

# World Health Organization Member States and Open Health Data: An Observational Study

Charles J Greenberg <sup>(1)</sup>, Sangeeta Narang <sup>(2)</sup>

(1) Master of Education, Wenzhou-Kean University

(2) Master of Philosophy, B.B Dikshit Library, All India Institute of Medical Sciences

**CORRESPONDING AUTHOR:** Charles J Greenberg, University Library Director, Wenzhou-Kean University, 88 Daxue Lu, Ouhai, Wenzhou, Zhejiang Province P. R. CHINA 325060 +86-57755870651 (office) +86-18267770875 (cell), FAX +86-57755870101 Email: [cjgberg@wku.edu.cn](mailto:cjgberg@wku.edu.cn)

**DOI:** 10.2427/11950

Accepted on September 5, 2016

## ABSTRACT

**Background:** Open health data has implications for clinical care, research, public health, and health policy at regional, national and global levels. No published attempts have been made to determine, collectively, whether World Health Organization (WHO) member states and governments have embraced the promise and effort required to officially share open health data. The observational study will provide evidence that WHO member states individually and collectively have adopted open data recommended principles, providing access to open health data.

**Methods:** Using the WHO list of member states (n=194), the researchers identified the presence of open health data or initiatives. With each country, the following types of official government web pages were recorded: a Ministry of Health web page; a conspicuous link on a government web page to open health data; additional government health web sites; national government-sponsored open data repositories; unique attributes of national health data web sites; and adherence to the principles of open government data for health. A supplemental PDF file provides a representation of data used for analysis and observations. Our complete data is available at: <https://goo.gl/Kwj7mb>

**Observations and Discussion:** Open health data is easily discoverable in less than one-third of the WHO member states. 13 nations demonstrate the principle to provide comprehensive open data. Only 16 nations distribute primary, non-aggregated health data. 24% of the WHO observed member states are providing some health data in a non-proprietary formats such as comma-separated values. The sixth, seventh and eighth open government data principles for health, representing universal access, non-proprietary formats, and non-patent protection, are observed in about one-third of the WHO member states. While there are examples of organised national open health data, no more than a one-third minority of the world's nations have portals set up to systematically share open health data. At least 15 WHO member states are observed to not even have a government health ministry representation online.

**Conclusions:** We hope the data collected in our Google Sheet and the discussion provided in this paper will generate international interest and advocacy for open health data.

*Key words:* open data, World Health Organization, data collection, health information systems

## BACKGROUND

Health care data is produced from clinical, observational and experimental research. A variety of clinical data, formerly inaccessible and untraceable, now hold the key to new approaches to diagnosis, treatment, prevention, and health promotion.

Health researchers, practitioners, and consumers have begun to demand that governments develop online resources with these new data sources that are citizen-centric, transparent, and accountable, with greater participation, engagement, and collaboration to improve health care. However, no attempts have been made to determine, collectively, whether WHO member states and governments have embraced the promise and effort required to officially share open health data.

Open government data advocates met in December 2007 at Sebastian, California [1], a moment in time when much existing public data was in aggregate, non-primary formats. Eight guiding principles for optimal open data were established:

1. **Complete:** All public data is made available i.e. data is not subject to valid privacy, security or privilege limitations.
2. **Primary:** Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. **Timely:** Data is made available as quickly as necessary to preserve the value of the data.
4. **Accessible:** Data is available to the widest range of users for the widest range of purposes.
5. **Machine processable:** Data is reasonably structured to allow automated processing.
6. **Non-discriminatory:** Data is available to anyone, with no requirement of registration.
7. **Non-proprietary:** Data is available in a format over which no entity has exclusive control.
8. **License-free:** Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

**SOURCE:** [https://public.resource.org/8\\_principles.html](https://public.resource.org/8_principles.html)

Since 2007, governments around the world began to develop web-based open data platforms, where both raw and aggregated data is being preserved and distributed, drawing more attention and inquiry. International institutions, non-governmental organisations, research funders, agencies, and journal editors are encouraging scientists to place their research data available on open data platforms. The United States National Institutes of Health (NIH) established BD2K [2] in 2012, as part of their data science portfolio, to provide appropriate tools and training to facilitate discovery and support for sharing data online.

Several studies and web sites were reviewed to document the value of open health data. We reviewed the following resources:

John-Arne Røttingen et al. [3] provide a comprehensive description and policy discussion of health and development data at a global level, on the basis of gross domestic product, volume of clinical trials, and investment in neglected diseases.

