Patient data as medical facts: social media practices as a foundation for medical knowledge creation

Patient Data as Medical Facts: Social Media Practices as a Foundation for Medical Knowledge Creation

Jannis Kallinikos and Niccolò Tempini
London School of Economics

Abstract: This paper investigates a web-based, medical research network that relies on patient self-reporting to collect and analyze data on the health status of patients, mostly suffering from severe conditions. The network organizes patient participation in ways that break with the strong expert culture of medical research. Patient data entry is largely unsupervised. It relies on a data architecture that encodes medical knowledge and medical categories, yet remains open to capturing details of patient life that have as a rule remained outside the purview of medical research. The network thus casts the pursuit of medical knowledge in a web-based context, marked by the pivotal importance of patient experience captured in the form of patient data. The originality of the network owes much to the innovative amalgamation of networking and computational functionalities built into a potent social media platform. The arrangements the network epitomizes could be seen as a harbinger of new models of organizing medical knowledge creation and medical work in the digital age, and a complement or alternative to established models of medical research.

Key words: Medical practice, medical knowledge, social data, social media, computation, patient participation, networking.

Introduction

In a seminal article, Susan Leigh Star showed how we might uncover the means and processes by which a scientific fact 'emerges which is simultaneously stripped of its complexities and isolated from its relationship to a larger work/historical context' (Star 1983: 224-225). In the wake of the so-called Internet revolution, with many organizations experimenting with unconventional approaches to knowledge making beyond the traditional boundaries of research and commercial institutions (e.g. citizen

science, peer-to-peer production, crowdsourcing, social media), social scientists must renew this commitment. There is a need to capture and document what, in such contexts, would otherwise remain invisible or untold, in new web-based approaches to science making.

Some argue that medicine is about to be revolutionized by new technological capabilities that allow new ways of conducting research, and providing therapy and care (e.g. Topol 2012). With this paper, we present a study of an organization that draws on the social networking platform it has developed to pursue medical research that relies on data collected from a distributed, open, user base through patient self-reporting. At one end of this research process, there stand, as a kind of raw material, a myriad of patient observations about their life experiences. The final product at the other end is a number of peer-reviewed articles and other scientific publications. The outcome seems striking. Producing medical knowledge through the routine online involvement of patients provides a stark contrast to the complex, expert-dominated, prestige-laden, and costly institutional arrangements characteristic of medical research. It is thus reasonable to wonder: *How does this process actually happen? How can unconventional, Internet-based organizational forms address traditional expert problems (medical research) through the systematic involvement of non-professionals (patients)?*

At the least, we are aiming to explain the case in such a way that will address the following three interrelated concerns. First, we would like to know the conditions under which "non-experts" are involved in expert work. *How is patient participation organized and governed, to provide information on a reliable and continuous basis such that it can be used as the raw data for medical research?* Second, we search for the technological underpinnings of such an enterprise. *How are data collected and aggregated so as to document and analyze patient experience? How are social media and information*

technologies shaping human communication in this context? Third, we seek to identify the broader implications. *Are traditional medical research practice and institutions going to be transformed by emerging research practices and organizational forms, and if so how?* This paper seeks to address these fundamental issues.

The central node of the network we focus on is *PatientsLikeMe*, a company that runs the key operations underlying the network.[1] Data collection relies on electronic questionnaires and forms that are made available online to network members. As data are collected, they are immediately and automatically aggregated and analyzed, on a continuous basis. In essential respects, the network epitomizes what the current literature (e.g. boyd and Ellison 2008; Gerlitz and Helmond 2013) construes as social media or social networking platforms. Patients are encouraged to enter data about their health status on a regular basis. The data thus made available are used for understanding and describing the patient experience at aggregate levels, with the aim of documenting the effects of medication, illness progression (or remission), and other medical conditions or relations of interest.

PatientsLikeMe uses the data thus collected for research purposes. To date, members of the staff have published 37 outputs – peer-reviewed articles in established journals, reports, editorials, and other formats. From the data collected from patient contributions, the research staff has been able to research a range of subjects. To name a few examples: symptom distribution discoveries (Wicks 2007; Turner et al. 2010); omissions in patient education by medical practitioners (Wicks and Frost 2008); distribution of social issues (compulsive gambling) across patient populations, and association to drugs (Wicks and Macphee 2009); drug efficacy discovery through virtual

[1] *PatientsLikeMe* is a for-profit company based in Cambridge, Massachusetts, USA. It was founded in 2004 and connects more than 250,000 patients (accessed July 28th, 2014). Further information is available at www.patientslikeme.com/about/

clinical trials (Wicks et al. 2011). As a pledge to the patient communities it engages, the organization makes most of the research publicly available (open access). In addition, the web-based system generates a wealth of information based on the patient-reported data and feeds it back to the community in the form of a large range of report pages, each dedicated to specific medical entities represented in the system (conditions, symptoms, treatments). No money is exchanged between the patients and the organization. Patient participation in the network is voluntary, motivated by whatever rewards (cure, socialization, ailment knowledge, recommendations) patients can hope to obtain with respect to the serious conditions they are living with.

The rest of the paper is structured as follows. In the next section, we describe the standard practices of medical data collection and their institutional settings, and contrast them with the data collection arrangements of the network we study here. This is followed by an account of the research strategy, and the data collection and interpretation methods. We subsequently provide a general overview of the network and its features. We then narrow down our focus to the details of one exemplary process of health reporting, that of symptom data collection. We describe how symptom-reporting processes take shape in the organization, paying attention to how technological resources are leveraged to reframe the standard research practices of symptom recognition and recording. Following this empirical case, we discuss our findings in the light of the three fundamental questions we raised above. In doing so, we place our findings within a broader framework that links this case to some fundamental issues of technological and institutional change.

Data and Data Collection Practices in Medical Research

PatientsLikeMe offers research services based on aggregating, packaging, and analyzing patient self-reported data. The organization has been able to use its underpinning

technological infrastructure to construct a unique offer in terms of the scale, longitude, and real-world reference of its data sets. The novelty and uniqueness of the network emerges against the background of the traditional conditions of medical research that we will now briefly characterize.

Medical data management has a long and complex history of non-medical expert involvement. Similarly perhaps to many other fields (e.g. Yates 1989), structured health data collection has, over the last century, become a progressively more complex enterprise that has involved specialists other than doctors or nurses. It has had to take place in specific institutions. Only the realization of hospital services in large scale has enabled the development and systematization of clinical statistics (Shryock 1961). At the same time, stenographers, data editors, and data librarians have all played an increasingly important role in the standardization, circulation, and storage of medical data in hospitals and other medical care settings (Berg 1997; Bowker and Star 1999; Timmermans and Berg 2003). Medical data management specialists have helped systematize the recording, storage, and availability of data produced by medical experts, and have improved the comparability of records across units and contexts, an essential requirement for medical research (Timmermans et al. 1998). Yet, these specialists have mostly not been directly involved with the generation of medical data, which has remained a prerogative of medical experts and, crucially, the ineluctable outcome of expert knowledge application and expert judgment (Dodier 1998; Conrad 2005; Timmermans and Oh 2010).

In the empirical part of this paper, we focus on symptoms data collection as the primary object of analysis, and it is worth briefly referring here to the differences between symptom detection in this network versus that in standard research settings. Traditionally, symptoms have to be discussed, assessed, and filtered through a clinical

interview that takes place where and when the clinicians operate. In most situations, loci are traditional research and health care institutions (research hospitals, laboratories, etc.), and time is limited to the availability of the clinical professionals. Even when data collection concerns physical biomarkers and is automated through machinery, an operator needs to be available to operate the machine at the end of the data collection exercise.

In either a case study or a randomized control trial (RCT),[2] the patient shares and discusses the situation *in situ* with a clinician (nurse or physician). Only through this negotiation can a symptom become a legitimate, recognized fact. A symptom officially enters an information system as data only by the hand of an expert. By controlling data entry, clinicians have the ultimate word on what a symptom really is. The patient plays a dependent role in data collection, and only so far as perceptions and feelings are part of the phenomena under investigation, such as when reporting symptoms. The patient is otherwise excluded from the assessment of all other reportable and observable medical entities (clinical signs) and has no relevant role to play in measurement, nor in inference. The investigation of biomarkers and other observable clinical signs is performed by the clinician and their entourage of tools, the machines of the profession, through the full epistemological authority the clinician commands. This clinical, as we may call it, apparatus (Agamben 2009) largely operates as an engulfing epistemological regime. It defines and interprets the evidence and prescribes the strategy and the

[2] The RCT is upheld by the evidence-based medicine (EBM) movement as the gold standard for medical research. The strengths of this quantitative experimental design for measuring the effects of a treatment derive from a number of features that aim to neutralize possible sources of bias. These broadly include the random assignment of study subjects to different groups, and the designation of each group to either the testing of a treatment or not. By comparing the results of the treated groups against the non-treated (control) groups, a hypothesis can be validated. Moreover, stable processing of experimental protocols can be protected by 'blinding', i.e. not disclosing particular information. Study subjects can be blinded (not told) as to whether they are receiving treatment or not (checking for placebo effects). Similarly, caregivers and researchers can be blinded as to who is administered what. Once blinded, actors are likely to refrain from altering protocols in order to obtain the favored outcome. For further elaboration, see the exhaustive guidelines by the group for the consolidation of trial standards, CONSORT (Schulz et al. 2010; Moher et al. 2010).

objects of the clinical investigation. In the context of limited and fragmented clinical encounters, reductions in scope are necessary so as to obtain consistency and economy of efforts. Thus, the attribution of a local clinical situation to an illness profile, medical category, or classification system is established by the clinician, and their expert knowledge. Against this backdrop, the routine generation of medical data by patients themselves represents an entirely new development that breaks with the history of medical records being used for research purposes and the institutional settings within which these records have commonly been generated.

