PatientsLikeMe

Niccolò Tempini
Email: n.tempini@exeter.ac.uk
Department of Sociology, Philosophy and Anthropology & Egenis, Centre for the Study of the Life Sciences, University of Exeter, Byrne House, St German's Rd, Exeter EX4 6TJ, UK

This is the author-accepted manuscript of the chapter published in the Springer Encyclopaedia of Big Data. Please cite as follows:
Tempini, Niccolò. 2017. "PatientsLikeMe." Edited by Laurie A. Schintler and Connie L. McNeely. *Encyclopedia of Big Data*. Springer International Publishing. doi:10.1007/978-3-319-32001-4_162-1.

Introduction

PatientsLikeMe is a for-profit organization based in Cambridge, Massachusetts, managing a social media-based health network that supports patients in activities of health data self-reporting and socialization. As of January 2015, the network counts more than 300,000 members and 2,300+ associated conditions and it is one of the most established networks in the health social media space. The web-based system is designed and managed to encourage and enable patients to share data about their health situation and experience.

Business Model

Differently from most prominent social media sites, the network is not ad-supported. Instead, the business model centres on the sale of anonymized data access and medical research services to commercial organizations (mostly pharmaceutical companies). The organization has been partnering with clients, in order to develop patient communities targeted on a specific disease, or kind of patient experience. In the context of a sponsored project, *PatientsLikeMe* staff develop disease-specific tools required for patient health self-reporting (Patient-reported outcome measures – PROMs) on a web-based platform, then collect and analyse the patient data, and produce research outputs, either commercial research reports or peer-reviewed studies. Research has regarded a wide range of issues, from drug efficacy discovery for neurodegenerative diseases, or symptom distribution across patient populations, to socio-psychological issues like compulsive gambling.

While the network has produced much of its research in occasion of sponsored research projects, this has mostly been discounted from criticism. This because, for its widespread involvement of patients in medical research, *PatientsLikeMe* is often seen as a champion of the so-called participatory turn in medicine, the issue of patient empowerment and more generally of the forces of democratization that several writers argued to be promise of the social web. While sustaining its operations through partnerships with commercial corporations, *PatientsLikeMe* also gathers on the platform a number of patient-activism NGOs. The system provides them customized profiles and communication tools, with which these organizations can try to improve the reach with the patient population of reference, while the network in return gains a prominent position as the centre, or enabler, of health community life.

Patient Members

PatientsLikeMe attracts patient members because the system is designed to allow patients to find others and socialize. This can be particularly useful for patients of rare, chronic, or life-changing diseases: patient experiences for which an individual might feel helpful to learn from the experience of others, whom however might be not easy to find through traditional, “offline” socialization opportunities. The system is also designed to enable self-tracking of a number of health dimensions. The patients record both structured data, about diagnoses, treatments, symptoms, disease-specific patient-reported questionnaires (PROs), or results of specific lab test, and semi-structured or unstructured data, in the form of comments, messages, and forum posts. All of these data are at the disposal of the researchers that have access to the data. A paradigmatic characteristic of *PatientsLikeMe* as social media research network is that the researchers do not learn about the patients in any other way than through the data that the patients share.

Big Data and PatientsLikeMe

As such, it is the approach to data and to research that defines *PatientsLikeMe* as a representative “Big Data” research network – one that, however, does not manage staggeringly huge quantities of data nor employs extremely complex technological solutions for data storage and analysis. *PatientsLikeMe* is a big data enterprise because, first, it approaches medical research through an open (to data sharing by anyone and about user-defined medical entities), distributed (relative to availability of a broadband connection, from anywhere and at anytime), and data-based (data are all that is transacted between the participating parties) research approach. Second, the data used by *PatientsLikeMe* researchers are highly varied (including social data, social media user-generated content, browsing session data, and most importantly structured and unstructured health data) and relatively fast, as they are updated, parsed, and visualized dynamically in real time through the website or other data-management technologies. The research process involves practices of pattern detection, analysis of correlations, and investigation of hypotheses through regression and other statistical techniques.