The Anthony Celi et al. editorial [4] in *American Journal of Respiratory and Critical Care Medicine* suggests that the data driven tools utilised during intensive care admission-to-discharge helps in unravelling complexity and variability of underlying diseases, leading to improved predictive models and health care delivery.

The Open Knowledge Foundation (OKF) [5] hosts a handbook on the legal, social and technical aspects of open data, maintaining its currency in a GitHub shared development project [6].

I.A. Sadkovsy, writing in 2014 [7] on predictive, preventive and personalised medicine (PPPM), suggests that the accumulation of large amounts of health-related data could stimulate the transformation of reactive medicine towards PPPM.

## METHODS

Using the World Health Organization (WHO) list of member states [8] (n=194), an observational inspection of government web sites was conducted by the authors, two experienced health information professionals, to identify and record the presence of open health data or planned initiatives to that effect. The authors worked on their own computers with access to the Internet. Google tools provided efficient ways to collect and store research results. The ability of the Google Chrome™ web browser to translate a number of non-English official government languages assisted observations. A shared Google Sheets™ [9] page allowed both authors to work independently, while being able to observe each other's work in progress. Besides the ability to readily locate a functioning Ministry of Health page for 93% of the WHO member states, the authors also compiled a list of intended national Internet domains and used the advanced Google search [10] feature to limit searches to official domains of each country.

With each country, an attempt was made to locate the following types of repositories on an official government web page:

- A Ministry of Health or equivalent government agency
- A conspicuous link on a government web page to open health data
- Additional national government health web sites, with or without data
- National government-sponsored open data repositories, not necessarily health

- Unique attributes of national health data web sites or conspicuous absences.

It should be noted that in the ungoverned and commerce-driven world of Internet domains, no government domain is reserved exclusively for a nation, so methodical checking, within a search restricted to a country domain, was necessary.

## OBSERVATIONS AND DISCUSSION

A supplemental PDF file for this article provides a representation of data used for analysis and observations. Our Google Sheets data is available for viewing or downloading at:

**FIGURE 1. QR code for data collected, also accessible at <https://goo.gl/Kwj7mb>**



The stated goal of this investigation was not only locate open health data repositories but also scrutinise national health data pages and repositories for the presence of open government data principles and open health data activity. How do WHO member states directly or indirectly comply with the original open data government principles established in 2007?

The authors observe that in less than one-third of the WHO member states, open health data is easily discoverable through links on a website or a separate open data collection. Data collected during this study is summarised in Figure 2, providing an overview of WHO member states that demonstrate observable open data principles.

## DISCUSSION

Government open data repositories are unequally spread across geographic regions, and the subset health data is not available in all these repositories.

Countries with well-developed national health data collection mandates, including the United Kingdom, the United States of America, Sweden, Germany, and Canada, have a variety of open health data portals on disease incidents, directories of health centres, and detailed descriptions of regulatory bodies that collect data.

**FIGURE 2. Summary: Number (#) and Percentage (%) of WHO member states that have adopted Open Data Principles.**

### Summary: Number (#) and Percentage (%) of WHO Member States That Have Adopted Open Data Principles

Open Data Principle	# of Member States with Evidence of Principle	# Without Observable Evidence	% with Observed Evidence
<b>1. Complete:</b> All public data is made available i.e is not subject to valid privacy, security or privilege limitations.	13	181	7%
<b>2. Primary:</b> Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.	16	178	8%
<b>3. Timely:</b> Data is made available as quickly as necessary to preserve the value of the data.	14	180	7%
<b>4. Accessible:</b> Data is available to the widest range of users for the widest range of purposes.	64	130	33%
<b>5. Machine processable:</b> Data is reasonably structured to allow automated processing.	47	147	24%
<b>6. Non-discriminatory:</b> Data is available to anyone, with no requirement of registration.	66	128	34%
<b>7. Non-proprietary:</b> Data is available in a format over which no entity has exclusive control.	68	126	35%
<b>8. License-free:</b> Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.	65	129	34%

European countries such as Estonia, Denmark, Greece, Iceland, and Netherlands have extensive datasets on health and incidence of disease. Although Afghanistan, Bahrain, Brunei, India, Indonesia, Japan, Kuwait, Palestine, and the Philippines have open data repositories, these Middle East and Asian countries do not have well represented data on health. African countries such as Kenya and Nigeria have taken interest and are building data repositories, and in the South Pacific, Australia and New Zealand have well established government open data repositories. It was observed that many open data repositories do not have current or even census data online.