Patient-Network Data Collection

PatientsLikeMe has developed a social media platform that a patient can join free of subscription fees. As the name suggests, the platform offers the opportunity to socialize with other patients going through similar life experiences. A patient manages a profile provided with common social media features: private messaging, broadcasting, and commenting features in addition to the self-representation tools of a profile picture, username and *'About me*' textbox. In addition, the system provides the patient with a set of health-tracking tools, whereby she can capture several aspects of her own health status. Examples of tracked aspects include the symptoms she is suffering from and their severity, the treatments being taken and related dosages or frequency, weight, labs and tests, and so on, along with many other health-related aspects. Patient members or their caregivers participate in the network voluntarily and generate data that are shared with the network.

The network that *PatientsLikeMe* has built contrasts with canonical models of medical research data collection (see e.g. Berg 1997; Timmermans and Berg 2003; Marks 1997) in a number of ways. *First,* the network breaks away from standard methods of generating medical facts, such as clinical interviews and RCTs, and the institutional

environment of a hospital or other health care unit in which medical facts are commonly embedded. The online platform represents a straightforward arrangement with rather few and simple patient network participation rules. The collected data are all generated through distributed input by the patients, from locations of their choice, and commonly from home. Through these arrangements, the network trespasses on the rigid boundaries separating medical expert practice and research – traditional loci being hospitals, primary care, and laboratories (Shryock 1961; Star 1986) – from the contexts of everyday living in which illnesses commonly manifest and patient experiences are lived.

Second, the network puts patients at the center of the task of data generation. In so doing, it violates or, at any rate, tweaks one of the pervasive customs of medical research, whereby data entry has been the exclusive prerogative of experts (medical doctors and nurses) as the ineluctable outcome of expert judgment. In several instances, the data collection features patient-generated health definitions. Original patient observations are assessed, further pursued, refined, and tested through in-house specialist-patient online interactions, before being incorporated into the system routines for further data aggregation. Still, most of the system routines related to data collection and analysis occur without the routine and direct involvement of clinical professionals. This is a clear departure from traditional medical data management practices in which clinicians are in control of data entry and clinical assessment while patients are relegated to a marginal and dependent position.

Third, data collection in the network is predicated on an inclusive, holistic understanding of health that goes far beyond the medically recognizable conditions of particular diseases. Data collection is, to a degree, use-agnostic, open to the recording of rather broad aspects of patient life. In the hope that all data might turn out relevant, the

network seeks to capture a wide range of circumstances, beyond those that medical researchers would traditionally earmark for data collection in the context of specific research undertakings. As we show in the context of symptoms data collection, patients can choose to track a range of symptoms that is much more granular and extensive than expert terminologies often allow for.

Fourth, data collection is longitudinal, encouraging reporting at all stages of patient life. It is also continuous, seeking to obtain patient inputs as frequently and regularly as possible. The longitudinal and continuous data collection is based on the assumption or belief that it is worth capturing the patient experience in significant detail, transcending the standard focus of most institutional care and research. Patients are free to enter data as often as they believe necessary, as the technology automates many of the transactions involved.

Taken together, these attributes describe a new and different way of organizing data collection for medical research. They lie at the heart of the network and the value it generates for several network stakeholder groups, including the company owners and employees, patients, medical research communities, and pharmaceutical companies. Little wonder that such attributes have been variously anticipated by the contemporary medical research and care practice. Giving patients greater leeway in diagnosis, therapy, and even disease management, observing the progression of diseases and patients over longer time scales, and integrating facts about life and disease have all been developments, in varying degrees, of current practice (Berg 2004; Timmermans and Berg 2003; Clarke et al. 2003; Conrad 2005). Similar views have been characteristic of the wider political discourses in which health care has been embedded over the last few decades (Hasselbladh and Bejerot 2007; Tousijn 2002). In this respect, the network we study both reflects and embodies wider assumptions that are diffused throughout

current practice but also society at large. Yet, through the coordinating framework of social media, these distinctive attributes have been catalyzed in new and interesting ways (Prainsack 2014). The network exemplifies a new architecture for organizing data collection, and new capabilities for analyzing and assembling evidence that require in-depth investigation (Star 1986). As we hope to demonstrate throughout this article, the distinct configuration of the network we study here derives from the flexible forms of interaction enabled by social media and the innovative deployment of the functionalities afforded by current computing and communications technologies (Jonsson et al. 2009).

Research Design and Methodology

A participant observation case study was conducted between September 2011 and April 2012, over 26 weeks, at the headquarters of the organization. One of us participated in work activities, mainly as an R&D team member. He was involved in several projects, while at the same time allowed to exercise great discretion over the time and resources committed to each project. Participation took the form of regular office hours, five days a week, and occasionally entailed acting as a delegate, representing the organization at conferences and in meetings or calls with external guests or partners. The researcher had access to resources that a regular research team member would have.

Such an intensive involvement in the organization allowed the researcher to join forces with most of the employees working at the company's headquarters (around 30-40 members during the period of observation). Beyond the informal observation of work and conversations, the researcher participated in numerous formal meetings – 128 in total – of different kinds, from project-specific task force meetings to stand-up developer meetings, release demo meetings, and company meetings. In addition, he was able to collect data from documents, screen snapshots of user- and admin-facing systems, slide-show presentations, internal e-mail messages and conversations, and the work that the

researcher himself produced for the organization. During his time on site, the researcher logged his observations, in the form of notes typed on a laptop using dedicated note-taking software. This software log was constantly at hand for recording immediate observations and reflections. Even during regular working hours, the researcher was relatively free to detach himself from the regular workflow, to develop notes and reflections that he felt needed prompt recording and elaboration. Additional reflections were logged off site – at evenings and weekends. Tentative interpretations of what he felt were compelling observations and events in need of further explanation were developed *in situ*, crosschecked, and stored (Aaltonen and Tempini 2014; Sayer 2000; Van Maanen 1979, 1993).

Due to the size of the workforce, most of the employees of the company, at all levels, were interviewed, some twice, based on their perceived proximity to the issues under research, and institutional knowledge and memory.[3] Interviews were semi-structured, yet the interview guide was prepared anew for each interviewee to accommodate their role and work, and the progression of the empirical study and collection of facts to that point. Interviews were held throughout the empirical study, but with more than half of them concentrated over the last month. Following the fieldwork, most of these interviews were transcribed and analyzed together with other written and documentary evidence.

With participant observation being the key vehicle of data collection, this should indicate that the empirical investigation featured an exploratory case study research design (Flyvbjerg 2006; Yin 2009). The state of the field on such novel developments did not provide us with firm theoretical propositions with which to link our data collection (Yin 2009). Embedded, observational case studies are an adequate research approach

[3] The researcher held 30 individual interviews, with an average duration of 60 minutes. Snapshots and written notes added up to 665 analytical episodes stored in the electronic log.

for developing new explanations in the absence of a theoretical framework that stipulates the conditions for research (Sayer 2000).

In the context of medical research carried out through social media and patient involvement, our first immediate goal was to assemble empirical observations with the view of addressing the questions we raised in the introduction. The ways these questions were framed (see our introduction above) directed our attention to the means, processes, and techniques by which the company and the network organized, fragmented, and distributed its data collection work, and its data processing and aggregation. Intermediating social interaction through text, measurements, categories, and classifications, the network had to be studied by putting the processes of the construction of health descriptions and symptom detection at the center. The stage of data collection followed by and large what, in current grounded theory jargon, is referred to as theoretical sampling (Corbin and Strauss 2008): the period of participant observation entailed a steady calibration of data collection with emerging interpretations. Our data analysis and interpretation continued after the fieldwork period, mainly through the crosschecking of the empirical material with a view to identifying a consistent narrative about the phenomena under investigation. In this process, we compared our empirical findings on the processes of data collection and analysis used at the field-site to data collection processes depicted in the literature on medical research and medical knowledge creation. Much of that comparison took place against a wider understanding of the role of social media, data, and computation. After several iterative readings and analyses, we selected the most relevant pieces of evidence and assembled them into a case study narrative, following retroductive logic to produce our explanations (Sayer 2000).

In such a unique and innovative case as *PatientsLikeMe*, it was clear from the beginning that many different questions could and should be asked. The network presents itself as a disruptive and unique organization at the crossroads of patient advocacy, evidence-based activism, health care provision, and the pharmaceutical industry. In this purview, it is compelling to prefigure issues of democracy and representation, for instance, against which much of the literature has contrasted similar organizations and initiatives (e.g. Epstein 2008; Rabeharisoa et al. 2013). However, it became clear to us that none of the central issues with which we were concerned could be satisfactorily pursued in the field without first accounting for the premises of systematic patient involvement in medical knowledge creation, and the role technology plays in this process, both as a platform of sociality and as a computational force supporting data collection and, critically, data aggregation and analysis. Both research interests (sociality and computation) shaped our interpretations of the documents we collected, the viewpoints we recorded in the interviews, and the explanations we advance in this paper.

