TILL DATA DO US PART: UNDERSTANDING DATA-BASED VALUE CREATION IN DATA-INTENSIVE INFRASTRUCTURES

Niccolò Tempini

Egenis Centre, Department of Sociology, Philosophy and Anthropology, University of Exeter

Byrne House, St German's Rd, Exeter EX4 6TJ, United Kingdom

n.tempini@exeter.ac.uk

http://www.tempini.info

ABSTRACT

Much of the literature on value creation in social media-based infrastructures has largely neglected the pivotal role of data and their processes. This paper tries to move beyond this limitation and discusses data-based value creation in data-intensive infrastructures, such as social media, by focusing on processes of data generation, use and reuse, and on infrastructure development activities. Building on current debates in value theory, the paper develops a multidimensional value framework to interrogate the data collected in an embedded ethnographical case study of the development of *PatientsLikeMe*, a social media network for patients. It asks when, and where, value is created from the data, and what kinds of value are created from them, as they move through the data infrastructure; and how infrastructure evolution relates to, and shapes, existing data-based value creation practices. The findings show that infrastructure development can have unpredictable consequences for data-based value creation, shaping shared practices in complex ways and through a web of interdependent situations. The paper argues for an understanding of infrastructural innovation that accounts for the situational interdependencies of data use and reuse. Uniquely positioned, the paper demonstrates the importance of research that looks critically into processes of data use in infrastructures to keep abreast of the social consequences of developments in big data and data analytics aimed at exploiting all kinds of digital traces for multiple purposes.

1. INTRODUCTION: VALUE QUESTIONS IN DATA INFRASTRUCTURES

After little more than a decade of wondering at their might, we have come to think of web-based social media networks as quintessential environments of emergent collaboration. Social media have powered innovative social and organizational experiments (Aaltonen & Lanzara, 2015; Benkler, 2007; Howe, 2008; Shirky, 2008, 2010) in which users collaborate with each other or with provider organizations (Boyd & Ellison, 2008; Treem & Leonardi, 2012) and employees collaborate with colleagues or other organizations' employees (Kane, 2015; Leonardi, 2014). The constant updatability of web applications allows social media organizations to let users get to grips with technology, make meaning, discover affordances and invent workarounds that can later be integrated into the formal offering of the social media platform (van Dijck, 2013; Faraj, Jarvenpaa & Majchrzak, 2011; Majchrzak et al., 2013; Gillespie, 2012). Users are able to access information resources generated by the network that otherwise would be very cumbersome, if not impossible, to generate. New forms of community life and sociality can emerge and be supported on an unprecedented scale (Bowker, 2013; Kallinikos & Tempini, 2014). The capacity of collaborative interactions to cut across boundaries has helped elevate these phenomena through the promise they hold for a more connected society and new forms of emancipated sociality (Feenberg, 2010; Feenberg et al., 1996; Josefsson & Hanseth, 2000). Explaining in detail the forces and phenomena that shape this fabric of new infrastructures is still, however, a largely incomplete task (Edwards et al., 2013).

One way in which the literature has been trying to address this gap is through analyses of social media that theorize the creation and exchange of value between the end user and the managing organization. A number of scholars have recently developed arguments based on Marxian understandings of value, asking questions as to whether or not user activity on social media constitutes exploitation and labor captured in the absence of remuneration (Ekbia & Nardi, 2015; Nardi, 2015; Jin & Feenberg, 2015; Rigi & Prey, 2015). Zuboff (2015) has coined the term *surveillance capitalism* to describe the institutional shift brought into being by companies that construct web-based infrastructures to extract value by (silently) witnessing, predicting and generating the unfolding of their users' lives (Cheney-Lippold, 2011; Negri, 1999). Social media are "inclusionary" (Ekbia, 2016:168) towards the surveilled public, and organizations relentlessly tweak and refine the technology to improve their grip on the data and shape user behavior (Tempini, 2015; Aaltonen & Tempini, 2014; Kramer, Guillory & Hancock, 2014; Gerlitz & Helmond, 2013). Organizations are entirely dependent on data to fuel value creation processes and, as such, control of the infrastructure is key (Andrejevic, 2015; Tilson et al., 2010). Social media technology is seen as an "empty container" that depends on user participation to be able to generate value for the managing enterprise (Ekbia, 2016:170; Ekbia & Nardi, 2014:9).

And yet, Ekbia (2016) observes that users are excluded by the economic riches that these infrastructures generate. They trade economic value for emotional satisfaction (Ekbia, 2016). Other *dimensions of value,* besides the economic, underpin their actions (see also Stark, 2000; Heaton & Proulx, 2015). Jin and Feenberg (2015) point out that the value generated in a social media network also relates to a certain potential for users to organize community life. Value is generated by enabling social processes and interactions that would otherwise be impossible. This dimension of value relates to the intermediation and enablement of the public sphere for political and community living (Jin & Feenberg, 2015:53). However, as some have argued, we need empirical confirmation (Ekbia, 2016; Andrejevic, 2015) of the convergences that make or

unmake social media-based communities. We do not fully understand how these spaces are constructed and how they survive over time. This is what this paper addresses. A first step to do this is to try and relate economic value creation to other dimensions of value creation. The existence of value dimensions other than the economic is not *alien* to its system of commensurations; rather, it is the basis for its legitimacy (Stark, 2009; Boltanski & Chiapello, 2005).