Open data principle 1 (Figure 3) is probably the most difficult to achieve and speaks to a greater good and value that comprehensive data release brings to form an accurate picture that can be verified. Health data certainly carries considerable privacy concerns and an ethical obligation to create de-personalised data.

Which WHO member states were represented in our observation of nations attempting to provide complete health data? Australia, Canada, Estonia, Japan, Lithuania, Republic of Moldova, Singapore, Spain, Sweden, Switzerland, the United Kingdom, and the United States of America.

Some WHO member states offer health data without a clear adoption of open data principles. For instance, Peru does not have open data repository but its government health website provides health statistics. [11]

Open Government Data Principle 2 (Figure 4) speaks to the strength of information evidence in primary unfiltered formats. For open health data, this principle suggests to these authors the power of electronic health record systems to provide primary exportable data that can at the same time protect the privacy of individualised patients. The WHO member states distributing primary unfiltered data

include Australia, the Cook Islands, Estonia, Finland, Germany, Japan, Mexico, Peru, Republic of Moldova, Sierra Leone, Singapore, Spain, Sweden, Switzerland, the United Kingdom, and the United States of America.

A very select group of nations have been able to broadly adopt and widely implement electronic health record systems to leverage open health data in research. In some cases, a smaller national population for some WHO member states provides opportunities to do more with less.

The third principle of open data (Figure 5), to quickly distribute data in order to preserve the value of currency, is a significant concern for questions of health policy and public health.

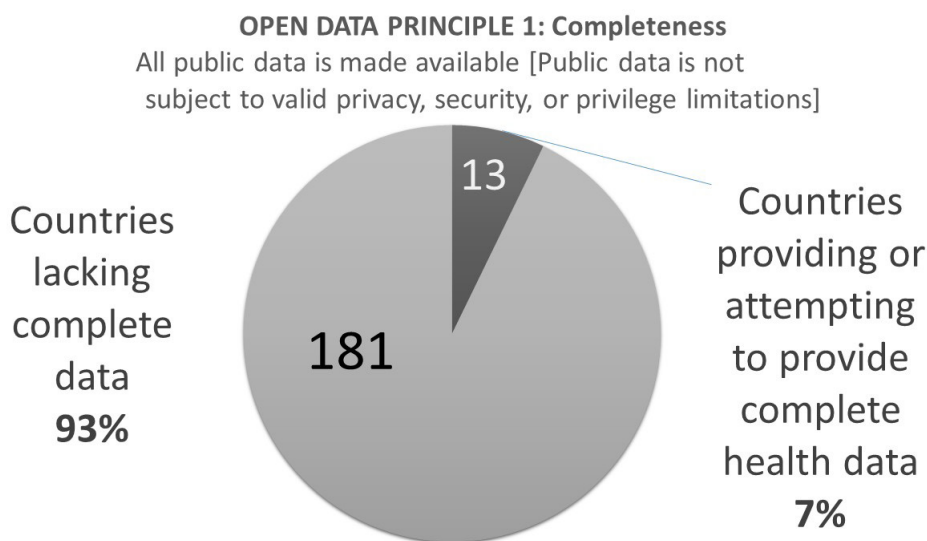
A major shortcoming in the current global provision for open health data we have observed is that only 7% of the nations provide timely, immediate access to collected health data: Australia, Cook Islands, Estonia, Finland, Germany, Japan, Mexico, Republic of Moldova, Singapore, Spain, Sweden, Switzerland, the United Kingdom, and the United States of America. Many factors could be cited to explain a slow or untimely release of data, but the global snapshot is discouraging.

The fourth open government data principle, the accessibility of data (Figure 6), shows a large, significant percentage of positive evidence that WHO member states are attempting to put health data into public view.

32% of WHO member states (n=64) make data widely accessible, but 68% do not make a visible, concerted effort. Many nations have been collecting selected statistics on health incidence and outcomes and releasing the information as a public service. Portable document formats simplify the publishing and online distribution of all types of government reports, including health.

The use of the most popular ubiquitous publishing format, PDF, does not help to reach the fifth principle

**FIGURE 3. Open Data Principle 1: Completeness, as expressed by WHO member states.**



of open government data, the ability to automate the processing and analysis of downloaded data (Figure 7). Nevertheless, there is a strong showing, with 24% of the WHO member states providing data in a format such as CSV or even Excel™.