The vision of scientific discovery that is underlying the *PatientsLikeMe* project is one based on the assumption that given a broad enough base of users and a granular, frequent and longitudinal exercise of data collection, new, small patterns ought to emerge from the data and invite further investigation and explanation. This assumption implies that for medical matters to be discovered further, the development of an open, distributed and data-based socio-technical system that is more sensitive to their forms and differences is a necessary step. But also, the hope is that important lessons can be learned by opening the medical framework to measure and represent a broader collection of entities and events than traditional, profession-bound medical practice accepted. The *PatientsLikeMe* database includes symptoms and medical entities as described in the terms used by the patients themselves. This involves sensitive and innovative processes of translation from the patient language to expert terminology. Questions about the epistemological consequence of the translation of the patient voice (until now a neglected form of medical information) over data fields and categories, and the associated concerns about reliability of patient-generated data, cannot have a simple answer. In any case, from a practice-based point of view these data are nonetheless being mobilized for research through innovative technological solutions for coordinating the patient user-base. The data can then be analysed in multiple ways, all of which include the use of computational resources and databases – given the digital nature of the data.

As ethnographic research of the organization has pointed out (see further readings section, below), social media companies that try to develop knowledge from the aggregation and analysis of the data contributed by their patients are involved in complex efforts to "cultivate" the information lying in the database – as they have to come to grips with the dynamics and trade-offs that are specific to understanding health through social media. Social media organizations try to develop meaningful and actionable information from their database by trying to make data structures more precise in differentiating between phenomena and reporting about them in data records, and make the system easier and flexible in use in order to generate more data. Often these demands work at cross-purposes. The development of social media for producing new knowledge through distributed publics involves the engineering of social environment where sociality and information production are inextricably intertwined. Users need to be steered towards information-productive behaviours as they engage in social interaction of sorts, for information is the worth upon which social media businesses depend. In this respect, it has been argued that *PatientsLikeMe* is representative of the construction of sociality that takes place in all social media sites, where social interaction unfolds along the paths that the technology continuously and dynamically draws based on the data that the users are sharing.

As such, many see *PatientsLikeMe* as incarnating an important dimension of the much-expected revolution of personalized medicine. Improvements in healthcare will not be limited to a capillary application of genetic sequencing and other micro and molecular biology tests that try to open up the workings of individual human physiology at unprecedented scale, instead the information produced by these tests will often be related with the information about the subjective patient experience and expectations that new information technology capabilities are increasingly making possible.

Other issues

Much of the public debate about the *PatientsLikeMe* network involves issues of privacy and confidentiality of the patient users. The network is a “walled garden,” with patient profiles remaining inaccessible to unregistered users by default. However, once logged in, every user can browse all patient profiles and forum conversations. In more than one occasion, unauthorized intruders (including journalists and academics) were detected and found screen-scraping data from the website. Despite the organization employing state-of-the-art techniques to protect patient data from unauthorized exporting, any sensitive data shared on a website remains at a risk, given the widespread belief – and public record on other websites and systems – that skilled intruders could always execute similar exploits unnoticed. Patients can have a lot to be concerned about, especially if they have conditions with a social stigma or if they shared explicit political or personal views in the virtual comfort of a forum room. In this respect, even if the commercial projects that the organization has undertaken with industry partners implied the exchange of user data that had been pseudonymised before being handed over, the limits of user profile anonymization are well known. In the case of profiles of patients living with rare diseases, which are a consistent portion of the users in *PatientsLikeMe*, it can arguably be not too difficult to reidentify individuals, upon determined effort. These issues of privacy and confidentiality remain a highly sensitive topic as society does not dispose of standard and reliable solutions against the various forms that data misuse can take. As both news and scholars have often reported, the malleability of digital data makes it impossible to stop the diffusion of sensitive data once that function creep happens.

Moreover, as it is often discussed in the social media and big data public debate, data networks increasingly put pressure on the notion of informed consent as an ethically sufficient device for conducting research with user and patient data. The need for moral frameworks of operation that overperform over strict compliance with law has often been called for, and recently by the report on data in biomedical research by the Nuffield Council for Bioethics. In the report, *PatientsLikeMe* was held as a paramount example of new kinds of research networks that rely on extensive patient involvement and social (medical) data – these networks are often dubbed as citizen science or participatory research.