Empirical Findings

Self-tracking can be useful to patients, not only for health monitoring, but also for socialization opportunities (Treem and Leonardi 2012). New lab results, disease courses, or other unfortunate health developments can be important subjects for interaction with other patients. For many patients, *PatientsLikeMe* is primarily a network for support, solidarity, empathy, and companionship. A patient can make use of a number of filters, provided by the system, to browse the network member base and find other patients confronting similar health situations. The efficiency of the system in connecting a patient to other patients with whom they share relevant characteristics (e.g. condition, co-morbidities, treatment regimes) very much depends on the amount of data that the patient inputs into the system. The more data a patient enters about her own situation, the more the system is able to draw connections across the member base.

For *PatientsLikeMe*, health self-tracking represents the possibility to collect valuable views on patients' health status. A host of tracking tools is at the patient's disposal. The patient can enter data autonomously and continuously, generating data over time – traditionally a very expensive and difficult-to-accomplish feat. This can happen whenever the patient finds it most feasible or useful, according to her own routine. System features do encourage data input at regular intervals through user interface (UI) notifications,[4] but the network aims nonetheless at maximizing data collection opportunities. Depending on their condition, patients may lack the time, energy, or even the opportunity to enter data at consistent intervals and volumes. The system therefore allows data inputting at irregular intervals, privileging input volume over timeliness.

Patients can explore information about their own health through various forms of data output. Through data aggregation techniques, the system dynamically constructs and displays profile pages on specific kinds of medical entities (conditions, symptoms, treatments, labs, and others), represented in the form of scores, descriptive statistics, and visualizations. Patients can thus browse a range of reports that put their profile data in perspective and offer a complex picture of the individual. Patients can also browse data representing the health aspects of entire patient populations. Patients access a wealth of information that the system generates by aggregating the data contributed by patient members across the network. Browsing a complex and dynamic network of links, a patient can quickly navigate from her own individual profile to population-level 'symptom (or condition, or treatment) report' pages. A 'symptom report page' shows, for instance, descriptive statistics such as the distribution of symptom severities (number of patients reporting severe, moderate, mild or no effect), the demographics of

[4] Only under particular conditions, such as for instance when a patient has not updated her symptom data for more than 30 days, does the system activate constraints on the UI that limit functionality. In the context of symptom data collection the system will require patients to update their symptom data before performing other operations like tracking new symptoms.

the affected population, and a list of the most popular treatments that patients associate with the symptom. It also shows links to the profiles of other patients suffering from that specific symptom. Because of this webpage structure, the platform provides the patient with information that can help her to understand her health situation, and links promoting and aiding social interactions with others who are similar.

The system generates these up-to-date statistics dynamically.[5] The patient can thus explore parts of the organization's database, in 'sliced and diced' form, by navigating a web of interlinked pages. Patients should then be able to access information that could help her to make sense of specific health situations. A patient can add specific items (e.g. a symptom) that she wants to track on her profile by following links in the item's report page. In so doing, the patient tailors the system to track all the aspects that she deems relevant to her life experience.

Enabling patients to track all aspects that they judge relevant is a strategic goal for *PatientsLikeMe*. The potential for clinical discovery – for collecting the *'gems out there'*, as one top executive defined the rare or insightful correlations or events the company hopes to discover – makes the case for this ambitious distributed data collection architecture. An underlying assumption is the idea that, in respect to a given medical issue, there can be revelatory cases out in the world, and these cases can be documented if the appropriate communications infrastructure is developed. In order for these cases to be discovered, however, the system will be more effective if its data collection remains open and, at the same time, sensitive to a wide range of phenomena, avoiding the over-fitting of events into pre-existing categories. This goal of distributed research

[5] When patients navigate through pages such as symptom report pages, report data are aggregated 'live' through database queries triggered by the execution of the web-application code. The aggregated data are then stored in the cache for as long as they are still up to date, to improve performance (i.e. lower page load time). Depending on the kind of data, they can be cached for between three hours and a month.

data collection adds both promise and burden to organizing collection by means of systematic patient input.

The data collection architecture has therefore been developed with the goal of detecting and mapping clinical diversity in patient experience. Important clinical events could easily escape being recorded. Often, they do not manifest evenly in a patient's life. Also, they may be thought of as singular or irrelevant, and forgotten by the time of the next data collection opportunity. Even small symptom signs, which may seem *prima facie* irrelevant, may have significance and eventually amount to a premonition of future developments (Tempini forthcoming). The open and distributed data collection architecture makes it possible for phenomena to be documented that usually escape recording in traditional settings because they may seem irrelevant, ephemeral or are not easily mapped onto medical experts' categories and classifications. An open architecture can also empower patients. One top executive commented that the system has a fundamental capability to record the patient's voice, with its concerns and insights, in the form of data entries: *'That data is a rich field of information to look at and understand patient concerns'.* The data that patients input into the system can represent needs and concerns that, in the past, were left unvoiced: *'Some of the stuff is not necessarily categorized today in medicine'*. As we show in detail in the next section, the system architecture allows patients to create new symptom categories and to aggregate data inputs into new categories, affording a bottom-up development of a medical categorization system.

Obviously, the potential for open and sensitive data collection is difficult to realize. As the aforementioned top executive said in an interview, '*In the long tail of our data there's probably three things. There's probably patient error, fraud (although I don't think we have a lot of that) or really interesting stuff'.* Successful data collection requires not only

that the system adapts to the life contexts of patients, but also that the researchers devise strategies for reducing biases, errors, and conflicting interests. This is the concern of the complex processes of category review and validation we discuss in the next section.

Symptom Data Input

Each kind of medical entity (symptoms, conditions, treatments, etc.) is described by some defining characteristics. These inform the development of the data collection system. Symptoms are ontologically simpler than other medical entities and for that reason suit the purpose of the present paper. Indeed, conditions have more cumbersome and ambiguous ontological histories (Bowker and Star 1999), while treatments require more complex data models specifying many parameters (dosage, form, frequency, etc). In the context of symptom tracking, the *PatientsLikeMe* system allows the patient to input severity ratings (none, mild, moderate, severe) and add treatments with which the patient associates the symptom. Figure 1 depicts the possible associations that patients can draw between a symptom and a treatment.

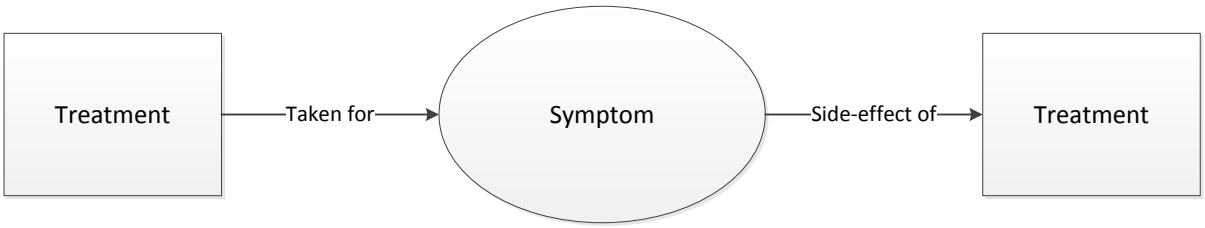

Figure 1: Two possible ways of drawing a symptom-treatment association

A symptom can be associated with a patient profile in three ways. In the first two ways, the system automatically assigns symptoms to a patient profile. First, there are general symptoms. These are symptoms that are expected to cut across the spectrum of all

patient experiences and are assigned to patients of all conditions: *anxious mood, depressed mood, fatigue, insomnia,* or *pain*. The patient is encouraged to track these generic symptoms, because they constitute a common denominator of basic patient life experience. Second, another set of symptoms are automatically added by the system to a patient's tracked symptoms. These are condition-specific symptoms and depend on the conditions that a patient adds to her profile.

Conditions represented in the system are administered through configuration files. The configuration file holds the 'genetic code' of a condition: it stores a number of relevant pieces of information that trigger a number of links or features across the system. Among other things, the staff can store in the configuration file a list of condition-specific symptoms – these are symptoms deemed to characterize the common experience of patients suffering from that condition. As an informant explained, in this way the system is able to automatically adapt and *'serve up'* symptoms to patients: *'The only way we have to serve symptoms up for patients in relationship to a condition is to identify them on the admin tool as the primary symptom'.*

When a patient then adds a condition to her profile, the system assigns the set of condition-specific symptoms to the symptoms to be tracked. If a patient adds a condition that does not have condition-specific symptoms stored in the configuration file, the system refrains from assigning additional symptoms to the patient profile. The identification of the symptoms that are specific to a condition is a labor-intensive task requiring a considerable amount of research. Only a fraction of the conditions stored in the system have so far been assigned condition-specific symptoms. This usually occurs through funded projects that allow the staff to undertake the required research. The list of condition-specific symptoms is compiled from various sources that describe the common experience of a specific condition (more on this later). As a member of the

integrity team explained, *'we are trying to pull those symptoms from an architecture of reference in science; it's sort of saying "ok, what are the ones* [symptoms] *that most commonly people might have experienced".'*

The third way in which a symptom can be associated to a patient profile is by the patient herself, adding symptoms to her profile through a link in the symptom report page (the page dedicated to the dynamic description of a symptom). Symptom report pages can be found through a search feature, by which the patient can find out whether the symptom is already present in the database and, by accessing its report page, see how other patients experience it. By adding the symptom to the profile of the patient, the system enables the patient's experience to be linked to an already existing symptom category. In this way, it is possible for the data collection to aggregate data consistently. The experiences of different patients are thus made similar and comparable through the mediation of the system. The structured nature of the data – with labels and other data fields – makes it possible to aggregate and compare the data that one patient enters with that entered by other patients in the network.