There are open data website templates that provide immediate data distribution tools.

For example, the nation of Gambia uses a template [12] provided by a third-party organisation that has data download alternatives.

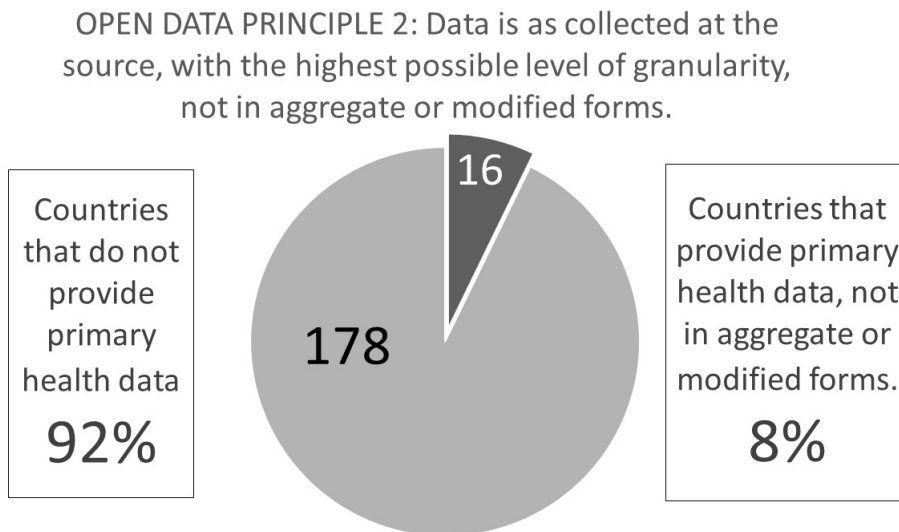
The sixth principle for open government data, universal access without registration or tracking, also has significant representation of agreement for those with open data collection (Figure 8). The large number opposing universal access is not a reflection of those with registration

requirements. Rather, the 66% is almost entirely WHO member states that have no visible open health data, and it is impossible to know how those member states will offer their data to the public, with or without a registration.

The seventh principle of open government data speaks to the elimination of proprietary formats in an open data environment (Figure 9).

The use of proprietary formats (e.g. Cricket Graph) [13] has been in decline as the open data movement has expanded. Just as in the responses to the previous sixth principle, the large number opposing the elimination of proprietary formats is almost entirely WHO member states that have no visible open health data, and it is impossible to know how those WHO member states will offer their data in a proprietary format, though common sense would

**FIGURE 4. Open data principle 2 expressed by WHO member states.**



**FIGURE 5. Open Data Principle 3: timely release of data.**

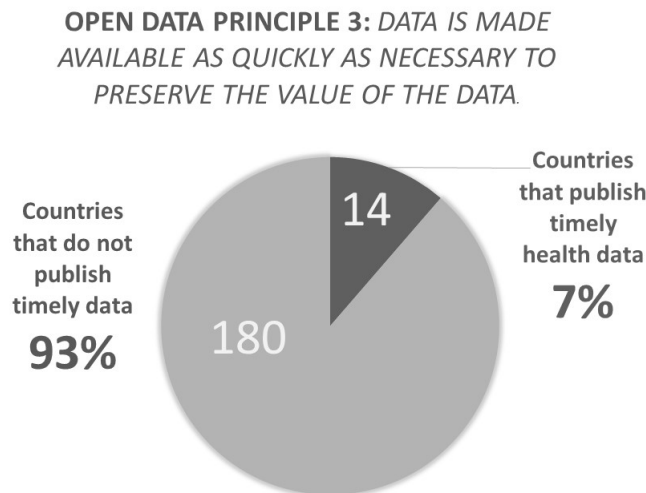


FIGURE 6. Open Data Principle 4: accessible data is available.