Crucially, in this respect, literature has not fully problematized the role of data in the simultaneous creation of multiple dimensions of value. This is peculiar since, ultimately, digital infrastructures are all about data generation, circulation and computation. And, as this paper will demonstrate, once these processes are taken into account, their role in shaping the social convergence that they support is impossible to ignore – as much as the shaping of the infrastructure as a result of data use practices. Much of the literature on value creation in social media has instead only touched on data where economic value extraction is concerned (Culnan, McHugh & Zubillaga, 2010). It assumes that the exploitation of data is mono-dimensional, unilateral and linear, and is owned by the organization developing the infrastructure. The value that users get out of participating in social media seems to result from unrestrained and unstructured social interaction (Jin & Feenberg, 2015) that takes shape despite data-based processes, such as surveillant tracking and profiling (Andrejevic, 2015), and not because of them. However, users are using data directly and indirectly when they engage in social interaction and self-representation over social media, and their use shapes infrastructure and data-based organizational processes (Tempini, 2014).

Helgesson and Krafve (2015) suggest looking at the *modes of data transfer* to unpack the workings of composite networks of data sharing and reuse, and trace the configuration of what data are made available, the participating partners, their relationships and the practical data sharing arrangements to understand what is considered valuable in each setting (Ankeny & Leonelli, 2015). This paper aims to fully take this suggestion in the context of social media networks. It looks at data generation, use and reuse through a longitudinal perspective (Pollock & Williams, 2009) to attend to the creation and making of value *in practice*: "articulations, choices, exchanges, hierarchizations, sortings, displacements, and commensuration of value lend themselves to understanding how values are made" (Dussauge, Helgesson & Lee, 2015).

If we are to understand the social implications of big data and datafication, we need to put data and their processes at the center of our analyses. Social media are just one, perhaps the most prominent, stage where these innovations are going to show their power and be appropriated. This is a great opportunity for the IS field, but one that cannot be fully taken advantage of if the movements, uses and powers of data are not put at the center of the conversation. Researchers of IS (and neighboring fields such as STS) must ask questions that open up the complexity of the relationship between the participating actors, the multiple dimensions of value, and the data and data structure forms through which situations and people are related with one another in data-intensive infrastructure projects such as social media.

2. THEORETICAL FRAMING

Against this backdrop and to illuminate an ethnographic case study of the prominent health research and social media network *PatientsLikeMe* (PLM), I set to ask *"How is a social*

media-based data infrastructure developed to create value in multiple dimensions and for different participating actors, and, conversely, what are the consequences of value creation for infrastructure development?" The aim of this paper is to further open up the relationship between value creation and the development of a social media infrastructure. When value judgments are many and respond to different needs, explaining infrastructure development becomes a complex task worth investigating.

A recently published study has started to make inroads in this same direction (Barrett, Oborn & Orlikowski, 2016) to analyze value creation in health social media. The authors conceptualize online community value as "performed through the ongoing sociomaterial configuring of OC strategies, digital platforms, and forms of stakeholder engagement" (2016:704), and draw from the sociology of worth (e.g. Stark, 2009; Boltanski & Thevenot, 2006) to see valuation processes as shaped by encompassing regimes of valuing practice (i.e. regimes of worth). In the empirical material, the authors show how the social media network they study gradually introduced new value propositions over time to create value of multiple kinds: *financial, epistemic, ethical, service, reputational,* and *platform*. The evolution of the platform is depicted as an additive process by which each value creation dimension was sequentially joined to an increasing portfolio of value dimensions and value propositions ("As new OC value propositions emerged, existing ones continued to be enacted in parallel, becoming more established and routinized over time" (Barrett, Oborn & Orlikowski, 2016:709)). The expansion of the value creation dimension portfolio can continue progressively, if properly managed ("this process of further OC value creation is likely to continue" (Barrett, Oborn & Orlikowski, 2016:717)). Despite the analytical narrative hints at exceptions, the model is one of a rather linear trajectory in terms of the development of value creation capability. The authors argue for the consideration of a much wider and diverse set of stakeholders to be included in the study of value creation in online communities. This enables a more complex understanding of the forces shaping the historical evolution of the network, and links value creation to changing configurations of stakeholder engagement. I take on board these points.

While following a rather similar approach, this present paper's theoretical framing differs mainly in the way it grounds the definition of data-based value creation and value dimensions *in data-based practices*. As Leonelli defines it, to study value in a situation necessitates attending to *"the modes and intensity of the attention and care devoted by given individuals, groups or institutions to given objects or processes, and the motivations underlying such attention and care, which can include a wide range of interrelated concerns"* (Leonelli, 2016:53). This definition helps to attend to value's multiple forms while focusing on actual practices. It can help to sensitize the empirical observer about the complex motivations of data users. This should allow avoiding a sense of unclearness in the way the study cited above defines and deploys a wealth of value concepts (regimes, kinds, propositions, flows) and relates them to one another. A lack of an explicit criterion or definition to underlie the chosen set of value dimensions (i.e. kinds) might make it more difficult to discern their interplay. In the conclusions, Barrett, Oborn and Orlikowski indeed highlight the importance of data for value creation processes: "the analysis and repurposing of data is an important part of how contemporary OC value is created, as evident, for example, in such value propositions as tracking and profiling" (2016:721).

The present paper instead maintains, from the outset, that attention to data is key. It aims to demonstrate the importance of understanding the creation of value in infrastructures as multidimensional and, most importantly, *in association with the use and reuse of data*. This is

necessary to understand how social media-based infrastructures that emerge through fast-paced, boundary-crossing interactions are shaped, and how they shape knowledge about their users. As Rajan and Leonelli (2013) have argued (see also Leonelli, 2014), questions of data circulation and use are imbued with questions of value in ways that must be understood multidimensionally: to a market concept of value, we must contrast social, normative and other concepts of value. All come into play in shaping how data move across different situations and what uses they are brought to bear. Thus, we must pay attention to both 1) what is valued in data by the participating actors and 2) what are the knowledge and worldview upon which actors make valuations of data. A perspective that centers on data practices allows us to better understand what different kinds of actors can get out of their participation in the network. To do so, and following Stark (2009:13), I specify the evaluative principles that apply to the case at hand.

