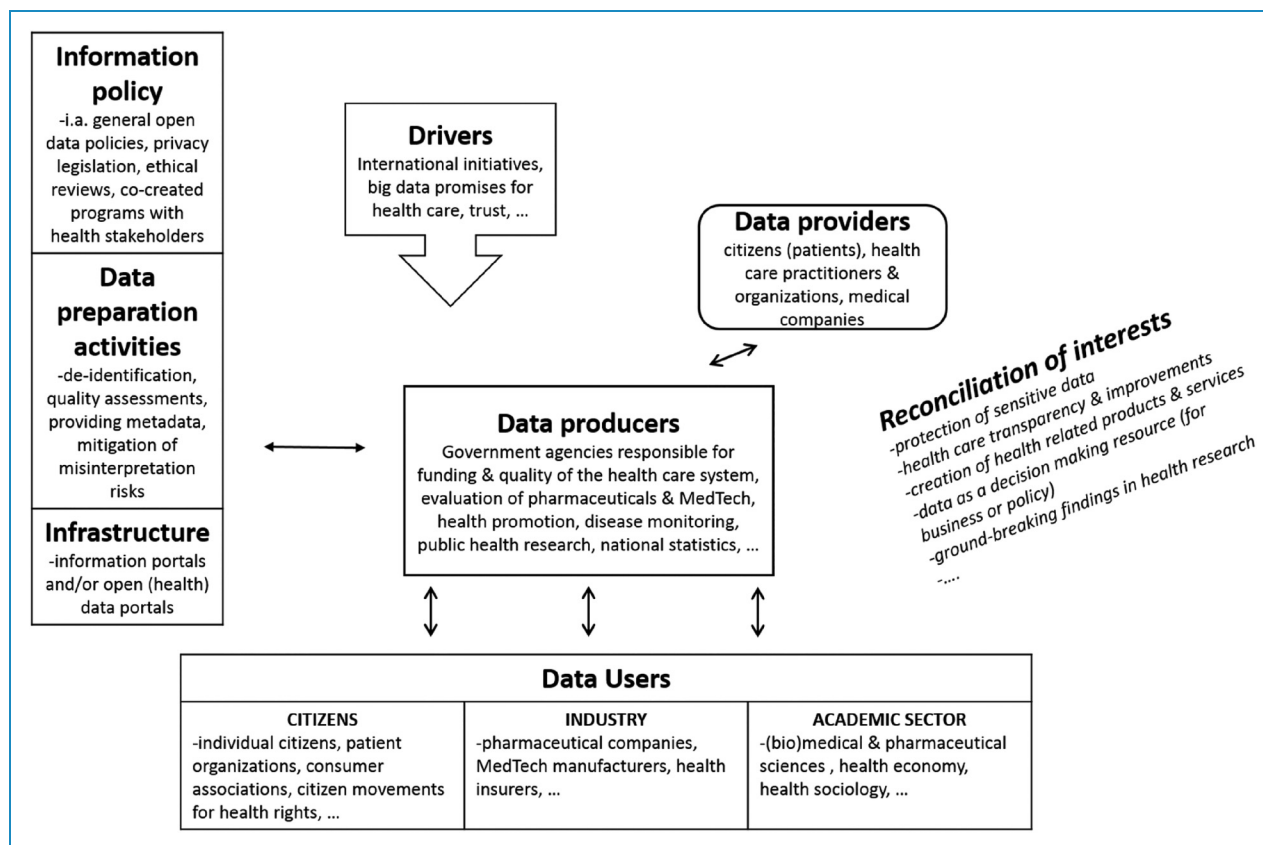
Figure 3. The open health data ecosystem.