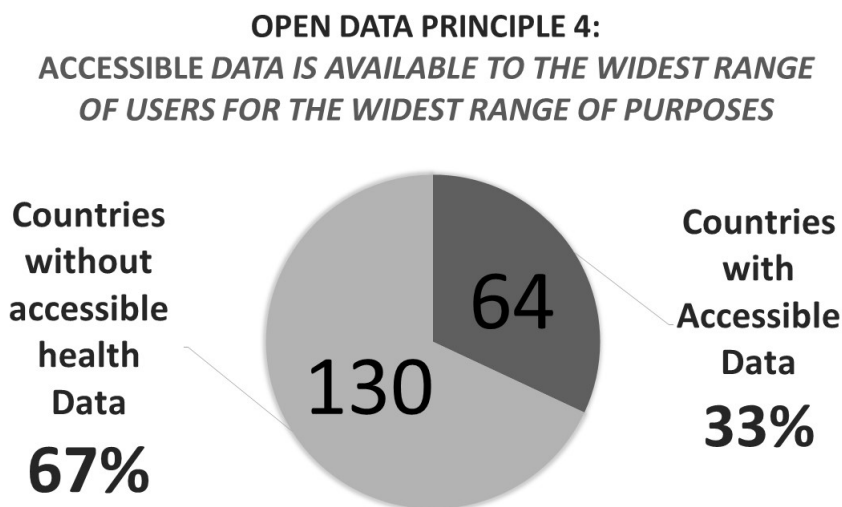


FIGURE 7. Open Data Principle 5, as expressed by WHO member states.

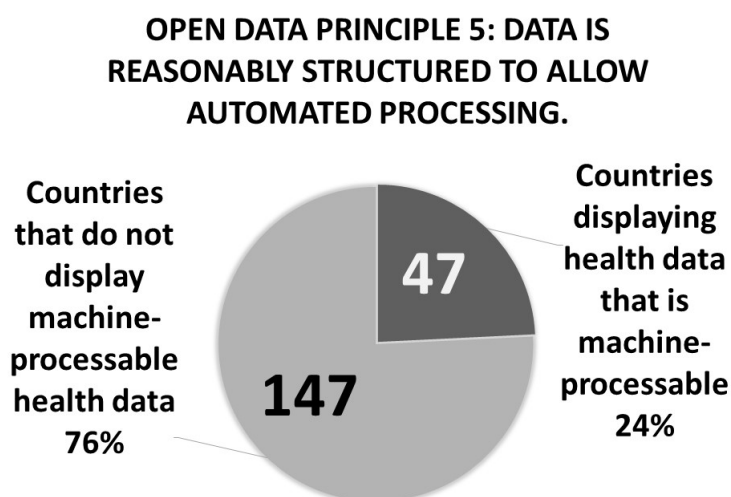
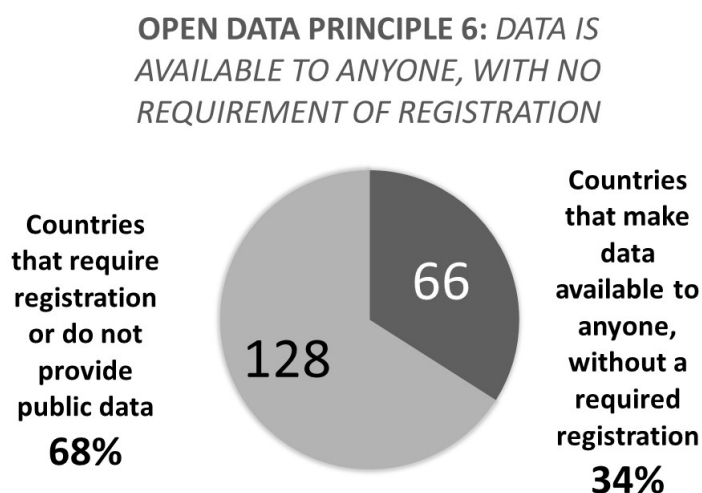
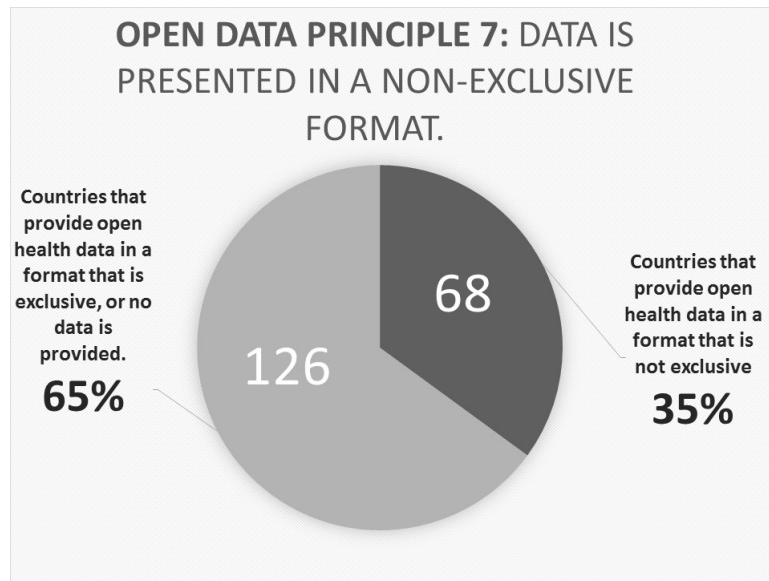


FIGURE 8. Open data principle 6, as expressed by WHO member states.

