The Disputed Inequality by Open Linked Data: A Position Paper

Kevin Matthe Caramancion
College of Emergency Preparedness, Homeland Security, and Cybersecurity
University at Albany, State University of New York
Albany, New York, United States
e-mail: kcaramancion@albany.edu / kevincaramancion.com

Abstract—**This paper examines Open Government Linked Data (OGLD) and places a position in one of the most controversial disputes in its research—is it subjected to the ways of unsustainable capitalism that it enables inequality among a nation's citizens? The polarization in advocacies for the use of these is solidifying. A history recall is made through an in-depth literature review. Followed by successive overviews of arguments for pro and anti-stances, the author's verdict concludes this position paper.**

Keywords- Open Data; Linked Data; Government Data

I. OVERVIEW OF OPEN DATA

Data access is highly essential for research development. The need for data to be reproduced, verified, and be built from results of existing literatures are all crucial to the development of knowledge for any particular domain [1].

In theory, Open Data as a phenomenon means that governmental data should be available to anyone, especially in its constituents, and that it should be allowed to be redistributed regardless of its form, free of any copyright restrictions [2]. Open Data as a phenomenon is not new as the practice is widely performed not necessarily in the context of governments but mainly in data that applies to public organizations and institutions, however, a formal definition attached to it is what's relatively new - definition from OKD's Manuel explicitly added the "open" as an adjective to describe data that can freely be used, reused, and redistributed without restrictions and the need for referencing and/or copy licensing. Although it may sound liberal, any form of restriction, no matter how small, renders a dataset to be "closed". Highlight to the most widely accepted definition of Open Data is it emphasizes the need to be able to be used by all and accessed by anyone. The utopian hope, the dream is all who access it will eventually share back the body of knowledge to others as well [3].

The objective of the founders of open-sourced data initiatives are similar to those of other "open(-source)" movements including but not limited to open-sourced software, open-hardware(standard malleable), open content, open education, open educational resources, open government, open knowledge, open access, open science, and the open web.

The challenge to this, however, is that the growth of the movement that calls against in these open standards also grows in parallel more so especially in an era of rise in intellectual property rights. Furthermore, data laws and policies are becoming highly pervasive and creators are becoming more conscious of their rights [4]. In modern practice, access to, or re-use of the data is regulated by organizations, public and private alike. The controls placed in it may be through restrictions in accesses in the form of licenses, copyright, patents, and charges for access or re-use.

It is crucial to also note that open data does not necessarily mean textual data. Non-textual data including maps, genomes, connectomes, chemical compounds, mathematical and scientific formulae, medical data, and practice, bioscience and biodiversity. As a matter of fact, the Open Data Forum classifies an item as "open" based not on the content itself, but rather the set of metrics it meets which are usually the characteristics of accessibility, rather than what is in it i.e. the content.

II. OVERVIEW OF LINKED DATA

The hope of its founders is that linked data will be the catalyst for a movement that allows the Internet to become a global database [6]. Linked Data advanced through web development by the application of its principles of sharing data to the web and disseminating the contents so that an even more advanced analysis can be performed on it before being put up on one huge master file. Instead, all and every unit of item in a data or file can now be given an individual identity or URI. The very property of Linked Data that couples with Open Data are the URIs. The following are the effects to Open Data of URIs' application [7]:

a.) the data is contained in a relevant environment, in this context, a unit of item has a web address that can be annotated and referenced hence allowing the whole contents and its implications to be linked back directly to the data.

b.) the data is further enhanced by attaching it to its information model and to related data resulting into the enabling of information to be intertwined across silos resulting into an even better compatibility with third party data sources and higher contextualization; and finally,

c.) by rendering a dataset accessible, even in its most atomic form over the web - applications can choose what to do with these raw intact live data, and in a way the users act as a verifier if it is up to date by reporting it to the moderator if it is not.

The convergence of Linked Data and Open Data is usually described as Linked Open Data (LOD). The term "linked open data" itself has been used since as early as February 2007, when the "Linking Open Data" mailing list was created. This said mailing list was initially hosted by the SIMILE project [8] at the Massachusetts Institute of Technology.