When not generated directly by the government agency they are dependent on the legal obligations concerning data provision towards them or willingness of the data providers when the data collection is not mandatory. The level of cooperation by the data providers—potentially influenced by data sensitivity, ethics, or reimbursement of health care—is crucial to obtain qualitative data. The open health data interests of the governments agencies include the protection of sensitive data, transparency concerning the health care system, intergovernmental data sharing, and the external creation of economic value from government data. In contrast, the interests of the data providers can be altruistic (e.g. contributing with data to new medical insights), legal (e.g. executing legal obligations concerning data registration) or protectionist (e.g. on the level of privacy, ethics, or disclosure of one's health care service performance).[44,54]

On the side of the data users, we distinguish citizens, the private sector, and the academic sector. Regarding citizens, a further division into individual citizens and citizen groups is made. An individual citizen, for instance in need of health care, could use open health data to obtain information about the availability of caregivers or the quality of the health care system. However, the open datasets could be too raw for data laypersons and/or difficult to interpret. Therefore, the

individual citizen might require support from open data products (e.g. a health app or informative website) possibly intended to help him make use of the data according to his specific needs and capacities. Data products are offered by creative citizens, firms, government agencies, or nonprofit organizations. Individual citizens could also utilize the open data or open data products for participating in health policy. However, citizen groups are because of their unified forces and pooling of resources presumably better organized and strengthened to exploit open health data. Such pooling of financial and human capital facilitates participation in (complex) health policies and gaining useful knowledge concerning the performance of the health care system.[55] In the case of health data, patient organizations, consumer associations, and/or citizen movements striving for the right to health should be able to profit from open health data considering activities in their interest such as informing citizens about health related subjects and lobbying for health care improvements.[56] On the other hand, they might also have interest to prevent data sharing because of privacy concerns.

The second type of data users concerns the industry. In the case of health data, the industry likely to be interested in open health data consists of pharmaceutical companies, firms related to medical devices and technologies (the

so-called MedTech manufacturers), private health insurance companies, etc.[47,48,57] Their use of open health data could either immediately lead to value as is the case when its utilized to decide about areas worth investing or the creation of uncomplicated health products or services. Nevertheless, as the development and market introduction of pharmaceuticals and medical devices is due to safety reasons highly regulated and subject to lengthy trials, some products partially generated with the aid of open health data could take years to become market proof and available to end-consumers. The interest of the industry concerns the facilitation of business activities by employing open health data as a free-available resource.[58]

In case of the third user, the academic sector, first thoughts go out to (bio)medical and pharmaceutical sciences which could employ the data for research concerning inter alia the causes, the evolution, and treatment of health-related problems or the organization and adjustment of medical practices.[33,57] However, considering the broad definition of health policy and existing health datasets, other academic disciplines should be added. Open health data are for instance also valuable for health economists, investigating which investments in health (care) give the greatest return,[59] or health sociologists interested in health inequities within a population. The academic use of open health data is linked to their interests concerning research publications and groundbreaking findings related to health(care).

Information policies

An open health data ecosystem is the one hand surrounded by general open data policies developed by a national, regional, or local government. Within these general open data policies concerning legal and policy frameworks to stimulate the (re)use of government data, nonsensitive health datasets (e.g. birth statistics or availabilities of hospitals) are potentially included leading to a relatively trouble-free data accessibility. However, due to the peculiar characteristics of several health datasets some specific policy frameworks shall be applicable. Within these frameworks items such as the goals, resources and responsible actors of the open health data policy need to be addressed.

A first framework comprises privacy legislation aimed to protect citizens with regard to the processing of their personal data. Although open data only includes anonymized data, health data often result from personal data. Privacy legislation determines strict conditions under which personal data such as data on an individual's health are whether or not allowed to be collected, stored, adapted, consulted, combined, etc.[22] Furthermore, these legislations might impose certain data protection measures and principles such as storage limitation having an impact on the management and opening of health datasets. Before opening up anonymized health datasets, risks assessments concerning the reidentification of individuals might be direct or indirectly imposed by privacy legislation.[17] Furthermore, advice or authorizations from national data protection authorities will in some cases be recommendable.