The system uses a number of techniques to help the patient match their symptom to one recorded in the database. As the patient searches for a symptom in the search box, typing the search query letter by letter, a drop-down list starts to show dynamically parsed, instant results.[6] The tool, powered by spelling-correction features, highlights the matching words in the instant results.[7] Clicking one of these results takes the patient to

[6] This feature is similar to the instant results outputted in Facebook's or Google's drop-down search menus.

[7] For instance, if one types the wrongly spelled 'ancious' in the search box, the drop-down menu offers the following results with associated patient populations. The highlighting shows the matching element:

Anxious mood	*251331*
Less anxious	*4*
Stiffness in legs when anxious	*2*
Anxious mood in the morning	*2*

the symptom report page. On that page, she can review the information the system displays about that symptom, consisting of the following elements: first, a symptom description, presented in a verbal, free-text form; second, the distribution of symptom severities on the NMMS scale (none, mild, moderate, severe) across the member population; third, the distribution of treatments associated with the symptom by other patients across the member population; fourth, links to a few profiles of patients experiencing the same symptom and a link to the complete list of all symptom-related patient profiles; fifth, links to a few forum posts related to the symptom and a link to all symptom-related forum posts.

If the patient is not successful in matching her individual case to an existing symptom, the system allows her to initiate the creation of a new symptom. The new symptom first undergoes a review by *PatientsLikeMe* staff. After the review process, a new symptom record is created and fed back into the system. A symptom report page is automatically generated, and other patients will then be able to add this symptom to their tracked symptoms list. In Figure 2, we depict the different mechanisms by which a patient profile's list of tracked symptoms is completed.

---

ancious isn't in our system. Submit a request to add it

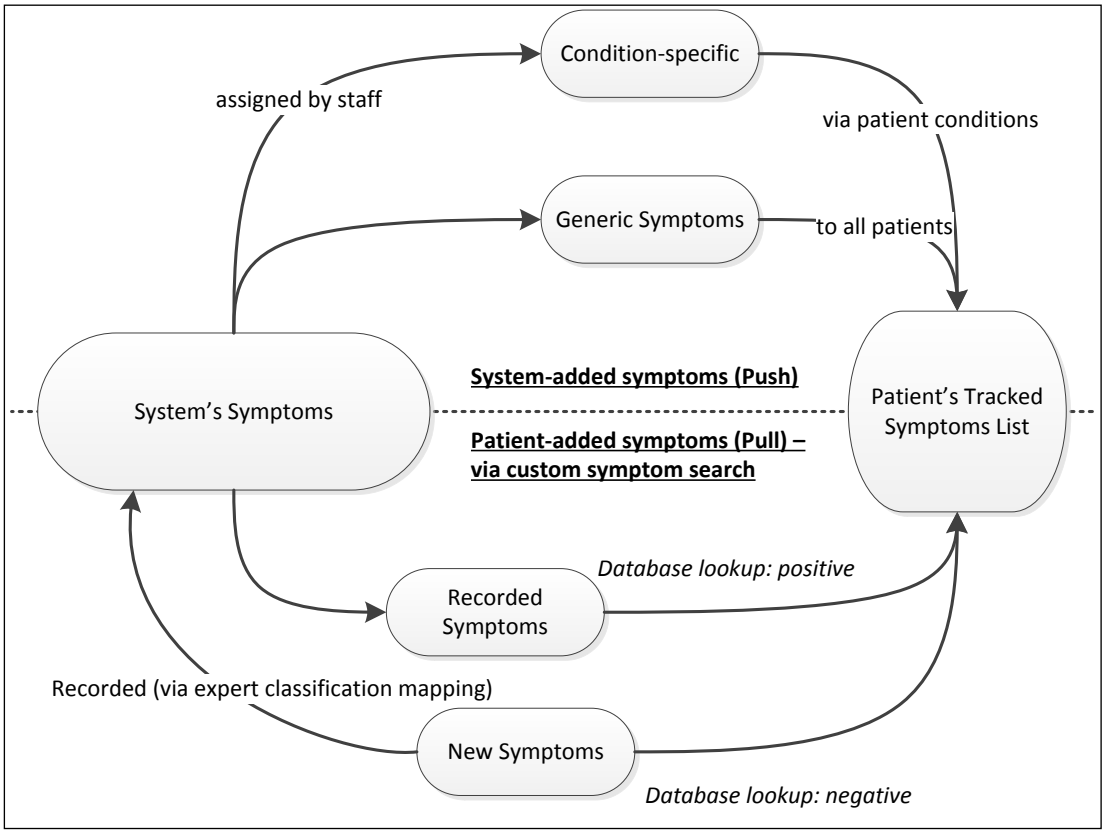

Figure 2: Mechanisms for adding a symptom to a patient profile

As referred to above, if the patient is unable to find a symptom matching her experience, a link in the drop-down menu of the search box allows her to request the creation of a new symptom and to provide its definition: '[symptom] *isn't in our system. Submit a request to add it.'* Upon submission of the request, the patient is informed that the new symptom is pending review from the staff.[8] At the other end of the system, the new symptom shows up in a dashboard used by the *PatientsLikeMe* staff in charge of the ongoing curation of the medical database.

On this dashboard, staff review new symptom requests as well as other new item requests (treatments, conditions, hospitalizations, etc.). The open, distributed, and

[8] The message reads: *'You have successfully added [X]. A patientslikeme staff member will soon review your addition and add it to our global symptom list. You will receive a private message when this process has been completed, and you will be able to add it as a symptom.'*

bottom-up categorization strategy is applied to all kinds of medical entities in the system. When a staff member (most often a registered nurse, pharmacist, or biologist) reviews a case, she can follow a number of alternative courses of action. Often, the definition a patient provides of a symptom is not self-explanatory. Patients might describe something using unclear wording. Sometimes, they may propose as symptoms things or events that are not symptoms. It then becomes necessary for the staff member to reconstruct the context of the patient's request. The staff member can send a private message to the patient asking for clarification or more detail. Through a number of messages, the staff member performs a short, mediated interview seeking relevant evidence so that she can decide how to manage the request. As a nurse and clinical informatics staff member explained,

I have to iterate with them: "[…] *you know, based on what you're showing me and what your picture is, this is what I think it* [the new item] *might be but I could be totally wrong; just let me know".* […] *I'm guessing what's happening based on my nursing background, and helping them to paint a clearer picture for everybody else.* […] *There are certain pieces that they* [the patients] *don't necessarily think are important to add to their profile that are helpful for other people if they know the whole story.*

The staff member also looks up the requesting patient's profile, in order to find clues that could explain what the patient is experiencing and trying to communicate. The staff member embarks on an investigative task, drawing possible connections between the conditions a patient is suffering from, the treatments the patient is taking, the surgeries undertaken, the number, sequence, and dates of diagnoses received, and other relevant information the patient has spontaneously stored in the '*About me*' textbox. Useful context can be provided just by some biographical information – '*there's just something about knowing about their age, about the other conditions they have'* – and health history:

Symptoms may be different because of their [patients'] *condition. So, if someone puts in something vague that might be condition-related and I can check the profile, and I see they've got this condition, it means she is probably talking about the symptom in this context.*

To complement the information about the patient context sourced from conversations and the patient profile, the staff member consults external resources that can range from PubMed and other E-Medicine portals to Wikipedia, UMLS meta-thesaurus and results from Google searches. Through this process, the staff member seeks to progressively define the nature of the item that she is negotiating about with the patient. The ontological status of medical entities themselves – conditions, syndromes, symptoms – is often ambiguous and disputed: *'There is one thing that we are always parsing around here. What's a symptom and what's a condition… Sometimes the patients do not necessarily make the distinction.'* Sometimes the staff cannot clarify the case and it becomes apparent that a prompt decision will not be reached any time soon, such as when a patient simply does not reply to a staff member's questions. In this case, the staff member *archives* the item into a dedicated folder for eventual future follow-up.

When, instead, a decision is reached about a symptom request, the staff member takes one of a number of different actions in the dashboard. One is to *merge* the new symptom into an already existing one. It can happen that a patient fails to notice that the symptom already exists in the database. Mishandling the search feature through a major misspelling error or an incomplete definition may lead a patient to submit a symptom request that can easily be solved. In this case, the staff member merges the new record to the original one: *'I know the context and I have a couple of different pieces of the equation that I might be able to say "yeah, ok, merge".'* The new label created by the patient upon submitting the symptom request is discarded, and the patient's data are

aggregated with the data for the group of patients associated with the existing symptom. Often, merge actions are laborious, and involve the inspection of the patient profile or interaction via messaging features. The staff member continues searching to find out whether the patient experience corresponds to and can be assigned to a specific symptom. For instance, one staff member realised that 'swelling' in fact meant 'injection site swelling' by looking at the patient profile and noticing that the patient's treatment entailed subcutaneous injection:

I could check their profile and I could say 'Oh, that person's on Copaxone'. […] *So you can bring those patients together in those reports; so now these patients are grouped together; it's not just this person has got a side effect of swelling, it's injection site swelling; you get that context from the profile.*

A second course of action the staff member can take is to approve the request and *create* the new symptom. The staff member produces a short description of the symptom, based on the information the patient provided and what it was possible to obtain from other medical sources. The new symptom becomes part of the symptoms database, a symptom report page is automatically generated, and other patients will be able to search for and add the symptom to their own profiles.