As I will demonstrate, this approach yields granular empirical results. In previous work, I have highlighted interesting, nonlinear dynamics at play when studying the link between data and data structures and user engagement (Tempini, 2015). With a different focus but a similar interest in data practices, in this paper, I highlight a set of nonlinear, vagarious dynamics that make infrastructure development for concerted and multidimensional data-based value creation an inherently challenging proposition. The paper shows how elusive and problematic achieving data-based value creation can be, when the historical evolution of the underlying infrastructure is taken into account.

2.1 SOCIAL MEDIA AS DATA-INTENSIVE INFRASTRUCTURE

To do this, the paper further leverages knowledge from information infrastructure studies to help problematize the evolution of the infrastructure at the center of the case study. It is inherent to the basic understanding of both physical and digital infrastructures to be open to multiple uses (Tilson et al., 2010; Hanseth, Monteiro & Hatling, 1996; Star & Ruhleder, 1996) and, at the same time, to place constraints on user activity. Hanseth and Lyytinen (2010:1) representatively define an information infrastructure as "shared, open, heterogeneous and evolving socio-technical system of information technology capabilities" and stress how information infrastructures are "controlled by emergent, distributed and episodic forms of control", while their "evolutionary dynamics are nonlinear, path dependent and influenced by network effects and unbounded user and designer learning" (see also Star & Ruhleder, 1996). A fine balance between constraint and enablement is what enables infrastructures to have generativity – a capacity to support a multiplicity of previously unrealized, and situated, possibilities of use (Tilson, Lyytinen & Sorensen, 2010; Zittrain, 2008) and activate processes of adoption, innovation and scaling (Henfridsson & Bygstad, 2013).

Achieving and maintaining this balance is a paramount problem for an infrastructure's long-term evolution. Issues of infrastructural change and control are paradoxically involved (Tilson et al., 2010), as they are both necessary elements and sources of uncertainty. Decentralized control and modular (Hanseth & Lyytinen, 2010), loosely coupled architecture (Henfridsson & Bygstad, 2013) are seen as preconditions. Hanseth and Lyytinen (2010) highlight as core design problems: bootstrap (igniting early user adoption) and adaptability (serving increasingly diverse use needs). They liken information infrastructures to complex adaptive systems, which exhibit nonlinear behavior, emergent order, state irreversibility and unpredictable outcomes (Hanseth and Lyytinen, 2010:5). This understanding of infrastructure

and its evolution is an invaluable resource in the case study that this paper presents, as it provides a set of conceptual checkpoints for relating infrastructural configurations to value creation practices.

Yet, this paper is no mere application of information infrastructure wisdom. While the information infrastructure literature has always held, at its core, an understanding of infrastructures as a place where multiple value questions converge, its explicit articulation through a structured, multiple value creation framework has been largely missing. This paper instead explores information infrastructure evolution through such a theoretical structure. At the same time, by mobilizing knowledge from information infrastructure studies, the paper also aims to extend its scope through an explicit interest in data and data use. Despite their obvious centrality to the basic operation of any digital infrastructure, data and data practices have not been looked at in detail. Use (and value) questions are articulated at the more abstract level of infrastructure or IT capability use. However, the complexity of data use practices (computation, aggregation, linkage, analysis, inferencing) converging in projects such as social media is of such importance that a focused perspective is a timely and much-needed contribution.

An immediate consequence of this paper's theoretical stance is to differentiate between information infrastructures in general and infrastructures for continuous, (semi-)automated data use and computation, such as social media. The empirical narrative describes social media organizing centered on an information infrastructure. There are novel features to this arrangement, I argue, that make it useful to think about this kind of social media-based information infrastructure through the concept of *data-intensive infrastructure* (Gerlitz and Helmond, 2013:1349, but not defined), a calculable, interactive social environment that I here define as an *infrastructure that is developed to collect and circulate the same data that it computes to serve and organize its users' contributions and interactions*. Indeed, at the center of PLM's evolution stand the data and their processing through infrastructure.

It is impossible to overestimate the importance of data in a social media network like PLM, which was entirely set up to gather and connect patients and make available patient-reported data for multiple uses and users. Data and data categories are used as *organizational devices* (Tempini, 2014), driving system behavior to guide user navigation, foster user interaction and sociality, and coordinate data collection. Data-intensive infrastructures are characterized by an interactive organizational behavior, which changes based on the same data that the infrastructure collects and circulates. This is a powerful way of understanding something specific about social media *qua* information infrastructures. Data-intensive infrastructures are "data-centric" (Leonelli, 2016) but they are a specific kind of information infrastructure because *part of the data generated by its users are continuously recirculated, after transformative computation, to guide the system's own interactions and to serve information to different users along pre-structured paths*. This is to support, coordinate and organize a diversity of loosely distributed participating actors, whose data contributions are *critical* to the survival of the infrastructure and organization itself and whom are, at the same time, interested in using the data themselves.

3. METHODOLOGY

The case presented here is based on empirical data collected in the context of a long, in-depth observational case study (Yin, 2009). Between September 2011 and April 2012, I visited the headquarters of PLM in Cambridge, Massachusetts, three times, spending a total of 26 full-time weeks at the fieldwork site (Tempini, 2014). Data collection was ethnographic in method, active, interventional and open to both purposefully occasioned and unexpected events that invited reflection. I participated in regular work activities, acting as an R&D team member. I was involved in several projects and had regular team member access to internal resources. I occasionally liaised with external partners or guests, in conferences, meetings or conference calls. My free collaboration (no financial exchange was involved) granted my access to the field site and the permission to conduct the fieldwork.