The opening up of health datasets is also be bounded by ethical frameworks considering questions such as "Can the data be (re-)used for the Public Good?" Like in medical science projects, government agencies could request an opinion from an ethical committee regarding the opening up of certain datasets. Otherwise, relevant recommendations of authoritative bodies such as (inter)national bioethical boards or deontological commissions of medical caregivers might influence government agencies in their decision process regarding the free availability of particular health datasets.

Lastly, as health data are provided by a large set of stakeholders with their own institutions, an open health data policy depends on a negotiation framework geared towards consensus in order to receive and publicly share data related to health. As health data is considered to be sensitive data, agreements or conventions with the health care sector and cocreated action plans assist governments to avoid misperceptions and negative publicity regarding the opening of health data sets.

Data preparation activities

Keeping the characteristics of health data in mind, expected data processing activities pertain to privacy safeguards and data quality. Some health datasets involve nominative or pseudonymized data. Prior to opening up these datasets, a process of data de-identification is necessary. This involves the detection, removal, or modification of information that could lead to personal identification.[40] Some information requires for instance a higher level of aggregation to avoid unintended disclosure of personal data. When working with smaller populations (for instance patients with a rare disease) challenges can arise or even inhibit the opening up of certain health data sets.[14,60]

Another important preparation activity preceding the opening of health data includes the assessment of quality. Medical records of caregivers are in several cases the primary source of the government health dataset. As this source is originally intended to support clinicians in diagnosing and treating their patients other objectives such as reporting for public health matters might not always correspond. Furthermore, skepticism exists concerning the completeness and accuracy of medical records.[42] Consequently, the manager of the open health dataset needs to evaluate if the dataset is fitted enough to use in an open data context where (re)use of data could be used for a wide variety of purposes.

Metadata is required to facilitate correct interpretations of the data and to give potential data users information about the opportunities and limitations of the health

dataset. Multiple studies have already proven cases regarding public misinterpretation of health data.[41] As misinterpretation might do harm, the opening of a dataset should evaluate the perils of misinterpretation. Sathianathen et al.[61] showed for instance how misinterpretation of publicly reported surgical outcomes could lead to disadvantageous care choices by (vulnerable) citizens. The provision of sufficient contextual information and assistive data visualization tools may possibly mitigate the risks of misinterpretation.

Infrastructure

Concerning infrastructure we refer to part 3 of this article where it was observed that portals serve as a useful instrument for open health data policies. Two types of portals were detected: on the one hand open data portals giving direct access to health datasets managed by the government, on the other hand information portals designed to give an overview of available data on the government side and by whom these datasets are managed.[13] A combination of both types is also possible, namely a portal providing direct access to some health datasets and an overview of other existing health datasets which do not have direct accessibility.

Drivers

Various factors drive the creation or enhancement of open data policies. A government can be inspired by the practices of another government, recommendations by supranational organizations, or international data sharing trends The United States was one of the first countries to explicitly, actively, and visibly include health in their OGD ambitions. Around 2010, their opening up of several health datasets, formerly closed to the public, were for instance widely applauded.[32]

Concerning other drivers for open health data policies, one can think of the current huge promises regarding the potential of big data projects related to inter alia the analysis of disease patterns, a faster tracking of disease outbreaks, health claim fraud detection, and the development of customized treatments for individual patients.[38,42] These projects require accessibility to lots of health data. Do governments believe in such data innovation projects and are they willing to share health data with actors wanting to execute such projects like academics, pharmaceutical companies, and health insurers?