Sometimes patients enter multiple symptom entities in the same text string.[9] In this case, the staff member *splits* the symptom into more than one symptom. If necessary, a new symptom is created, but in most cases splitting a new symptom item involves summoning existing symptoms. Through merging and splitting symptom requests, the patient profile can be redirected or subsumed under appropriate categories and thus become an object of aggregation. Tools for merging and splitting symptom requests were not part of the early features for administrating the *PatientsLikeMe* system. They

[9] For instance, one symptom for review could in fact be a string containing two symptoms, such as 'toothache cognitive impairment.'

were developed in order to streamline and automate some standard, common operations that previously depended on patient actions:

Let's say they accidentally entered a treatment as a symptom. There is no way for me to [change it to a] *treatment from* [a symptom] *entry and I didn't want to code it up as a symptom and you can't delete it because it's patient data.* [...] *I would have to message the patient. We then helped build tools like splitting and merging.* [...] *We now have the ability to merge something. If someone puts in 'Fibromyalgia, head pain, headaches', now we can split it into these different categories of already existing databases and make new ones out of it too.*

As a course of action unfolds, the staff member keeps the patient informed and provides an explanation of the action taken. Patients often react if they believe the label they provided still best describes their experience, and it can happen that a staff member will make an incorrect guess. Keeping the patient informed on changes encourages feedback for the actions taken. In the following two diagrams we summarize the interactions we have just described. Figure 3 depicts the operations involved when a patient adds a symptom to her profile that is already present in the system. Figure 4 depicts the operations involved when a patient adds a new symptom to the system. We highlight as 'controlled computation' the steps of the routine that come under expert review. Through the reconstruction of these flows, the organization has engineered a pattern of mediated and linked interactions that utilize advanced data representation techniques to support the patient in the process of data collection. In the cases where automated support breaks down, technology enables, as we have seen, the intervention of a clinical professional and the repairing of the process through several techniques of disambiguation, including patient-staff remote interactions and the use of a range medical resources and data representations. Breakdowns can happen because the

automation is not sufficient to help the patient find the appropriate category, or because the patient experience does not conform to other experiences captured in the database.

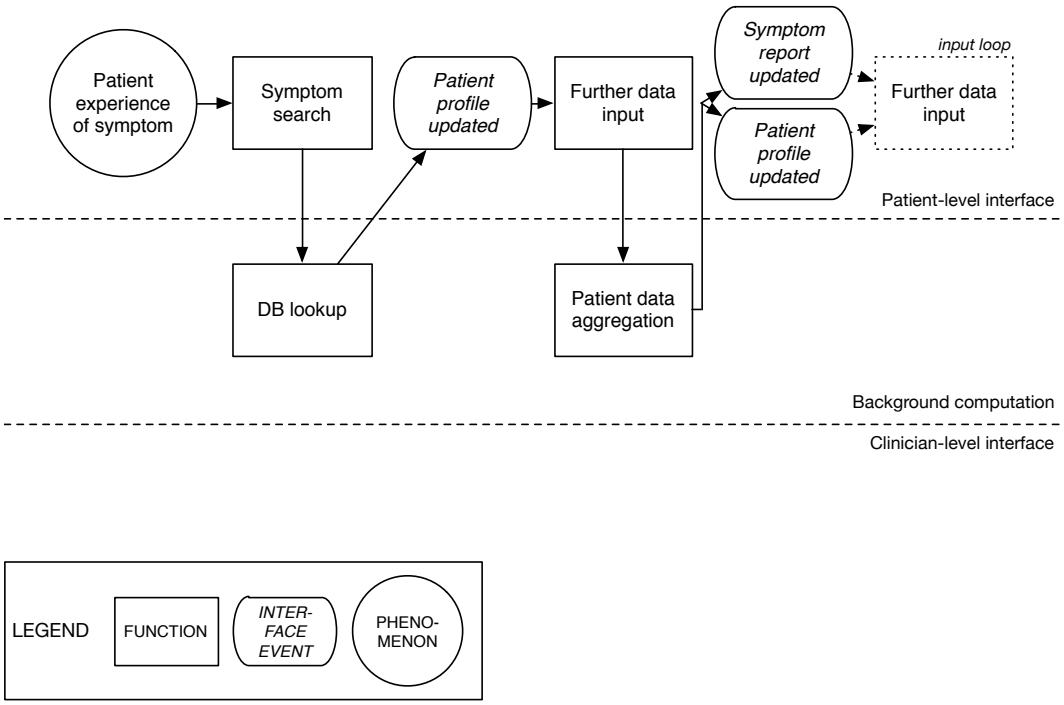

Figure 3: Operation of adding to a profile a symptom already present in the database, divided between patient-level interface, background computation, and clinician-level interface

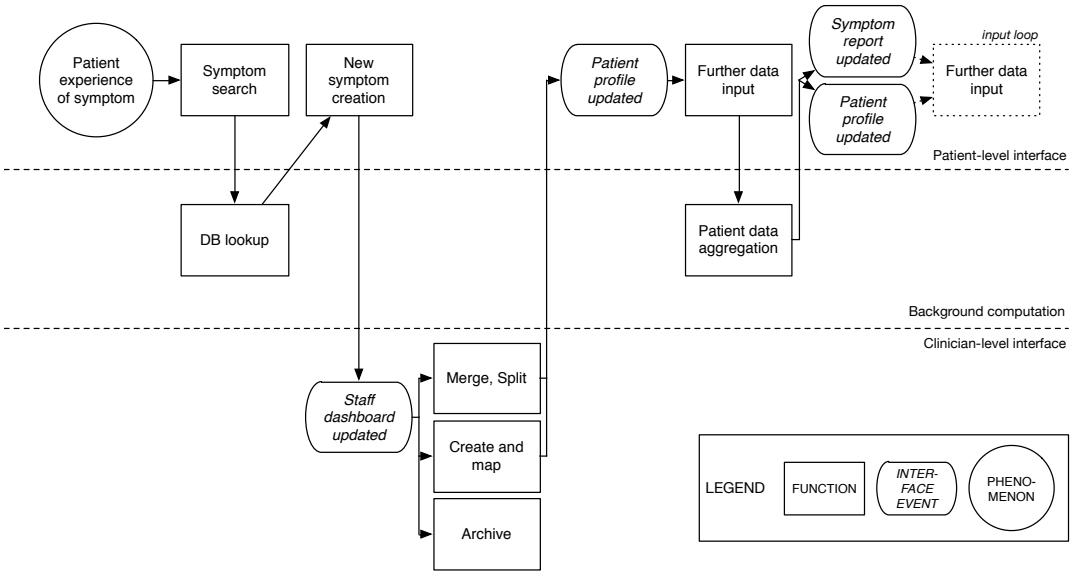

Figure 4: Operation of adding to a profile a symptom not already present in the database, divided between patient-level interface, background computation, and clinician-level interface

To allow new phenomena (patient experience forms) to be detected and emerge through a bottom-up process is a Janus-faced accomplishment. Indeed, to be useful in medical research, new labels need to be made sense of – and meaning arises only if connected to medical knowledge. Therefore, on the one hand, the novel aspect (difference) of a phenomenon needs to be brought to the fore and highlighted, through dedicated definitions and data representations. On the other hand, it is important to place the phenomenon among what is known. New phenomena are only new to a limited extent: at a cost, much can be reduced and subjected to the existing ontology, if need be. Based on their interaction with the patient and apprehension of the illness details that make up the context of the patient's life, the staff member maps the new symptom record to a symptom represented in the expert classification systems (SNOMED, ICD10, ICF and Meddra LLT) through coding. This operation enables the dovetailing of the patient-generated definitions to established expert definitions, that is, of the patient experience language to the clinical professional language.[10] Often patient-generated symptom definitions describe experiences with more nuances or detail than the definitions employed by expert classification systems. For many patients, some nuances are relevant that experts would not recognize as such. Preserving patient definitions means preserving information that can be meaningful not only to patients but also to researchers. As one informant explained, while the system allows the researcher to see the hidden associations between analogous symptoms, it also preserves the patient voice that could be a source of further differentiation: *'You get down to the one that the patient actually told us about in their own words.'* The coding

[10] For instance, the symptom 'anxiety with telephone' is mapped to 'specific (isolated) phobias'.

enables researchers to aggregate different symptoms under a more generic, expert category, and combine the respective data as instances of the same phenomena.