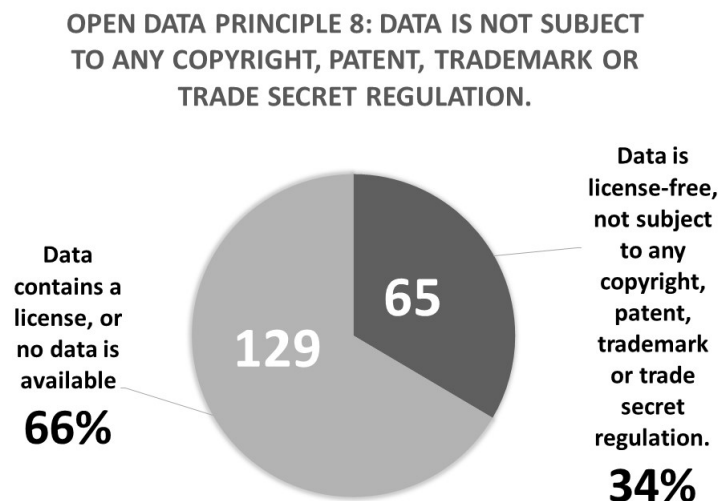




**FIGURE 9. Open Data Principle 7: Non-exclusive data formats.**



**FIGURE 10. Open Data Principle 8.**



suggest that open formats will predominate future WHO member states that decide to share their data openly.

The final open government data principle, addresses how a license, patent, or trademark would be inappropriate to encourage maximum use of open data (Figure 10).

As in the previous two principles, nearly the same proportion of observations are apparent: about one-third of the WHO member states do not mention or suggest any kind of license or patent or trademark. We can make no assertion about the status of the two-thirds majority, the overwhelming majority of whom do not yet have an open health data collection or program. It also seems counterintuitive to think of health data having patent or trademark value.

The opening of health data in public domains has given individuals, organisations, and corporations

the freedom to use, reuse, and redistribute data while adhering to attribution and share-alike Creative Commons [14] provisions adopted by many portals. National governments have improved ministerial performance, enhanced transparency, increased visibility, contributed to confident decision making, greater operational efficiency, cost reduction and reduced risk, greater citizen participation, stronger democracies, fostered innovation, and documented providing better services.

Open data health repositories have enormous potential for improving economies and the welfare of citizens. While studying the repositories it was apparent that some national governments are encouraging and facilitating the maximum use of data by their citizens, developing tools, applications, and services to improve the

welfare of communities. This includes governments in many middle incomes WHO member states around the world that have well-established open data repositories and are slowly compiling datasets on health.

Governments in many developing WHO member states have launched open data initiatives and are attempting to bridge the inequities of income distribution, customising the types of open data on the basis of the needs of their community.

In a systematic observation of the WHO member states and their visible presence online, what can we say to characterise the current state of open health data?

While there were examples of exceptionally well-organised national open health data distributors, there was no more than a one-third minority of the world's nations with principled, organised portals set up to systematically share open data. Attempts to address that gap with shared portals and shared expertise may help those WHO member states that have competing priorities and lack expertise. As the authors have observed, at least 15 WHO member states do not even have a health ministry representation online.

It was also observed that with scarcity of any kind of authoritative health data for some WHO member states, both non-profit and for-profit data entrepreneurs see an opportunity. The Institute for Health Metrics and Evaluation (IHME) [15] has been launched at the University of Washington, with support of the Bill and Melinda Gates Foundation [16]. The IHME is scouring the globe, to identify existing data sources to put together complex pictures of health. A market-based approach is also being undertaken by an online company called Quandl [17]: buying and selling data, including health data, receiving a commission for matching providers of data sets with potential buyers.