As health data is due to its characteristics considered to be sensitive, trust will be a key driver promoting or confining the opening of health datasets. Is there a strong belief that the health datasets will generate societal benefits (for instance breakthroughs in cancer research)? Or is the perceived danger regarding misuse of health data stronger? Positive or negative attention concerning health data

initiatives could influence further policy outcomes. History already showed that distrust in projects concerning the sharing of health data has the potential to scuttle these intended projects.[44]

Reflection on the mapping result

The common elements of generic open data ecosystems proved to be useful to customize an ecosystem dedicated to the theme health data. It allowed to identify fundaments of an ecosystem namely people, infrastructure, and institutions related to health(care). Certain particularities within the ecosystem dynamics related to health data lie in

- the diversity of governmental data producers and their high dependency on (nongovernmental) data providers to gather qualitative data for several datasets
- lower visibility of derived open data products because certain health products require heavy market authorization procedures due to safety concerns
- the presence of bioethical frameworks and relevant governance authorities that might influence data access
- the execution of stringent de-identification processes (if the data are derived from personal data) compared to many government datasets with none or low identification risks
- the existence of portals that make existing datasets transparent but not directly accessible
- the impact of societal debates about the benefits or dangers of data sharing (in comparison to other data types that are less sensitive)

Testing the open health data ecosystem

Having mapped the open health data ecosystem, the next research goal is to test the system via a case study in which an open health data policy was introduced. The selected policy concerns the initiative "Data for Better Health" of the federal government from Belgium launched in 2018. The policy aims to facilitate the reuse of the many existing data resources in health and health care. It believes that a more intensive use of these health (care) databases, developed by or on behalf of the government, will lead to more knowledge and innovation in the field of public health.[62]

Being a recent policy, the case allows us to examine the ecosystem dynamics at an early stage where several stakeholders laid the first foundations of an open health data environment. The sources used to study the case comprise the website dataforbetterhealth.be, the portal fair.healthdata.be, government documents *(related to topics such as open data, ehealth, collaboration with the pharmaceutical industry),* and publications from interest groups *(e.g. patient organizations, federations of the technological industry).* Furthermore, contacts with civil servants directly engaged

in the *Data for Better Health* provided additional information to comprehend the policy actions.

Data for Better Health was initiated by the Federal Minister of Public Health, the Federal Minister of Digital Agenda, and the State Secretary for Privacy. It was supported by Sciensano (a governmental institute with expertize in public health research), NIDO (the innovation lab of the federal government), and DigitYzer (a digital Innovation Hub within Belgian's capital Brussels). The initiative wanted to be inspired by a broad network consisting of (1) health care administrations *(Federal Agency for Medicines & Health Products, National Institute for Health and Disability Insurance)*, (2) patient organizations, (3) knowledge institutions (Inter-university Micro-Electronics Center & the Belgian Health Care Knowledge Center), (3) public health insurers, (4) representatives of hospitals and (5) federations of the pharmaceutical and technological industry. Actors within this network were involved in workshops and working groups giving further development of the open health data policy. Other stakeholders, which were invited and present at workshops and hackatons, included open data communities and data driven startups in health care. The policy was inter alia preceded by general government open data ambitions,[63] hearings in the House of Representatives about big data in health care, aspirations from industrial actors, and several government strategies concerning Belgian's future health care. Influenced by examples abroad (for example the US portal HealthData.Gov) and inspired by the "FAIR Guiding Principles for scientific data management and stewardship" the *Data for Better Health* initiative took shape. The FAIR data principles which refer to making data findable, accessible, interoperable and reusable guided the policy objective to improve the reusability of health data.[64] Although *Data for Better Health* was affiliated with open data policies, it should already be noted that FAIR data are not always open data. For example, while open data should be available for everyone the accessibility of FAIR data can be limited to appropriate people, at an appropriate time, in an appropriate way.[65]

Data for Better Health consisted of several events and actions. Examples include a workshop to introduce government data managers to the principles and practices of FAIR and open data and a hackaton where the National Institute for Health and Disability Insurance opened the DOC PH database containing information about medication delivery in all Belgian hospitals since 1994. The events and actions were always aligned with the European Union General Data Protection Regulation in order to guarantee privacy and legislation concerning the reuse of public sector information.

