Forming a baseline for open health data.

Towards a scorecard based on a preliminary mapping study.

Michael Haider
2017

Abstract

"Data, and in this case, big data will become the third fundamental window into humanity after the microscope and the telescope. It's that important."
- John Nosta (John Nosta, 2016)

In the algorithmic age, the use and availability of big data sets in healthcare has become not only essential but increasingly an imperative. Health data, not only complex but also arbitrary, faces unique challenges like no other domain. Transforming health data into knowledge can range from contributing to medical breakthroughs to taming unsustainable health system costs. This dissertation deals with the availability side of open health data, more particularly how to evaluate open health data. The outcome is a scorecard that, naturally, also acts as a guideline on what to consider when publishing open health data. While much research has been done on open data generally, limited research is done on open health data specifically. This work tries to fill this research gap and aims to establish a baseline for open health data. Given the limited amount of research focusing on the health domain, a preliminary mapping study had to be executed initially and is central to this dissertation. The clusters identified were "Open data applications, platforms, portals and initiatives", "Open data enablers" and "Open data evaluations and guidelines". From the three clusters evidence was collected that could be relevant for the scorecard. Based on these insights the author created the open health data scorecard. This scorecard shall not be understood as a complementary (in the sense of another scorecard evaluation already existing points of analysis (such as data quality)) one to open data generally but one that is tailored to the health domain. Since there is very little research that

concentrates on evaluating open health data, benchmarking the outcome is only possible in a very limited way. Only further research and the research community's discourse can holistically evaluate the outcome. What is more, weighting the individual components of the scorecard (e.g. is data quality or data openness more important) exceeded the scope of this research and could become the focus of further studies. Also, different health regions could be evaluated with the new scorecard in a future publication and be compared to other evaluations on the same regions.

John Nosta (2016) *HIMSS16: Thought leaders preview this year's event*. Available at: http://www.philips.com/a-w/innovationmatters/blog/himss16-thought-leaders-preview-this-years-event.html (Accessed: 24 April 2017).