To sum it up, Linked Data breaks down the information silos that exist between various formats and brings down the hindrances between sources usually caused by inconsistent format.

III. THE EMERGENCE OF LINKED OPEN DATA IN GOVERNMENTS

Government data in its raw form alone is powerful, high in magnitude, and extremely relevant to each country's citizens. Since Open Data constructs can be applied to any form of data sets—government data actually is the ideal model for demonstrating the unimaginable benefits of this said "openness". Transforming it to be open fosters transparency, wielding democratic citizens' influence to hold governments accountable. It also enables data improvement through crowdsourcing addendum to the data and facilitation of the provision.

Further applying the constructs of Linked Data renders it to be significantly more useful and powerful—the constraints in resources and non-dynamic form can mean that open data can appear only in limited formats, restricting other technologies for its analysis — such as PDF, Excel, and CSV — but using linked data result in extraction of its maximum value, supporting reuse in unanticipated contexts [9].

A study [10] compared and contrasted a data's value with and without the influence of Open Linked Data. Through the use of RDF that supports several data models he retrieved the data not only as RDF, but also as JavaScript Object Notation or CSV. Highlight to his finding is just because a data is stored as RDF or any specific format, doesn't mean that it

Data-based and informed decision-making and service provisioning can be stemmed from these data and its implications can resonate even to the local government units, communities, the private sector, charities, nongovernmental organizations, and even an individual user. Accountability is unequivocally increased by referencing and keeping track of a government's performance.

Linked data allows easier comparison between two different nations & deep contextualization on the similarities and differences. It is clear from even a cursory examination of government portals that the amount of linked data in the world is widespread and this is due to the proportion of government open data worldwide, paralleled with the increase in technological developments [11].

Clearly, Open Government Linked Data combines both open data and linked data's benefits with less effort from users. The application of linked open data in governments, specifically in the context of the potential for publishing in organizations and the Linked Data Web itself has been positively regarded in its technical and technological dimension [12]. However, other aspects of its use and application including but not limited to the administrative and political issues of it has been raised. This paper discusses these in the next sections and places a position on the subject.

IV. Arguments For

The primal motive to which linked open government applications was created is to seek to empower citizens, aid small to medium (SMEs) businesses, and generally to create value effect for which the most part—in a positive and constructive way. An important note however, opening government data is only just a tool for guidance on the mission of improving education, improving government, and building tools to solve other real world mostly societal problems.

To be particular, several arguments have been made categorically for open data and it is worth noting that these arguments often depend highly on the type of data (since it can also include non-textual data) and its potential applications. The context to which an argument is made should be specifically considered on a specific format for context. Arguments made on behalf of open data includes the following but not limited to:

Data belong to the human race and as such owning the very rights to it should be bound to no one. Typical examples of this are genomes, data on organisms, medical science descriptors, and even environmental data. The Aarhus Convention paved way to an internationally recognized standards for Government Open Linked Data, one worth noting is that public funds should be ought to be used to work on these, including the acquisition and regulation, and so it should be universally available [13]. The logic follows that since it was created by or at a government institution (this commonly applies to developed countries and their government agencies), facts stemming from it cannot be legally copyrighted. The sponsors of research incepting the discoveries of these do not and should not get full value unless the resulting data are freely available. Restrictions, in any form, such as charging for access, on data re- use creates a monopoly to its ownership and thus shouldn't be put in place.

Moreover, government data are essential for the efficient

advocates argued that government funding should not be used to duplicate or create value in data in such a way that it benefits mostly the activities of the private sector e.g. the case of PubChem. It is the responsibility of governments to be accountable for the efficient use of each citizen's money paid through taxes in a way that it should benefit mostly the public rather than any private or commercial organization mostly. Or in cases where it is deemed inevitably that public funds will be used to aggregate the data and should these data bring commercial or private advantageous effects to only a few number of users, the users or sectors who benefitted should reimburse governments for the cost acquired of providing the data.