While the World Health Organization has been at the forefront of establishing minimum timely health reporting for basic global comparative health statistics, these authors feel the WHO should move even further than the worthy Global Reference List of 100 Core Health Indicators [18] to promote the success stories of open health data national efforts in places like Estonia [19] or the Republic of Moldova [20].

## CONCLUSION

We hope the data collected in our Google Sheet [21] and the discussion provided in this paper will provide contribute to international interest in open health data and more participation from WHO member states.

## Acknowledgement

An oral & slide version of this research project was

presented by the corresponding author at the Medical Library Association annual meeting in Toronto, Canada, 16-May, 2016.

No other publication of the research result has taken place.

## References

1. A Brief History of Open Data and GIS. GovLoop 2014. Available from: <https://www.govloop.com/a-brief-history-of-open-data-and-gis/> [Accessed March 13, 2016].
2. BD2K Home Page | Data Science at NIH. Available from: <http://datascience.nih.gov/bd2k> [Accessed March 6, 2016].
3. Røttingen J-A, Regmi S, Eide M, et al. Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory? *The Lancet* 2013;382(9900):1286–307.
4. Anthony Celi L, Mark RG, Stone DJ, Montgomery RA. "Big Data" in the Intensive Care Unit. Closing the Data Loop. *Am J Respir Crit Care Med* 2013;187(11):1157–60.
5. Open Knowledge: What is Open? Available from: <https://okfn.org/opendata/> [Accessed March 10, 2016].
6. okfn/opendatahandbook. GitHub. Available from: <https://github.com/okfn/opendatahandbook> [Accessed March 10, 2016].
7. Sadkovsky IA, Golubnitschaja O, Mandrik MA, et al. PPPM (Predictive, Preventive and Personalized Medicine) as a New Model of the National and International Healthcare Services and Thus a Promising Strategy to Prevent a Disease: From Basics to Practice. *Int J Clin Med* 2014;5(14):855–70.
8. WHO | Countries. WHO. Available from: <http://www.who.int/countries/en/> [Accessed February 21, 2016].
9. Google Advanced Search. Available from: [https://www.google.com/advanced\\_search](https://www.google.com/advanced_search) [Accessed March 13, 2016].
10. Google Sheets - create and edit spreadsheets online, for free. Available from: <https://www.google.com/sheets/about/> [Accessed February 21, 2016].
11. SITUACIÓN DEL DENGUE EN EL PERÚ (SE 43). Peru Ministerio de Salud. Available from: [http://www.app.minsa.gob.pe/bsc/detalle\\_indbsc.asp?lcind=59&lcobj=4&lcper=1&lcfr eg=12/11/2015](http://www.app.minsa.gob.pe/bsc/detalle_indbsc.asp?lcind=59&lcobj=4&lcper=1&lcfr eg=12/11/2015) [Accessed March 13, 2016].
12. Gambia Data Portal, Dashboards. Available from: <http://gambia.africadata.org/en/Dashboards> [Available February 4, 2016].
13. CricketGraph. In: Wikipedia, the free encyclopedia. Available from: <https://en.wikipedia.org/w/index.php?title=CricketGraph&oldid=700761595> [Available February 4, 2016].
14. Creative Commons. Creative Commons. Available from: <https://creativecommons.org/> [Accessed June 14, 2016].
15. Institute for Health Metrics and Evaluation. Available from: <http://www.healthdata.org/institute-health-metrics-and-evaluation> [Accessed February 4, 2016].
16. Bill & Melinda Gates Foundation. Available from: <http://www.gatesfoundation.org/> [Accessed February 21, 2016].
17. Quandl Financial and Economic Data. Available from: <https://www.quandl.com/>. [Accessed February 4, 2016].
18. WHO | Global Reference List of 100 Core Health Indicators, 2015. WHO. Available from: <http://who.int/healthinfo/>



- indicators/2015/en/ [Accessed February 4, 2016].
19. Estonia: Health statistics and health research database. Available from: <http://pxweb.tai.ee/esf/pxweb2008/dialog/statfile1.asp> [Accessed February 21, 2016].
  20. Moldova: 88 seturi de date gasite. Available from: <http://data.gov.md/ckan/ro/dataset?tags=s%C4%83n%C4%83tate> [Accessed February 21, 2016].
  21. Greenberg CJ, Narang S. Open data big health: Global Open Data for Healthcare Research - Google Sheets. Available from: <https://goo.gl/Kwj7mb> [Accessed February 21, 2016].