Similarly, three portals were introduced. The Belgian FAIR-portal (fair.healthdata.be) provides a transparent metadata-catalogue of public health databases. This information portal shows which health related databases exist,

the person of contact, information concerning access, etc. Related to this FAIR-portal a web application metadata-healthdata.be was developed. This metadata-portal allows data managers via user accounts to update the metadata of their database and to add new data collection projects. Their input is automatically published at the FAIR-portal. Data managers were invited to complete the metadata of the databases for which they are responsible. Although, several already did a large amount of them still proved to be little responsive, having doubts concerning data quality or not believing in the benefits of open data. Thirdly, a request-portal was put into production where interest data users can fulfill a request form to demand access to databases that are indexed at fair.healthdata.be. The awareness, accuracy, and user-friendliness of these portals are considered to be areas for improvement. In this early stage of the *Data for Better Health initiative* an open data portal giving direct access to several datasets still seemed a bridge too far.

Considering the case study results, the developed open health data ecosystem is viable to identify the elements that constitute a socio-technical environment promoting or obstructing the opening up of health data (see Table 4). While less emphasized in our mapped system, more general open data aspects like the presence of open data communities and data driven startups or the applicability of legislation concerning reuse of public sector information were also noticed within the case. Hence, one needs to remind that within a specific health-related ecosystem general open data aspects keep playing a role.

Conclusions and future research

Former research on OGD ecosystems stated that the actors and dynamics of such ecosystems could differ according to the policy domain in which data are produced. Based on knowledge concerning the peculiarities of health data, this article aimed to map and test an open data ecosystem concerning the policy domain of health. We managed to create and assess an open health data ecosystem consisting inter alia of particular stakeholders, interests, information policies, and activities. Particularities regarding an open data ecosystem for the policy domain of health include inter alia de-identification activities, (bio)ethical assessments, and the specific role of data providers (*e.g. patients, hospitals, caregivers)* who deliver the data for some government-managed databases. Due to privacy and ethical concerns, several health datasets will not be opened for every actor and for every use. This is reflected in some government portals making existing health datasets findable to the public but not directly accessible. As such the concept of open data does not always fully apply to the policy domain of health in which desires for data access are sometimes dependent on request procedures and in which some applications of data reuse are potentially more interesting with pseudonymized data including richer information

Table 4. Ecosystem elements of the data for better health initiative.

Drivers	National open data ambitions (Digital Belgium) Expectations concerning data driven innovation in Health care (e.g. hearings at the Chamber of Representatives, aspirations of Belgian industrial interest groups, ...) International initiatives (e.g. HealthData.gov US) Strategies concerning a sustainable, future-proof Belgian health care system
Stakeholders	Data providers such as representatives of patients and hospitals were involved in a workshop intended to gather challenges & concerns Data Producers: inter alia Federal Agency for Medicines & Health Products, National Institute for Health and Disability Insurance, Sciensano, Belgian Health Care Knowledge Centre Data users: the participants of the hackaton and the early users of the data request-portal consisted mainly of pharmaceutical companies, research institutions, medtech companies, data driven startups, and governmental organizations. In this early stage, use by citizens, patient organizations, or other civic organizations is rather limited
Information policy	National Open Data Program (Digital Belgium) European Union General Data Protection Regulation European Union and national legislation concerning the reuse of public sector information FAIR principles as guideline Cocreated policy initiatives with health care stakeholders (i.a. government agencies, patient organizations, industry, public health insurers, hospitals, knowledge institutions)
Data preparation activities	Providing metadata for the FAIR-portal (fair.healthdata.be) Anonymizing DOC PH database for hackaton
Infrastructure	FAIR-portal (fair.healthdata.be) Metadata-portal (metadata.healthdata.be) Request-portal (requesthealthdata.be)

then with completely anonymized data. The authors expect that potentially more data sets could be made open after an anonymization process, however, a high potential lies in conditional access to pseudonymized data to stakeholders with acceptable interests. Acceptance of specific interests will require societal debate. Another particularity lies in the fact that some benefits of open health data reflected for instance within services or products derived from the data are not always as quickly visible in comparison to other types of data. While some open government geographical datasets for instance resulted in multiple apps assisting people to find a city parking place, some products related to health are because of human safety reasons bounded by strict and lengthy market entry procedures.