I accessed the heart of the organization's life, taking part in the design and development of the infrastructure (Edwards et al., 2013) with tangible input. I collaborated with most of the employees based at the headquarters (~35) and participated in a variety of internal meetings (totaling 128) with all teams – from software development to R&D, the Health Data Integrity (HDI) team and the Community team, but also project task-force and company-wide meetings. I collected numerous documents and presentations, analyzed screen snapshots of both user-facing and admin-facing systems, and took notes from internal e-mail messages and conversations I took part in. Internal meetings were a particularly rich source of information, because members talked openly about issues, confronted perspectives, and hinted at past events of relevance. Embedded, first-person participation in meetings and projects allowed me to acquire a rich understanding of current situations and concerns. Then, when I identified interesting cues or themes, I could easily make use of internal resources and of focused interview time to trace back and reconstruct concatenations of events. Interviews were tailored to each interviewee, their role, and institutional knowledge and memory. Most of the employees were interviewed; there were a total of 30 interviews and an average duration of one hour. I collected field notes continuously, which often served both project collaboration and fieldwork purposes, and took note of apparent connections between the observed events and known theoretical issues. Analytical writing is useful for reflecting on observed events and attempting to interpret them (Aaltonen & Tempini, 2014). For initial interpretation and immediate storage and retrieval, I made use of tagging features in my note-taking software to code notes of events and conversations. These first layers of codes proved useful for immediate thematization and interpretation on site (Corbin & Strauss, 2008; Marton, 2013). Emergent codes were a useful prop for building a consistent memory of what I had seen, and sharpened my observation of what came next (Aaltonen & Tempini, 2014; Van Maanen, 1979).

I entered the field with the goal of understanding how social media data and the classifications and categories that give them structure are used to govern distributed user bases, select user behavior, and understand users and their life contexts. In order to understand the role of data and data structures in shaping organizing, it was necessary to search for and uncover the processes in which they mattered the most (Avgerou, 2013), uncovering the activities that actors undertake to use, reuse and repurpose data (Leonelli, 2016). Given the state of the literature, which did not provide strong middle-range theoretical frameworks, I understood my research as exploratory. Ethnographic data collection is a powerful research strategy when operating under such conditions (Sayer, 2000).

7

However, the start of my study happened in the months immediately after some important learning events (which I will describe below). Staff were still sorting through the aftermath. In the way informants talked and fleshed out work issues, there was a diffused sense of the importance of the organization's history and the legacy of its approach to patient involvement, business, and research. References to the past of the organization were continuous, often to explain development decisions or project requirements. It was impossible to ignore the way in which the past could help to explain the present, and I felt obliged to embark on a reconstruction of PLM's history and expand the research scope to be at once latitudinal and longitudinal (Foucault, 1989; Pollock & Williams, 2009). A broadened research perspective turned out to be critically important (Wynn & Williams, 2012).

After the fieldwork period, I continued the analysis by cross-checking the empirical material with the aim of reconstructing the evolution of the infrastructure in a web of concatenated data-related developments. Adopting an understanding of the evolution of infrastructure as an open, emergent and complex process, and keeping my commitment on data and data work, I carefully linked purposeful action from the system developers to its observed effects. Several times, residual evidence suggested the emergence of unintended effects – empirical material that required explanation. I tried and identified multiple causal chains and applied retroductive logic to this aim: once I had selected a relevant event, I analyzed the data to find evidence of converging activities that may have caused it (Mingers, 2004; Sayer, 2000). The rich variety of empirical evidence I was able to collect allowed me to triangulate explanations – an important warrant for retroductive explanations (Yin, 2009). While ethnographic data collection necessarily has a purpose and a limited set of select topics of interest, it is relatively rich and unbounded, and allows broad explorations of the material.

Of course, this paper retraces this history from a perspective situated within the organization, through the situations inhabited by the organizational actors and their understanding of events occasioned by infrastructural change. As such, this account of multiple perspectives and actors involved in the infrastructure is constructed from, as it were, a 'single' perspective – that of the developing organization. Evidence about patient data use and value creation activity has been constructed by triangulating secondary evidence, as gathered through knowledge openly shared among members within the organization, but also the knowledge that was shared with me, through interviews, by staff who spent dedicated years of their work lives to communicating with patients and understanding their needs and activity. Against this backdrop, I hold that this account must be taken seriously because, as I will show, it reports on years-long organizational efforts to observe and understand the status of the network by the one party that is most interested in producing value for all involved – at stake is the organization's very survival. This limitation of empirical access comes together with the strength of methodological consistency and depth.

In the effort to place this narrative in the most appropriate context, I pursued, and abandoned, links with the literature to find an ideal fit. I wrote an initial analysis that framed the historical development of the platform in the perspective of Feenberg's critical theory of technology (Feenberg, 2010), opening up the issues of first and second instrumentalization and the role of experience-in-context in technological design. This initial explanation of the evolution and innovation dynamics of the infrastructure was not sufficient to explain the role of data and the activities around data and data repurposing, which was my base commitment. The analysis was evolved further by connecting it to recent perspectives from theories of value and knowledge

production. The multidimensional framework of value was initially inspired by Bolin's (2009) analysis (building from Bourdieu) of value forms in public TV content production. A multidimensional understanding of value, as I explained in the theoretical framing section, was the key to unite and differentiate between all the observed data-use and infrastructure development activities – to observe effects of infrastructure evolution. With this approach, I identified four dimensions of data-based value creation to explain the empirical evidence: *business*, *scientific*, *individual*, and *community*. These are defined in the next section.