This open, distributed data collection architecture has, over time, come to host about 7,000 patient symptom definitions. Many of these definitions differentiate and specify phenomena along more ordinary medical dimensions; in others, social, personal, and emotional meaning prevail, testing the boundaries of established medical concepts and categories. Collecting and storing perceptions and experiences for the most varied and often multiple reasons, patients overlay the traditional and restrictive condition-treatment-symptom architecture of the patient experience with spurious – but phenomenologically connected – phenomena of everyday living. Coded against the ICD10 code R45.3 'demoralization and apathy'[11] are patient concerns of various kinds: *'loss of ambition, loss of interest, life appeal, not caring further if I die, apathy, environment, no motivation, inability to initiate tests, disorganized…'.* There, in this messy, laborious, and expensive data collection exercise stands the potential that the network is trying to cultivate, for grasping knowledge that lies at the boundaries of social and linguistic conventions, yet is linked to established medical definitions:

Certainly I think it's great that we have a less clinical database in here. […] *Because that's what we are trying to do is use your voice, patient voice, patient-centered data, all these terms we use. It would be kind of hypocritical to create databases that only we decided what would be the entries in them.* […] *you code against happy and unhappiness or social behavior* [issues]. *That is not something that is going to be in any clinical book; it is not going to be ICD. It is not going to be like that but you code it with something similar so that it gets grouped with socialization disturbances or behavioral disturbances and social stuff, and it is all in there.*

[11] http://www.icd10data.com/ICD10CM/Codes/R00-R99/R40-R46/R45-/R45.3

Discussion

The empirical evidence presented in the preceding pages describes the processes and arrangements based on which data on symptoms and patients are collected, ambiguities in the process of symptom mapping are negotiated or settled, and data are made sense of, at both the individual and aggregate levels. In what follows, we draw on this description to address the three fundamental questions we raised at the outset of our investigation, concerning (1) the premises of patient participation in the network, (2) the technological underpinning and organizational arrangements underlying patient data collection, and (3) the putative implications these developments carry for medical practice and institutions.

Network Patient Participation

What seems to strongly differentiate *PatientsLikeMe* and the network it governs from the canonical models of medical research reviewed earlier in this paper are the largely unsupervised data entry by patient populations and the concomitant modest expert contribution that underlies the online process of symptom mapping. The unsupervised data entry by patients establishes the conditions for a diversified information inflow that captures facets of patient life that have hitherto remained beyond the scope of expert medical work and research. It is this objective of capturing the details of patient life and the events that punctuate their everyday *en masse* (to obtain the '*gems out there*') that pervades the network and lies at the heart of the distinctive contribution it is making to medical research (see also Tempini forthcoming).

The objective of capturing the patient everyday in these terms requires the steady and reliable procurement of patient data. Organizing patient participation on this scale is a complex and delicate accomplishment. While massive and largely unguided, the data entry is nonetheless carefully crafted and architected. The mediation of patient life

occurs via an elaborate grid of data fields and categories (e.g. generic and condition-specific symptoms) through which the system and the platform encode existing medical knowledge and other facts of patient life (e.g. biographies, treatments etc).

At the same time, the process of symptom mapping remains open to recording aspects of patient life that do not fit the prescribed categories of medical knowledge. This is accomplished through patient-staff online interaction and a navigational structure through which the process of symptom mapping and creation is organized. Figures 3 and 4 illustrate the pattern of these interactions beyond established medical categories and the series of steps through which patients and staff members negotiate the reality of the patient experience. The objective of reaching beyond the boundaries of established knowledge is also assisted by the links between patients themselves. Through these links, patients can trace aspects of their patient life that might otherwise have escaped their own awareness or observation. The dual accommodation of the requirements of structured data input and the open character of the events that punctuate patient life is the distinguishing mark of the network.

All these vital operations are, in turn, critically dependent on the steady inflow of information, without which the entire system would collapse in one blow (like a spacecraft without fuel). Ensuring a steady level of patient contributions is a delicate task that is sustained, as we show below, by the inventive deployment of the social media platform on which the entire network relies. Web technologies make it technically possible to collect open and longitudinal data but how does this become practically and socially possible? By what means is patient activity in the network sustained? Patients contribute to the network voluntarily and for multiple and often unexpressed reasons, according to their life schedules and priorities, while many of them are dealing with the dramatic implications of their illnesses. Still, *PatientsLikeMe*

depends on patient contributions, as it does not source health data by any other means. With no patients contributing their data over time, the organization would collapse.

Elements supporting the steady inflow of data are, in the first instance, the very social features and interactions that the platform makes available. As indicated, patients enjoy a range of standard social media tools and features that facilitate communication with others in similar situations. However, more than the tools provided to the patients to sustain online conversation, what is critical is the way the platform supports their connection with other patients. In this sense, the platform is an environment that continuously generates possibilities for interaction (connections) and records their outcomes. Patients are linked to specific forum rooms according to the conditions they add to their profiles (they are free to participate in others too). Also, they search for other similar patients through the patient search feature. Patients can filter the user base according to health parameters. The feature is more effective when the patient has entered data about herself, as the system is then able to use those pieces of data to pre-select certain filters. Crucially, for a patient to find someone else, other patients must have entered data about themselves. Conversely, in order to be found by other patients, a patient must have entered data about herself.

Even more powerful than the patient search feature are the links to other patient profiles that pervade the platform on many of its pages, and by means of which data collection is strongly coupled to interaction possibilities. In our description of the symptom report page, we highlighted how the page embeds a host of links that allow a patient to navigate to other patient profiles. These links are as numerous as the number of patients taking a certain treatment, reporting a certain symptom severity, commenting about the symptom in the forums, and so on. The platform, through dynamically constructed pages and database associations, continues to reshape the

linkages between one patient's experience and the experiences of other patients. The range of links that reflect possible connections between patients, worked out on the basis of aggregated data operations, constructs a web of socialization possibilities that become a steady source of patient activity on the platform (see also Tempini forthcoming).

Technological Underpinnings and Organizational Arrangements

In some basic ways, the technological underpinnings of the network coincide with its social media platform, split into patient and clinical interfaces that are supported by a series of background database operations (see Figures 3 and 4). At first glance, one might, perhaps rightly, conclude that patient participation and the linking possibilities it affords depend on a set of straightforward networking options or capabilities typical of web technologies. Patients are put in touch with each other in various ways and explore their links to other patients themselves. The platform intermediates their exchanges.

In fact, much of the social media literature deals with these kinds of social links enabled by social platforms (boyd and Ellison 2007; Gerlitz and Helmond 2013; Morris 2012). Studies have shown how social media enables certain interactions that assist with knowledge production and collaboration within, across, or beyond organizations (Faraj et al. 2011; Majchrzak et al. 2013; Treem and Leonardi 2012). Actors in organizations use social media to reach out to heterogeneous, public tools that afford several kinds of associations. In this regard, social media platforms afford association of 'people to other people, people to content, or content to content', as Treem and Leonardi put it (2012: 162), to support social connections, provide access to information, or enable emergent connections through rankings and recommendations (see also Scott and Orlikowski 2012).

Yet, our reflection on the connections produced by *PatientsLikeM*e that we provide above takes these insightful observations a step further. In our case, patient links are made possible and realized through a series of computational operations, whereby data associations and data manipulation become the principal means for constructing social linkages. The links that are drawn through scores and numbers in the symptom report pages are produced through the filtering, juxtaposing, and aggregating of specific patient data. It is through these data computations – of two or more data tokens belonging to different patient profiles – that a third entity is produced (scores, counts – e.g. Desrosières 1999), whereby associations of one patient with the life paths and experiences of other patients are traced. In this way, back-end data computations and the data architectures on which they rely steadily interfere with front-end interactions, shaping and at the same time being shaped by them. This innovative combination of computational and networking solutions sets *PatientsLikeMe*, and perhaps recent social media platforms more generally,[12] apart from other forms of collaborative networking supported by information and communication technologies (Benkler 2007; Faraj et al. 2011; Majchrzak et al. 2013; Treem and Leonardi 2012).

Some might find this conceptualization unsurprising. At a very basic level, all networking services of information and communication technologies depend on computational operations. Routers and switches coordinating the flux of networking data through algorithmic computation and e-mail clients receiving and sending e-mails are typical examples. Clearly, at this general level, the interpenetration of networking and computation is intrinsic to the current technologies of computing and communication.

[12] To those familiar with social media such as Facebook, our explanation should draw to mind various fundamental features such as the 'Like' action.

However, our claim concerning the mutual implication of networking and computational capabilities is evidently much more specific. The links between patients and the patient activity in the network driven by those links are organized through a wide range of connections between patient data and patient profiles that the system is able to compute by relying on advanced data techniques. Database operations, we claim, lie at the heart of this *computed sociality*, as it were, which is realized by means of advanced representation and aggregation techniques that ceaselessly construct links between network members, here patients (Alaimo 2014; Van Dijck 2013). These observations of ours suggest that social media platforms are not vehicles of unconstrained socialization but complex technological arrangements that recast sociality in a network of social affinities that are shaped by computational operations. As we have shown above and in the empirical findings, patient interactions with one another and with staff are to a significant degree mediated by continuously updated links between network members previously unlinked and unaware of each other. It is this dynamic and constantly updated linking of patients to other patients via the intermediation of scores, counts or categories that shows the complex technological underpinnings of the network and makes it and similar ventures innovative and theoretically interesting (Kallinikos et al. 2013).

Institutional Implications: New Arrangements and Forms of Medical Work

The mutual implication of networking and computational operations generates the need for a specific kind of expert work, performed in the process of symptom mapping. The openness of the data collection to phenomena of various origins means that the system collects information on a much broader range of circumstances than traditional approaches would allow. This includes recognized medical entities such as symptoms, treatments, and conditions, mapped on a continuous and longitudinal basis. It also entails, though, as the symptom creation process demonstrates, data on everyday

experiences and events that evade prescribed categories and, not infrequently, test the boundaries of what is, or may become, relevant and meaningful.