A study [16] most argumentatively suggested that the revenue earned by publishing these data should be used to cover the costs of generating and/or distribution of the data and in effect this model will result into the dissemination of all data continuing indefinitely. The resulting revenue generated by publishing data can also be used to fund non-profit organizations for their activities and operations for example the learned society publishing supports the society [15].

Concerns of privacy may result into the access to data that is limited only to a specific user group or to restrict it down only to the sub-sets of the data. The process itself of regulating the datasets i.e. collecting, preprocessing, formatting and disseminating data are most of the time, labor and cost- intensive processes and thus a user or organization that renders these services should receive fair remuneration for providing those services. The very same entity who acts as sponsors do not get full value of return unless their data is used appropriately—sometimes this requires quality management, dissemination and efforts of "branding" that can best be achieved by charging fees to the information consumers [17].

In most cases, the end-users who are specifically catered as the target consumers cannot use the data without additional processing i.e. analysis and/or visualization and since anyone has access to the data, a conception that none may have an incentive to invest in the processing required to make data useful and place efforts of improvement for it especially in the cases of biological, medical, and environmental data.

VI. Conclusion and Author's Position

This position paper aims to closely examine the arguments and inferences for and against Open Government Linked Data (OGLD). In essence, the research question that acts as its guide is: Is the Commercialization of Open Government Linked Data justified by the Freedom of Information?

The author, having considered both stances, and in the spirit of an Information & Computing Professional's lens is gearing towards the position of *pro-Open Government Linked Data*. The claims against the OGLD, although soundly reasonable, incrementally disintegrates the very principles and characteristics of Open Data and Linked Data alike. Putting any of the proposed restrictions in any part of the data, its linking mechanisms, and accesses - effectively halts all the benefits that acts as the primal purpose and all the enablers of OGLD. Moreover, the benefits resulted from government funding, should it positively impact a private sector more than the public or government units - should be of no concern to its citizens as it is one of the very hopes to

REFERENCES

[1] House of Commons Science and Technology Committee. (2011) Science and Technology Committee – 8th report. The Publications. Accessed via publications.parliament.uk

[2] Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. Government Information Quarterly. 30 (4): 508–513. doi:10.1016/j.giq.2013.05.012. ISSN 0740-624X

[3] Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *PLoS Biol*, *9*(12), e1001195.

[4] Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

[5] Berners-Lee, T. (2006). Linked data-design issues. *http://www. w3. org/DesignIssues/LinkedData. html*.

[6] Ontotext. (2011). What are Linked Data and Linked Open Data?

[7] Stavrakantonakis, I., Fensel, A., & Fensel, D. (2015, November). Linked open vocabulary recommendation based on ranking and linked open data. In *Joint International Semantic Technology Conference* (pp. 40-55). Springer, Cham.

[8] OpenOrg. (2017). Linked Data Basics for Techies. Open Data Pattern for Organizations. Accessed via http://openorg.ecs.soton.ac.uk/

[9] Heath, T. (2011). Linked data-welcome to the data network. *IEEE internet computing*, *15*(6), 70-73.

[10] Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, *27*(3), 16-24.

[11] Glaser, H., & Halpin, H. (2012). The linked data strategy for global identity. *IEEE Internet Computing*, *16*(2), 68-71.

[12] Hendler, J., Holm, J., Musialek, C., & Thomas, G. (2012). US government linked open data: semantic. data. gov. *IEEE Intelligent Systems*, (3), 25-31.

[13] Manocha, I. (2007). On the Road to Open Data. IDG Connect. Accessed via idgconnect.com

[14] Hilbert, M. (2013). Big data for development: From information-to knowledge societies. *Available at SSRN 2205145*.

[15] Norris, R. P. (2007). How to Make the Dream Come True: The Astronomers' Data Manifesto. *Data Science Journal*, *6*, S116-S124.

[16] Dickinson, B. (2019). Towards a Science Commons includes an overview of the basis of openness in science data. Creative Commons. Accessed via creativecommons.org

[17] Khodiyar, V. (2014). Stopping the rot: ensuring continued access to scientific data, irrespective of age. F1000 Research. F100