4. EMPIRICAL FINDINGS

4.1 RESEARCH SETTING

PatientsLikeMe (PLM is an online community and data-intensive infrastructure connecting more than 350,000 patients suffering from more than 2,500 conditions. It was founded as a for-profit venture in 2005 in Cambridge, Massachusetts, by three MIT alumni. It was a site for patients of amyotrophic lateral sclerosis (ALS) to connect, and give and receive peer support by sharing information – including treatment regimens and coping strategies. The three founders were trying to do everything possible to save brother and long-time friend, Stephen Heywood, who suffered from ALS. The website was the Heywood family's second major venture, having previously founded one of the most prominent patient organizations for ALS research, ALS TDI.

PLM sits at very complex crossroads. It is not powered by advertisements and its patient users voluntarily contribute self-report data for multiple reasons: for use in research, to make connections with other patients and socialize, and to keep a personal health record that captures their health trajectory. As such, the developing organization is tangled up in multiple operating principles. Started by patient activists and so-called 'guerrilla scientists', the for-profit company makes money by selling research services, including pseudonymized data and customized research reports. PLM also tries to exploit the data collected through the infrastructure to simultaneously serve the patient user base (manipulating data to construct digital spaces for social interaction and self-representation and self-understanding) and produce relevant information to fuel its own research (to build the legitimacy it needs to prove its model). Through patient self-report data, the staff have published some 37 outputs, including peer-reviewed journal articles, on a broad range of topics (e.g. Turner et al., 2011; Wicks, 2007; Wicks & Frost, 2008; Wicks & MacPhee, 2009; Wicks et al., 2011). PLM is routinely mentioned by industry and government agencies alike as the poster child for visions of the future patient as empowered and connected (Manyika et al., 2011; Richards et al., 2015).

For PLM to be a sustainable project, the value that is generated from the data cannot be exclusively for business or research, but must also be of collective and individual use and align with norms and moral values. There is a complex convergence of value dimensions at play to shape data, data structures and infrastructure evolution. Importantly, these dimensions should not be thought of as exclusively related to a stakeholder. Different data users are interested in multiple value dimensions at once. The four dimensions of value in PLM are defined by the *situation of expression and use of data-based information*. Digital data are symbolic objects and, as such, they bear different informational value depending on the situation of use (Tempini, 2015; Bateson, 1972; Kallinikos, Ekbia & Nardi, 2015). Accordingly:

- *Business value* can be generated when the data give good information for creating commercial research reports (i.e. what patients say about a particular drug).
- *Scientific value* can be generated when the data give good evidence for conducting health research, including construction of claims and publication of peer-reviewed articles.
- *Community value* can be generated when the data can be used to foster social interaction within the network to create a thriving and inclusive community.
- *Individual value* can be generated when the data can be accessed by the patient user in a way that is useful or rewarding, to understand the health situation or to develop self-representations.

Exploring the ways in which standardized digital data can be collected, circulated, used and reused is the underlying strategy upon which PLM's attempts at coordinating such a complex assemblage of social actors relies.

Through the website, patients self-report their health status, logging data over time about a number of health dimensions. They manage their personal profiles, designed to provide a snapshot of the patients' present and past health developments through infographics, scores and text. Website features include private messaging, broadcasting, commenting and spaces for self-representation (such as the profile picture or the *About Me* text box). Many patients find *individual value* in compiling and reviewing this data-intensive diary of their journey: it helps them build narratives and interpretations of the self. The platform offers several health-tracking tools. Tracked dimensions include conditions (primary and co-morbidities), symptoms (severity, associations with treatments or conditions), treatments (dosage, form, frequency, side-effects), lab measures (for instance, blood cell counts or forced vital capacity), hospitalizations (reasons, dates), disease-specific patient-reported outcome measures (PROs), and others. To develop a system that successfully captures health data through this approach, it is essential to develop a template of patient experience that is meaningful to the patients (regarding their health and life experience), such that they will be encouraged to input data.

At the same time, the system must be able to support the research (health economics, drug evaluation) that the company has been developing for its clients. Data must support the creation of *business value*. Moreover, the network must also be able to produce scientific research to peer-review standards: data are used in *scientific value* creation and need to support original arguments and outputs. Demonstrating the scientific basis of this data collection approach has been a critical passage point for the organization to be able to sell its research model and, ultimately, its business services. Research, business, community building and patient involvement are entangled goals that shaped system requirements as PLM developed, one release after another, over the years.

Patients also participate in the platform to connect with other patients, for socialization, support, to source information about alternative treatment regimens or coping strategies, or to learn more about others' dosages of a drug or health risk thresholds (Wicks et al., 2010). However, to construct the possibility for community and foster social networking, the system is designed to display data-intensive information nearly everywhere on patient-facing web pages (see also Kallinikos & Tempini, 2014). It relies on data-intensive functionalities to host and foster sociality and to represent the health of patients through visual, numerical and textual analytics that continually compute the underlying database of patient data. Interfaces are

replete with interactive links that guide the user through the virtually infinite social space of the network.

The system automatically computes and renders web pages that display scores and charts about one's profile or the 'state' of specific medical phenomena (e.g. a particular symptom) across the platform as a whole. In the first instance, pages such as *My Symptoms* put the log of one's symptoms in context by displaying a longitudinal view of the tracked symptoms and related severity scores. In the second instance, pages such as the *Symptom Report* page or the *Treatment Report* page give an overview of the information that has been shared about that particular medical entity on the platform, including the distribution of a symptom's severity as experienced by the platform members, demographic information about the patients experiencing that symptom, the treatments that those patients use to fight it and conversations on the forum where the symptom is discussed. These automatically generated information items are interactive and can be clicked by the patient to connect to other patients suffering from the same symptom or talking about it on forums and so on. Similarly, and in addition to the report pages, the website offers a number of forum rooms and threads, organized hierarchically by condition type (plus a few miscellaneous forum threads for topics such as politics or platform announcements): there are forum rooms for cardiac diseases, neurological diseases, etc., all provided with interactive links, as just described.