For patients, tracking everyday experiences can represent opportunities to communicate with other, similar patients, along dimensions that they find meaningful or worthy of pursuing – in addition to the possibility of personal health record bookkeeping. What is captured in the system of representations becomes a matter of convergence or divergence between experiences and life histories. By adding a symptom to her profile, the patient establishes the sameness between her experience and those of many other patients. The patient converges towards others via the intermediation of a standardized reference of experiences. Alternatively, by creating a completely new symptom, the patient marks the uniqueness, or difference, of her own patient experience from that of anybody else. Through the creation of a new category, the patient creates an experiential signpost through which other patients might start to connect.

For the organization, data of this sort represent the potential for making clinical discoveries, and identifying and storing meaningful information on medical phenomena and events that could otherwise be difficult to detect. Due to the idiosyncratic, ephemeral, or mundane character of many patient observations, turning these data into something meaningful depends on laborious, expert work. As the symptom disambiguation process described in the preceding pages shows (see Figures 3 and 4), such expert work includes interacting with the patients online, seeking to nail down the precise meaning and reality of patients' observations. In this process, expert medical staff link patient observations to medical categories and definitions whenever possible. When it is not, they establish new medical items, which, once integrated into the

routines of the system, will have their relevance tested by future patient observations and associations.

A few things are worth pointing out in this context. The symptom disambiguation and detection process occurs online without physical contact with the patient. By the same token, the process is mediated by verbal means and other communication cues, at the expense of bodily examination and the focus on bio-chemical markers. These things occur in an environment marked by the absence or, at any rate, the minimal presence of the emblematic figure of clinical research, the doctor. In *PatientsLikeMe,* doctors figure as data collection architects and researchers. They influence data collection through activities such as research projects, participation in the system's long-term strategic planning, and leading frequent, internal, data collection process review meetings. Clinical professionals such as nurses and pharmacists conduct the expert work of data integration that we have depicted. Where technology alone can suffice to provide them support, patients are independent, namely in reporting, selecting, and recording their experiences in standard forms. In exceptional circumstances, the system requires the labor-intensive intervention of clinicians to collaborate and control the completion of the data entry process according to organizational standards. A new kind of division of labor is thus established whereby the tasks underlying medical research are differently distributed across the range of clinical professionals. Also, the alternative architecture through which data are collected transforms the very shape and nature of this work. While it is hard to assess the stability and practical embedding of these changes, the pervasive nature of social media across the social and economic fabric suggests that they may well be part and parcel of wider institutional and organizational changes (Benkler 2007; Faraj et al. 2011; Majchrzak et al. 2013; Treem and Leonardi 2012; Zittrain 2008).

While the process of symptom mapping is often laborious, requiring extensive forays on the part of staff into medical knowledge (e.g. classification systems and definitions), it is essentially aided by computational facilities and advanced database and representation techniques. Exploiting the editable, open, interactive, and distributed nature of digital data (Kallinikos et al. 2013), these computational means and resources enable the expert to draw links between varying phenomena. In many respects, this expert work is data work as Zuboff (1988) depicted it some time ago (see also Kallinikos 1995, 1999). Of course, as our study shows, the social and technological conditions through which data are generated and analyzed have shifted dramatically since the publication of her influential work. However, the nature and implications of the work processes Zuboff associated with work environments infused by a variety of disembodied data tokens, the challenge of what she called 'mastering the electronic text' (Zuboff 1988; ch. 5), persist. In some respects, the changes we have outlined in this paper suggest that the work environments Zuboff perceptively described two and half decades ago have become even more pervasive today (Borgmann 1999, 2010; Kallinikos 2010).

The involvement of broad audiences, enabled by social media platforms and web technologies (Zittrain 2008), is the driving force behind the changes we have sought to depict in this paper. Crucially, the changes we refer to extend beyond industrial or routine work settings, and concern expert work and the processes through which one of the most emblematic of expert pursuits, namely the construction of medical knowledge, is carried out. The punctuation of the patient everyday, the mapping of patient experiences, and the wide reach of phenomena *PatientsLikeMe* is able to access are all made possible through vicarious descriptions, and the medical entities they represent. In this process, social (patient) data become the raw materials transformed into medical facts through the series of operations we have documented in this paper. As shown in the example of symptom data collection, the clinical professional can manipulate links

between entities through data actions such as coding, merging, splitting and so on. Specific, advanced data techniques underlie these operations that would otherwise be so demanding as to render their execution unfeasible. Technology and the data management techniques it embeds underpin the routinization of a range of fundamental expert operations through which patient data are transformed into medical facts. These are no meager changes.

Conclusion and Suggestions for Further Research

In this paper, we have studied the processes through which a social media platform, *PatientsLikeM*e, draws on patient self-reporting to pursue medical research. Using social (patient) data for scientific purposes is, in many respects, an extraordinary accomplishment. The production of medical knowledge has commonly been based on collective processes in which professional skills in data generation, analysis, and validation have figured prominently (Bowker and Star 1999; Timmermans and Berg 2003). In medical research in particular, these processes have taken place in a dense institutional context characterized by established organizational arrangements such as hospitals and health care units and the modes (routines, tasks, standard operating procedures) by which such formal schemes operate. The social media platform we have described in this paper sidesteps these fundamental conditions on which medical research has relied, and provides an alternative path to medical, and more generally expert, knowledge creation.

A network such as *PatientsLikeMe* embodies organizational developments that escape the dichotomies of industrial versus grassroots organizations, and formal versus open, life contexts. It has been pointed out that innovations facilitated by information and communication technology enable 'greater organizational and institutional reach' (Clarke et al. 2003: 162). Also, these innovations power heterogeneous initiatives of

knowledge production on the part of groups such as patient advocacy organizations (Marks 1997; Clarke et al. 2003). Thus, it was correctly foreseen that 'the heterogeneity of knowledge sources can be interpreted as disrupting the division of "expert" versus "lay" knowledge and enabling new social linkages' (Clarke et al. 2003: 177). However, the case of *PatientsLikeMe* attests to the coming together of expert and lay actors through the interconnecting facilities of a new socio-technical system. What this seems to suggest is the advent of the lay actor not as a challenge or substitute to the expert in the production of knowledge, but as a stable collaborator – as an operator upon which expert organizing depends.

At present, it is difficult to assess the stability, promise, and possible drawbacks of the web-based arrangements we have studied here. There is no doubt that the access to the patient everyday that social platforms such as *PatientsLikeMe* facilitate carries significant promise for making use of facets of patient reality and experience that have so far remained beyond the reach of medical practice and research. However, there may too be drawbacks associated with professional turf battles and social conflict (Abbott 2001). It is also difficult to ignore the suspicion that something important may well get lost when medical expertise is cast in the role analyzed in this paper (Dreyfus and Dreyfus 1986; Bowker 2005; Zuboff 1988). These important questions necessitate further research into these alternative modes of pursuing medical knowledge and their implications. In this paper, we have sought to carefully document the *terra incognita* of pursuing medical research via social media platforms and patient self-reporting. While the precise resources and solutions by which such a task will be pursued in the future may vary, the need for documenting patient experience through the means offered by social media platforms and web technologies will persist and possibly grow. The diffusion of these social technologies across the social and economic fabric suggests that they may well be part of wider cultural change in which the boundaries of institutional

and organizational practices and arrangements are refigured (Benkler 2007; Faraj et al. 2011; Majchrzak et al. 2013; Treem and Leonardi 2012; Zittrain 2008).

Extending previous research on social media platforms and drawing on our empirical evidence we have been able to further theorize on the nature of these social technologies. Social media platforms, we have claimed, are not solely places of congregation (socialization) but of aggregation as well. A variety of data is constantly brought into new configurations via aggregation techniques, producing new possibilities for interaction that, in turn, feed back into the data generation process (Tempini forthcoming). Not much is currently known about this computational, as it were, rendition of sociality (Kallinikos 2009) mediated by back-stage operations in social media platforms (Alaimo 2014; Van Dijck 2013). The social relevance and realism of social objects (e.g. averages, aggregates) constructed by statistical operations has been a pervasive theme in contemporary scholarship (Bowker and Star 1999; Desrosières 1999; Hacking 1990, 1999; Porter 1995). It would be interesting to draw on these path-breaking works of literature to reflect on the ontological nature and implications of a sociality that is considerably mediated by computational means and instrumented via social media platforms.

Acknowledgments

This research was partially funded by the Engineering and Physical Sciences Research Council (EPSRC).

References

Aaltonen, A., N. Tempini. 2014. Everything counts in large amounts: a critical realist case study on data-based production. *Journal of Information Technology* **29**(1) 97-110.

Abbott, A. 2001. *Time matters: On Theory and Method,* The University of Chicago Press, Chicago.

Agamben, G. 2009. *What is an Apparatus? And Other Essays*, Stanford University Press, Stanford.