Considering our observations and the obtained understanding of the broad domain of health policy, it needs to be mentioned that the dynamics of an open health data ecosystem could vary according to the health dataset. An administrative dataset concerning for instance an overview of hospitals in a region will likely face less constraints to open up than a dataset related to a disease registry containing for instance information on the prevalence of a health problem affecting only a small population. Depending on the data sensitivity and risks regarding reidentification subdivisions in the open health data ecosystem may arise. Within existing literature on the topic of open health data distinction between types of health data is often neglected, nevertheless, it could be recommendable for future research. Which sort of health datasets are easily to publicly publish and which ones should have restricted access policies because of privacy risks?

The selected case study did not allow to more deeply investigate the degree of participation by potential data users. At first instance, most interested users of the health data seemed to come from the side of the industry to support business activities. Do other potential users have in

practice enough capacity, time and skills concerning data processing? Additional research could focus on how other potential users such as patient and citizen organizations may employ the data to inform citizens about health(care) aspects and to strive for health care improvements. Knowledge of the resources required to exploit the benefits of open health data can enrich the described ecosystem.

Finally, an in-depth comparison with other types of data that are considered to be sensitive like security data or financial data can be valuable in putting certain characteristics and dynamics regarding health data into a broadened perspective.

ORCID iD: Roel Heijlen ⓘD https://orcid.org/0000-0001-8653-3139

References

1. Harlan Y and Robinson D. The New ambiguity of 'open government'. *UCLA L Rev Disc* 2012; 59: 178–208.
2. Kitchen R. *The data revolution: big data, open data, data infrastructures and their consequences*. London: Sage, 2014.
3. Ubaldi B. *Open government data: towards empirical analysis of open government data initiatives*, *OECD Working Papers on Public Governance*, No. 22. Paris: OECD Publishing, 2013.
4. Van Loenen B, Vancauwenberghe G, Crompvoets J, et al. Open data exposed. In: Van Loenen B, Vancauwenberghe G and Crompvoets J (eds) *Open data exposed*. Switzerland: Springer, 2018, pp.1–10.
5. Attard J, Orlandi F, Scerri S, et al. A systematic review of open government data initiatives. *Gov Inf Q* 2015; 32: 399–418.
6. Open Knowledge International. Global Open Data Index, https://index.okfn.org/ (2019, accessed 28 July 2019).
7. Cerrilo-i-Martinez A. Fundamental interests and open data for re-use. *Int J Law Inf Technol* 2012; 20: 203–222.
8. Gonzalez-Zapata F and Heeks R. The multiple meanings of open government data: understanding different stakeholders and their perspectives. *Gov Inf Q* 2015; 32: 441–452.
9. Janssen M, Charalabidis Y and Zuiderwijk A. Benefits, adoption barriers and myths of open data and open government. *Inf Syst Manage* 2012; 29: 258–268.
10. Martin S, Foulonneau M, Turki S, et al. Risk analysis to overcome barriers to open data. *Electron J e-Gov* 2013; 11: 348–359.
11. Dawes SS, Vidiasova L and Parkhimovich O. Planning and designing open government data programs: an ecosystem approach. *Gov Inf Q* 2016; 33: 15–27.
12. Styrin E, Luna-Reyes LF and Harrison TM. Open data ecosystems: an international comparison. *Transform Gov: People, Process Policy* 2017; 11: 132–155.
13. Vancauwenberghe G and Crompvoets J. Governance of open data initiatives. In: Van Loenen B, Vancauwenberghe G and Crompvoets J (eds) *Open data exposed*. Switzerland: Springer, 2018, pp.79–100.