Based on computational operations and automated link generation, opportunities for interaction are identified and suggested to the patient in the form of interactive links. The interaction possibilities that have been engineered into the PLM environment realize a loop whereby the more patients share with the system about their health status, the more the system is able to produce (through continuous computational operations) useful and up-to-date links to other patients or pieces of information that have been shared elsewhere in the system. Data analytics becomes a tool here for understanding, constructing and taking part in the community that the network creates, and generating new data as social interaction unfolds. Data are then also processed to deliver *community value*.

4.2 A CONDITION AS HORIZON: TIGHT COUPLING OF USER AND PATIENT EXPERIENCE

Initially, PLM supported only one condition (amyotrophic lateral sclerosis – ALS) and it was designed based on the founders' direct knowledge and research of life experience with ALS. ALS is an extreme case of a patient experience. There is no cure and patients retain cognitive capacity throughout the convalescence while progressively losing control of voluntary motion over a short time period (life expectancy after diagnosis is two to five years). In addition, ALS belongs to the category of rare "orphan diseases", which attracts very little to no R&D investment. For these reasons, ALS patients are known for their activism and experimentation with anything that gives them hope. The experience of this condition dramatically dominates a patient's world and eventual co-morbidities recede into the background. As of September 2016, the website states that 9,180 ALS patient profiles are on the platform, an important proportion given that a very high majority of members are from the US and the total US ALS population is estimated at around 30,000.

The original website featured a forum focused on discussing all matters related to ALS. Patients could also track the course of the disease through an ALS-specific PRO questionnaire and a fixed list of 40 symptoms selected by the staff. Similarly, the treatments that could be tracked were also a fixed list of those that the great majority of ALS patients take – above all, the drugs

Riluzole and Baclofen. The website's focus on a single condition and the fixed lists of symptoms and treatments built an implicit, clear context into the user experience: the site was about ALS. However, the founders soon realized that the data collection model based on fixed lists of items was too restrictive and wasted opportunities for learning. With a view of growing the value (especially *individual* and *scientific*) of the database, they decided to continue promoting the tracking of the lists of "primary symptoms" and "commonly prescribed treatments", but also allowed patients to create and track other symptoms and treatments, opening the medical data architecture to user-generated categories. Quickly, patients started to generate a wealth of new symptoms and treatment definitions, many of which were not 'strictly medical' in the professionals' view, while others could be linked to existing expert terminologies but described the patient experience at a higher level of specificity (Arnott-Smith & Wicks, 2008).

The website grew with the addition of other disease communities, starting with multiple sclerosis (MS) and Parkinson's disease (PD). The staff had to research each condition to determine what set of tools would be better able to capture the patient experience in data and could be used for multiple purposes. They developed a PRO tool for MS due to the lack of an existing one, while they adopted a previously validated one for PD. As for the rest of the system, it was possible to use the same tools with little adaptation. ALS, MS and PD, very different conditions from each other, still belong to the same broad family of diseases (neurological). Some similarities emerge, especially in comparison with other globally existing medical conditions. PLM constructed a patient experience template to manage these different diseases by developing on top of the architecture initially developed for the ALS community. These three patient communities remain some of the most active and successful that PLM has ever launched.

Over the first five years of its existence, development continued along similar lines, with the launch of about 25 patient communities. These communities linked to conditions that could be very different (the group included, among others, epilepsy, fibromyalgia, HIV, transplants, and mood conditions). For each of them, the staff had to go through extensive and time-consuming research to select the set of self-reporting tools that would provide a user experience that could align successfully with the patient experience, and produce individual value to the patient members, while, at the same time, providing clinically meaningful and reliable data. When possible, the staff would adopt validated tools among those available. Otherwise, they developed and published new PRO tools for the purpose. Importantly, the choice of which disease communities to launch was dependent on developing business partnerships with clients that were interested in PLM's research services. Pharmaceutical companies have typically been starved of real-world information about patients' outcomes and their evaluations of competing drugs. PLM offered a unique access point.

4.3 Aging architecture: The crunch between network growth, community service and scientific understanding

Over time, it was increasingly felt that the website could serve a larger global patient population and diseases numbering in the thousands. Expansion of the platform was slow, as development was dependent on partnerships with clients who would fund the research necessary for developing a community. The organization received thousands of requests from individual patients asking for a community for their own disease, while the business development staff envisioned a transformation from a customized research services boutique to a one-stop shop that would also offer direct 'raw data' access to the clients' in-house researchers. The aim was to

satisfy all the needs of the big pharmaceutical companies for online patient community data, and thereby to obtain bigger, longer-term contracts. With an interest in improving their own understanding of the patients' lives and health to expand their own business offering, scientific understanding and patient services, the staff planned a solution whereby PLM would become a *universal online patient community*. However, to satisfy these requirements, the staff had to deal with the materiality involved in managing and analyzing data and data structures.

The installed base of legacy application and data architecture developed incrementally over five years and, from a humble beginning, was now an obstacle to innovation. Each disease-specific community was a walled garden of its own and accessed the website from different sub-domain URLs. Worse, patient accounts were tightly coupled with one disease as a result of the legacy database design, which had been responding to old business value requirements (i.e. creating a single community per project). Each patient account was associated with one fixed condition number, which could not be changed. The condition number was the pivotal data point coordinating system behavior.

The number would drive the execution of different portions of the codebase, resulting in different system interfaces and behavior, to reconstruct a context appropriate to the experience of a patient with a particular condition. The number also drove the automatic association of custom, disease-specific PRO tracking tools with patient profiles. Each number loaded a different configuration of the system. In this way, patients accessed different forum rooms. Moreover, the condition history questionnaire (the system's metaphorical equivalent of the clinical interview, in which patients share the date on which the first symptoms manifested, their family history, date of diagnosis, etc.) and the lists of *Primary Symptoms* and *Commonly Prescribed Treatments* to track were different for different conditions.