Alaimo, C. 2014. *Computational Consumption: Social media and the construction of digital consumers,* PhD Thesis. Department of Management, London School of Economics.
Benkler, Y. 2007. *The wealth of networks: How social production transforms markets and freedom*, Yale University Press, New Haven.
Berg, M. 1997. Of Forms, Containers, and the Electronic Medical Record: Some Tools for a Sociology of the Formal. *Science, Technology, & Human Values* **22**(4) 403–433.
Berg, M. 2004. *Health Information Management,* Routledge, London.
Borgmann, A. 1999. *Holding On to Reality: The Nature of Information at the Turn of the Millennium*, The University of Chicago Press, Chicago.
Borgmann, A. 2010. Orientation in technological space. *First Monday* **15**(6). http://journals.uic.edu/ojs/index.php/fm/article/viewArticle/3037/2568.
Bowker, G. C., S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*, MIT Press, London.
Bowker, G. C. 2005. *Memory Practices in the Sciences*, MIT Press, Cambridge, MA.
boyd, d. m., N. B. Ellison. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* **13**(1) 210–230.
Clarke, A. E., J. K. Shim, L. Mamo, J. R. Fosket, J. R. Fishman. 2003. Biomedicalization: Technoscientific transformations of health, illness, and US biomedicine. *American Sociological Review* **68** 161–194.
Conrad, P. 2005. The Shifting Engines of Medicalization. *Journal of Health and Social Behaviour* **46**(1) 3–14.
Corbin, J., A. Strauss. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory*, Sage, London.
Desrosières, A. 1998. *The Politics of Large Numbers: A History of Statistical Reasoning*, Harvard University Press, Cambridge, MA.
Dodier, N. 1998. Clinical Practices and Procedures in Occupational Medicine: A Study of the Framing of Individuals. *Differences in Medicine: Unraveling Practices, Techniques, and Bodies,* M. Berg, A. Mol, (eds.), 53–85.
Dreyfus, H. L., S. E. Dreyfus. 1986. *Mind Over Machine*. Free Press, New York.
Epstein, S. 2008. Patient Groups and Health Movements. *The Handbook of Science and Technology Studies,* eds. E. J. Hackett, O. Amsterdamska, M. Kynch, and J. Wajcman, MIT Press, London, 499-539.
Faraj, S., S. L. Jarvenpaa, A. Majchrzak. 2011. Knowledge Collaboration in Online Communities. *Organization Science* **22**(5) 1224–1239.
Flyvbjerg, B. 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry* **12**(2) 219-245.
Gerlitz, C., A. Helmond. 2013. The Like economy: Social buttons and the data-intensive web. *New Media & Society* **15**(8) 1348-1365.
Hacking, I. 1990. *The Taming of Chance*. Cambridge University Press, Cambridge.
Hacking, I. 1999. *The Social Construction of What?* Harvard University Press, Cambridge, MA.
Hasselbladh, H., E. Bejerot. 2007. Webs of Knowledge and Circuits of Communication: Constructing Rationalized Agency in Swedish Health Care. *Organization* **14**(2) 175–200.
Jonsson, K., J. Holmström, K. Lyytinen. 2009. Turn to the material: Remote diagnostics systems and new forms of boundary-spanning. *Information and Organization* **19**(4) 233–252.
Kallinikos, J. 1995. The Architecture of the Invisible: Technology is Representation. *Organization* **2**(1) 117–140.
Kallinikos, J. 1999. Computer-based technology and the constitution of work: a study on the cognitive foundations of work. *Accounting, Management and Information Technologies* **9**(4) 261–291.
Kallinikos, J. 2009. On the Computational Rendition of Reality : Artefacts and Human Agency. *Organization* **16**(2) 182–202.

Kallinikos, J. 2010. Smart Machines. *Encyclopedia of Software Engineering,* Taylor and Francis, London, 1097–1103.
Kallinikos, J., A. V. Aaltonen, A. Marton. 2013. The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly* **37**(2) 357–370.
Majchrzak, A., S. Faraj, G. C. Kane, B. Azad. 2013. The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing. *Journal of Computer-Mediated Communication* **19**(1) 38–55.
Marks, H. M. 1997. *The progress of experiment. Science and therapeutic reform in the United States, 1900-1990*, Cambridge University Press, Cambridge.
Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, D. G. Altman. 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical research ed.)* **340** c869.
Morris, J. W. 2012. Making music behave: Metadata and the digital music commodity. *New Media & Society* **14**(5) 850–866.
Porter, T. M. 1995. *Trust In Numbers : The Pursuit of Objectivity In Science and Public Life*, Princeton University Press, Princeton.
Prainsack, B. 2014. *Beyond Professional Expertise.* Unpublished manuscript: King's College, University of London.
Rabeharisoa, V., T. Moreira, M. Akrich. 2013. Evidence-Based Activism: Patients' Organisations, Users' and Activist's Groups in Knowledge Society. *CSI Working Papers Series*. Working Paper 033. Centre de Sociologie de l'Innovation, Mines ParisTech, Paris, France.
Sayer, A. 2000. *Realism and Social Science*, Sage, London.
Schulz, K. F., D. G. Altman, D. Moher. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical research ed.)* **340** c332.
Scott, S. V., W. J. Orlikowski. 2012. Great Expectations: The Materiality of Commensurability in Social Media. *Materiality and Organizing: Social Interaction in a Technological World,* P. M. Leonardi, B. A. Nardi, J. Kallinikos (eds.), 113–133.
Shryock, R. H. 1961. The History of Quantification in Medical Science. *Isis* **52**(2) 215–237.
Star, S. L. 1983. Simplification in Scientific Work: An Example from Neuroscience Research. *Social Studies of Science* **13**(2) 205–228.
Star, S. L. 1986. Triangulating Clinical and Basic Research: British Localizationists, 1870-1906. *History of Science* **24**(1) 29–48.
Tempini, N. forthcoming. Governing PatientsLikeMe: information production and research through an open, distributed and data-based social media network. *The Information Society.*
Timmermans, S., M. Berg. 2003. *The Gold Standard. The Challenge of Evidence-Based Medicine and Standardization in Health Care*, Temple University Press, Philadelphia.
Timmermans, S., G. Bowker, S. L. Star. 1998. The Architecture of Difference: Visibility, Control, and Comparability in Building a Nursing Interventions Classification. *Differences in Medicine: Unraveling Practices, Techniques, and Bodies,* M. Berg, A. Mol, (eds.), 202–225.
Timmermans, S., H. Oh. 2010. The Continued Social Transformation of the Medical Profession. *Journal of Health and Social Behavior* **51**(S) S94–S106.
Topol, E. 2012. *The Creative Destruction of Medicine*, Basic Books, New York.
Tousijn, W. 2002. Medical dominance in Italy: a partial decline. *Social Science & Medicine* **55**(5) 733–741.
Treem, J. W., P. M. Leonardi. 2012. Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Communication Yearbook* **36** 143–189.

Turner, M. R., P. Wicks, C. A. Brownstein, M. P. Massagli, M. Toronjo, K. Talbot, A. Al-Chalabi. 2011. Concordance between site of onset and limb dominance in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **82**(8) 853–854.
Van Dijck, J. 2013. *The Culture of Connectivity: A Critical History of Social Media*, Oxford University Press, Oxford.
Van Maanen, J. 1979. The Fact of Fiction in Organizational Ethnography. *Administrative Science Quarterly* **24**(4) 539–550.
Van Maanen, J. 1993. *Tales of the Field*, MIT Press, Cambridge, MA.
Wicks, P. 2007. Excessive yawning is common in the bulbar-onset form of ALS. *Acta Psychiatrica Scandinavica* **116**(1) 76–76.
Wicks, P., J. Frost. 2008. ALS patients request more information about cognitive symptoms. *European Journal of Neurology* **15**(5) 497–500.
Wicks, P., G. J. MacPhee. 2009. Pathological gambling amongst Parkinson's disease and ALS patients in an online community (PatientsLikeMe.com). *Movement Disorders* **24**(7) 1085–1088.
Wicks, P., T. E. Vaughan, M. P. Massagli, J. Heywood. 2011. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nature Biotechnology* **29** 411-414.
Yates, J. A. 1989. *Control Through Communication: The Rise of System in American Management*, Johns Hopkins University Press, Baltimore.
Yin, R. K. 2009. *Case Study Research. Design and Methods. Fourth Edition*, Sage, London.
Zittrain, J. L. 2008. *The Future of the Internet and How to Stop It*, Yale University Press, New Haven.
Zuboff, S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*, Basic Books, New York, NY.

Jannis Kallinikos is Professor and head of Information Systems and Innovation Group, Department of Management at the London School of Economics. His research focuses on IS and the governance of organizations. Kallinikos has published widely in IS, Organization Studies and Sociology journals. Recent books include *Governing Through Technology: Information Artefacts and Social Practice*, Palgrave, 2011 and *Materiality and Organizing: Social Interaction in a Technological World*, Oxford University Press, 2012 (co-edited with Paul Leonardi and Bonnie Nardi).

Address: London School of Economics, Department of Management, Information Systems and Innovation Group, Houghton Street, London WC2A 2AE, UK

Email: J.Kallinikos@lse.ac.uk

Niccolò Tempini is a PhD Candidate in Information Systems at the London School of Economics and Political Science, with a background in philosophy. His research focuses on organizations developing open and distributed networks for the generation, collection and analysis of big amounts of data. In his dissertation centred on the PatientsLikeMe case, he focused on how social media technology supports the emergence of new organization forms, epistemic practices and mechanisms of information production and network governance.

Address: London School of Economics, Department of Management, Information Systems and Innovation Group, Houghton Street, London WC2A 2AE, UK

Email: N.Tempini@lse.ac.uk