14. McGrail KM, Gutteridge K and Meagher NL. Building on principles: the case for comprehensive, proportionate governance of data access. In: Gkoulalas-Divanis A and Loukides G (eds) *Medical data privacy handbook*. Switzerland: Springer Internatonal Publishing, 2015, pp.737–764.
15. Harrison TM, Pardo TA and Cook M. Creating open government ecosystems: a research and development agenda. *Future Internet* 2012; 4: 900–928.
16. Segen JC. *Concise dictionary of modern medecine*. Michigan: McGraw-Hill, 2002.
17. Van Veen EB. Observational health research in Europe: understanding the general data protections regulation and underlying debate. *Eur J Cancer* 2018; 104: 70–80.
18. European Group on Ethics in Science and New Technologies. *Ethical issues of healthcare in the information society*. Opinion no 13, Brussels: European Commission, 1999.
19. OECD. *"Health data governance. Privacy, monitoring and research." Health policy studies*. Paris: OECD Publishing, 2015.
20. World Health Organization. Mental health: a state of well-being, https://www.who.int/features/factfiles/mental_health/en/ (2014, accessed 25 April 2019).
21. Professional Records Standard Body. *Standards for the structure and content of health and care records*. Leeds: NHS Digital, 2018.
22. Chassang G. The impact of the EU general data protection regulation on scientific research. *Ecancer* 2017; 11: 709.
23. Lin B and Chan HG. Managing data quality in the health care industry: some critical issues. *J Int Inf Manag* 2000; 9: Article 4: 33–55.
24. Open Government Partnership. Health & nutrition, https://www.opengovpartnership.org/theme/health-nutrition (2019, accessed 31 January 2019).
25. UK Government Cabinet Office. G8 open data charter, https://opendatacharter.net/g8-open-data-charter/ (2016, Accessed on 15 June 2019).
26. European Commission. INSPIRE data specification on human health and safety – technical guidelines, https://inspire.ec.europa.eu/id/document/tg/hh (2019, accessed 18 June 2019).
27. Lyseen AK and Hansen HS. INSPIRE compliance of public health information: a Danish case study. *Int J Spatial Data Infrastruct Res* 2015; 10: 84–102.
28. Dong H, Singh G, Attri A, et al. Open data-set of seven Canadian cities. *IEEE* 2017; 5: 529–543.
29. D'Agostino M, Samuel ON, Sarol MJ, et al. Open data and public health. *Rev Panam Salud Publica* 2018; 42: 1–7.
30. World Health Organization. Determinants of health, https://www.who.int/hia/evidence/doh/en/ (2019, accessed 15 June 2019).

31. Van Veen EB. *Patient data for health research. A discussion paper on anonymization procedures for the use of patient data for health research*. Den Haag: MedLawconsult, 2011.

32. Conway PH and VanLare JM. Improving access to health care data. The open government strategy. *JAMA* 2010; 304: 1007–1008.

33. Kostkova P, Brewer H, de Lusignan S, et al. Who owns the data? Open data for healthcare. *Front Public Health* 2016; 4: Article 7: 1–6.

34. Van Cauter L. *Government-to-government Information system failure in Flanders: an in-depth study*. PhD Thesis, Katholieke Universiteit Leuven, BE, 2016.

35. Tang C. *The data industry. The business and economics of information and big data*. Wiley, 2016.

36. Cole A and Towse A. *Legal barriers to better the better use of health data to deliver pharmaceutical innovation*. OHE Consulting Report, London: Office of Health Economics, 2018.

37. Cattel J, Chilikuri S and Levy M. *How big data can revolutionize pharmaceutical R&D*. McKinsey Center for Government, 2013.

38. Raghupathi W and Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014; 2(3): 1–10.

39. Coppen R, van Veen EB, Groenewegen PP, et al. 2015. Will the trilogue on the EU data protection regulation recognise the importance of health research? *Eur J Public Health* 2015; 25: 757–758.

40. Gkoulalas-Divanis A and Loukides G. Introduction to medical data privacy. In: Gkoulalas-Divanis A and Loukides G (eds) *Medical data privacy handbook*. Switzerland: Springer Internatonal Publishing, 2015, pp.1–14.