The tight coupling of patient profiles with individual conditions also meant that patient profiles were not designed to track co-morbidities. This was initially difficult to notice, as ALS and PD are so devastating that they make the existence of co-morbidities less visible – and reported. However, as the staff learned how many patients struggle with multiple conditions (e.g. MS and bipolar disease, or MS and epilepsy), it became clear that the data model was simplistic. The huge number of combinations – between a condition and its co-morbidities – that might be meaningful from a research point of view was neglected by the system architecture. The model was thus also limited from a scientific value point of view: this severely limited the scope of research that the staff could produce from the data.

Finally, the architecture fell short from a community value point of view too. Patients accessed siloed forums. Social features suggesting navigation to profiles of other patients were limited to patients with the same condition number. Some of the patients managed two separate, siloed accounts, one per condition. Patient accounts were associated with forum rooms dedicated to one condition exclusively, and patients could not access forum rooms dedicated to other conditions. While this focused many forum conversations and allowed condition-related sensitive issues to be discussed more freely, the staff felt there were aspects of the patient experience that patients might wish to share and discuss with patients suffering from other conditions as well. For example, it could be useful, especially for patients belonging to less active communities, to discuss wheelchair options with patients from other communities as well.

Patients invented workarounds so as to be able to track the events that mattered to them. Soon after the introduction of the open architecture for user-generated symptom categories, patients started to exploit it creatively. One of the ways in which patients accommodated the feature was to use it to track co-morbidities in their symptoms list as if they were custom symptoms. The workaround thus created value for the individual patient and for the community (the system could produce links to other patients), but the symptom data model was too simple for a condition, tracking its "severity" on a given date on a scale of "None", "Mild", "Moderate" and "Severe" (NMMS). The burgeoning growth of co-morbidities records in the symptoms database demonstrated the complex reality of the patient experience, but, as such, these data were not valuable for scientific or commercial research.

At these early stages of the platform's evolution, the combinations of user experience templates and data architecture were limited in number. Similar to other patient networks, both online and offline, a specific medical condition was the implicit background of the site's user experience and social interaction. Fig. 1 summarizes this former architecture, taking three conditions as an example. Two are relatively similar to one another (ALS and PD, both neurological) and one (HIV) is very different from them.
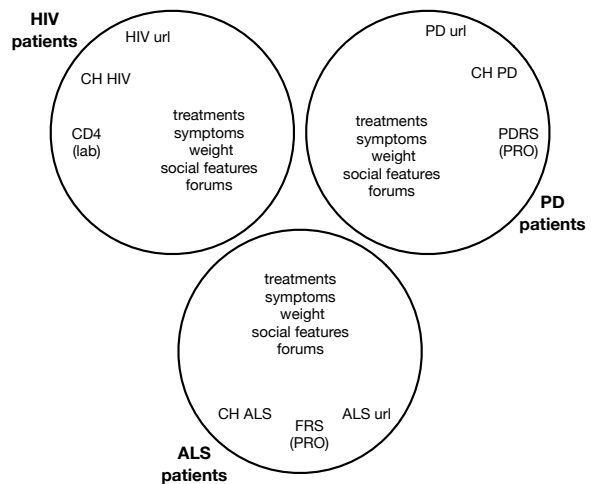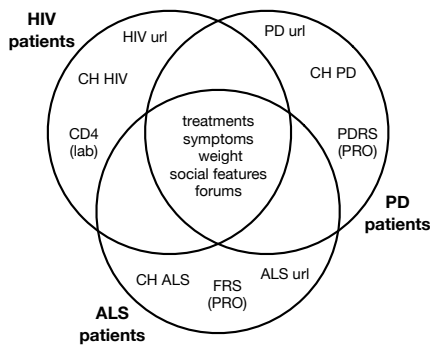

Fig. 1 – Pre-GP architecture of PLM.

4.4 THE GENERALIZED PRODUCT (GP) PROJECT: FROM TEMPLATES OF EXPERIENCE TO THE TEMPLATE OF EXPERIENCES

Running a more sophisticated configuration for each patient profile was crucial to collecting useful data that could be processed to foster interaction and create more community value, to drawing a meaningful personal profile for more individual value, to informing commercial research reports and producing more business value, and to providing more scientific value as a space for scientific investigations. With the Generalized Product (GP) project, implemented in

the spring of 2011, the organization embarked on an infrastructural renovation of the system that had no antecedent in the site's own history, employing all of the available development resources for more than six months.

The team set out to unbundle layers that were once glued together. The data architecture had to have a broader scope and become easier to manage and configure. It was no longer feasible to hardwire fixed configurations of patient profiles and conditions. First and foremost, patient accounts would no longer be tightly coupled with one condition. The system had to be able to flexibly capture the relationship between parallel conditions and the consequent ramifications for associated treatments, symptoms and other entities. It was necessary to establish and modify multiple relationships between patient accounts and conditions according to how the patient experience evolved. The system also needed to find a way to quickly scale up the number of conditions represented in the system and move away from the slow, in-depth and labor-intensive research process used to model conditions.

Once an implicit background of the user experience, the condition now had to be approached as a contextually embedded object. The GP was a new architecture of condition management designed to allow patients to add conditions to their profile or request the creation of custom conditions. Patients could now make a request for the creation of a condition after an unsuccessful search in the database – similar to how they created definitions for symptoms and treatments in the old architecture. The data architecture could now expand at a much faster rate. Less than a year after GP release, the system was hosting 1,400 conditions.