On another note, some have argued that *PatientsLikeMe*, as many other prominent social media organizations, has been exploiting the rhetoric of sharing (one's life with a network and its members) to encourage data-productive behaviours. The business model of the network is built around a traditional, proprietary model of data ownership. The network facilitates the data flow inbound and makes it less easy for the data to flow outbound, controlling their commercial application. In this respect, we must notice that the current practice in social media management in general is often characterized by data sharing evangelism by the managing organization, which at the same time requires monopoly of the most important data resources that the network generates. In the general public debate, this kind of social media business model has been linked as a factor contributing to the erosion of user privacy.

On a different level, one can notice how the kind of patient-reported data collection and medical research that the network makes possible to perform is a much cheaper and under many respects more efficient model than what the professional-laden institutions such as the clinical research hospital, with their specific work loci and customs, could put in place. This way of organising the collection of valuable data operates by including large amounts of end users who are not remunerated. Despite this, running and organizing such an enterprise is expensive and labour-intensive and as such, critical analysis of this kind of "crowdsourcing" enterprise needs to look beyond the more superficial issue of the absence of a contract to sanction the exchange of a monetary reward for distributed, small task performances. One connected problem in this respect is that since data express their value only when they are re-situated through use, no data have a distinct, intrinsic value upon generation; not all data generated will ever be equal.

Finally, the affluence of medical data that this network makes available can have important consequences on therapy or lifestyle decisions that a patient might take. Sure, patients can make up their mind and take critical decisions without appropriate consultation at any time, as they have always done. Nonetheless, the sheer amount of information that networks such as *PatientsLikeMe* or search engines such as Google make available at a click's distance is without antecedents and what this implies for healthcare must still be fully understood. Autonomous decisions by the patients do not necessarily happen for the worst. As healthcare often falls short of providing appropriate information and counselling, especially about everything that is not strictly therapeutic, patients can eventually devise improved courses of action, through a consultation of appropriate information-rich web resources. At the same time, risks and harms are not fully appreciated and there is a pressing need to understand more on the consequences of these networks for individual health and the future of healthcare and health research.

There are other issues besides these more evident and established topics of discussion. As it has been pointed out, questions of knowledge translation (from the patient vocabulary to the clinical-professional) remain open, and unclear is also the capacity of these distributed and participative networks to consistently represent and organize the patient populations that they are deemed to serve, as the involvement of patients is however limited and relative to specific tasks, most often of data-productive character..

The afore-mentioned issues are not exhaustive nor exhausted in this essay. They require in-depth treatment; with this introduction the aim has been to give a few coordinates on how to think about the subject.

Further Reading

Angwin, J. (2014). *Dragnet nation: A quest for privacy, security, and freedom in a world of relentless surveillance.* New york: Henry Holt and Company.

Arnott-Smith, C., & Wicks, P. (2008). PatientsLikeMe: Consumer health vocabulary as a folksonomy. *American Medical Informatics Association Annual Symposium Proceedings, 2008*, 682–686.

Kallinikos, J., & Tempini, N. (2014). Patient data as medical facts: Social media practices as a foundation for medical knowledge creation. *Information Systems Research, 25*, 817–833. doi:10.1287/isre.2014.0544.

Lunshof, J. E., Church, G. M., & Prainsack, B. (2014). Raw personal data: Providing access. *Science, 343*, 373–374. doi:10.1126/science.1249382.

Prainsack, B. (2013). Let's get real about virtual: Online health is here to stay. *Genetical Research, 95*, 111–113. doi:10.1017/S001667231300013X.

Richards, M., Anderson, R., Hinde, S., Kaye, J., Lucassen, A., Matthews, P., Parker, M., Shotter, M., Watts, G., Wallace, S., & Wise, J. (2015). *The collection, linking and use of data in biomedical research and health care: Ethical issues*. London: Nuffield Council on Bioethics.

Tempini, N. (2014). *Governing social media: Organising information production and sociality through open, distributed and data-based systems* (Doctoral dissertation). School of Economics and Political Science, London.

Tempini, N. (2015). Governing PatientsLikeMe: Information production and research through an open, distributed and data-based social media network. *The Information Society, 31*, 193–211.

Wicks, P., Vaughan, T. E., Massagli, M. P., & Heywood, J. (2011). Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nature Biotechnology, 29*, 411–414. doi:10.1038/nbt.1837.

Wyatt, S., Harris, A., Adams, S., & Kelly, S. E. (2013). Illness online: Self-reported data and questions of trust in medical and social research. *Theory Culture & Society., 30*, 131–150. doi:10.1177/0263276413485900.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30,* 75-89.