41. Bannister F and Connolly R. The trouble with transparency: a critical review of openness in e-government. *Policy Internet* 2011; 3: Article 8: 1–30.

42. Keen J, Calinescu R, Paige R, et al. Big data + politics = open data: the case of health care data in England. *Policy Internet* 2013; 5: 228–243.

43. Walravens N. Mobile business and the smart city: developing a business model framework to include public design parameters for mobile city services. *J Theor Appl Electron Commer Res* 2012; 7: 121–135.

44. Carter P, Laurie GT and Dixon-Woods M. The social licence for research: why care data ran into trouble. *J Med Ethics* 2015; 41: 404–409.

45. Manyika J, Chui M, Brown B, et al. *Big data: the next frontier for innovation, competition and, productivity*. Report, US: McKinsey Global Institute, 2011.

46. OECD. *Data-driven innovation: big data for growth and well-being*. Paris: OECD Publishing, 2015.

47. Stanford Medicine. *Harnessing the power of data in health*. Stanford: Health Trends Report, Stanford Medicine, 2017, med.stanford.edu/healthtrends.

48. Mehta N and Pandit A. Concurrence of big data analytics and healthcare: a systematic review. *Int J Med Inform* 2018; 114: 57–65.

49. Hakulinen T, Arbyn M, Brewster D, et al. Harmonization may be counterproductive – at least for parts of Europe where public health research operates effectively. *Eur J Public Health* 2011; 21: 686–687.

50. Zuiderwijk A, Janssen M and Davis C. Innovation with open data: essential elements of open data ecosystems. *Inf Polity* 2014; 19: 17–33.

51. Washington AL. Government information policy in the Era of big data. *Rev Policy Res* 2014; 31: 319–325.

52. Zotano M and Bersini H. A data-driven approach to assess the potential of smart cities: the case of open data for Brussels capital region. *Energy Procedia* 2017; 111: 750–758.

53. Acuff K. Definition of health care policy, https://healthfully.com/259661-definition-of-health-care-policy.html (2017, accessed 14 August 2019).

54. Bevan G and Hood C. What's measured is what matters: targets and gaming in the English public health care system. *Public Adm* 2006; 84: 517–538.

55. Tritter JQ and McCallum A. The snakes and ladders of user involvement: moving beyond arnstein. *Health Policy* 2006; 76: 165–168.

56. Christiaens W, Kohn L, Léonard C, et al. *Models for citizen and patient involvement in health care policy – part I: exploration of their feasibility and acceptability*. Brussels: Belgian Health Care Knowledge Centre (KCE), 2014. KCE Reports 195C. D/2012/10.273/105 (2nd print).

57. Geissbuhler A, Safran C, Buchan I, et al. Trustworthy reuse of health data: a transnational perspective. *Int J Med Inf* 2013; 82: 1–9.

58. Levitin AV and Redman TC. Data as a resource: properties, implications, and prescriptions. *Sloan Manage Rev* 1998; 40: 89–101.

59. Annemans L. *Health economics for non-economists. Principles, methods and pitfalls of health economic evaluations*. Kalmthout: Pelckmans Pro, 2018.

60. Department of Health. Agency standards for reporting data with small numbers. Report, Washington State, May 2018.

61. Sathianathen NJ, Albersheim-Carter J, Labine L, et al. Potential patient harms from misinterpretation of publically reported surgical outcomes. *BJU Int* 2019; 123: 180–186.

62. Sciensano. Data for better health. A Belgian federal government initiative, https://dataforbetterhealth.be/ (2018, accessed 3 March 2019).

63. Digital Belgium. Digital Belgium, http://digitalbelgium.be/en/ (2019, accessed 9 June 2019).

64. Wilkinson MD, Dumontier M and Aalbersberg IJ. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016; 3: 160018.

65. Go Fair. Go Fair, https://www.go-fair.org/ (2019, accessed 9 June 2019).