To achieve on-the-fly condition creation, the GP required the modularization of relevant features to make the code available to other condition entities. Before the GP, conditions used a shared codebase for a few functionalities (see Fig. 1). With the GP, more system features needed to be shareable and deployed on a needs basis. Employing software engineering resources to reuse code each time a condition had to be created was not scalable. Through a new "condition admin tool", conditions became objects that staff could create and model at any time in the admin area of the site. The electronic form generated and stored a configuration of system modules and expert classification system codes necessary to drive the system's behavior and aggregate patient data. A new database "condition_ID" key took the place of the once-fixed "condition number", which would no longer act as the overarching capsule of a patient community. Creating conditions on a scalable basis necessitated the standardization of the condition entity (along a limited number of dimensions) and the kinds of configurations of tools and modules to which it could be associated. In the new system architecture, it was not necessary to study each condition in depth before creating it. The staff analyzed a list of more than 3,000 conditions from the authoritative medical university Karolinska Institute, based on which they compiled a set of six condition types within which all conditions should fit: infections, chronic diseases, pregnancy-related, mental health, events and injuries, and life-changing surgery. Each condition type drove different behavior in the system, such as the questions in the condition history questionnaire: this would not be a highly contextual, disease-specific set of questions, but it would avoid the unnecessary, off-topic ones. As such, by just filling in a name, a short name and synonyms, choosing a condition type, and coding the condition against standard expert terminology codes (including MedDRA, ICD-10 and SNOMED CT), the staff could immediately create a new condition record to which patients could link their experience data. The large majority of conditions added in the first year of the GP launch were created with such minimal information assets.

The organization designed upgrading processes to "elevate" specific conditions through "condition levels" – from level 0 (condition "communities" created in a few clicks and linked to generic tools) to level 3 (fully researched, developed communities that are active and equipped with high-standard PRO tools). Due to resource constraints, opportunities for condition elevation were mostly dependent on research projects linked to business partnerships. Laborious research was needed to develop and configure disease-specific tools such as PROs and labs, and the lists of *Primary Symptoms* and *Commonly Prescribed Treatments* that could be tracked. The user experience for level 0 conditions was highly standardized,[1] but the staff hoped that it still offered enough tracking tools and socialization features to get a patient community started. The hope was to enable patients of new conditions to sign up and create profiles, eventually giving birth to a new community. The basic set of tools was intended to provide enough individual and community value for patients to stick around and use the website. The data thus collected was expected to be of basic scientific and business value, but still a useful starting point to develop valuable in-depth investigations. Importantly, this process flipped the order of investment in a community: now the organization followed after patients, instead of the opposite.

All requests for the creation of a condition were reviewed by the HDI team's staff, who could then create a new condition in a few clicks. The new architecture included the User Voice (UV) dashboard, displaying all the category creation requests that patients submitted when they wanted to add certain phenomena (conditions, symptoms, treatments) to their profiles that they could not find in the database. HDI staff used the dashboard to review requests and take action along a number of courses. In several instances, the intervention of the professional clinician was deemed necessary to settle a situation through professional judgment and offer a sensitive human touch when communicating with the patient.

The staff's main concern was avoiding duplication of the same condition under different labels: that would split the community and destroy community and individual value, as well as make it more difficult to produce robust research or commercial and scientific value. Many patients indeed simply inputted alternative definitions of existing conditions. Moreover, patients often inputted as "conditions" other medical entities such as symptoms, or made simple errors (e.g. complex misspelling errors of the sort that automated tools do not catch). If unguarded, these events would make the data too fragmented and practically unusable in the long run. At the same time, the staff could also opt not to create conditions (especially in the case of entities whose epistemological status is discredited in institutionalized science). This option could be applied to defend the reputation of the platform as a reliable source of scientific reporting (caring for scientific and business value). However, it could conflict with the value that the requesting patient would get from the profile and the potential community of patients that the rejected label could assemble.

---

[1] At level 0, the system pushed only five generic symptoms to all patient profiles, holding them as basic denominators of the patients' experience of all conditions: fatigue, pain, insomnia, anxious mood, and depressed mood. All patients were offered self-reporting tools for symptoms and treatments, together with other general tools – the quality of life (QoL) questionnaire, the daily mood–health tracker (InstantMe) and the weight tracker.

The organization had to walk a thin line between comprehensive capture and epistemological chaos. Monitoring was necessary to maintain desired levels of data aggregation, and to avoid the proliferation of duplicate definitions of the same phenomena in myriad patient languages. Fragmented data would be useless for commercial, scientific and community purposes. It would not be possible to use them for the elaboration of reports or scientific papers, nor to power the socialization features that promoted social interaction across the website by calculating relations based on shared data points.

In the months following the launch, the staff also undertook a laborious review and integration of all those conditions that had been inputted spuriously as symptoms, the legacy of the workarounds that patients used before the GP. The database held more than 300 such "conditions-as-symptoms" records, many of which linked to multiple patient profiles – several in the hundreds. These records were migrated into proper condition records. The migration resulted in a loss of patient data: the symptom severity data (NMMS scale, above) according to which these conditions had been tracked did not fit the condition data framework. Originally of some individual value, these longitudinal patient data were 'switched off', practically lost, as the infrastructure was unable to compute and display the data without misrepresentation. These individual data and their personal, hidden meanings were misaligned with the system's "medical framework" templates (e.g. condition report page representations, functionalities and links to other entities). The risk was alienating other patients (destroying community value) through confusing data representations and visualizations. Importantly, the chances of using these data for scientific or commercial purposes were negligible. An explanation of the change was shared with the affected user base and the patient profiles linked to the reviewed legacy categories were re-linked to the original conditions – in the new guise – so that those patients could start anew and share data again. Fig. 2 highlights the change implemented through the GP project, for comparison with Fig. 1.
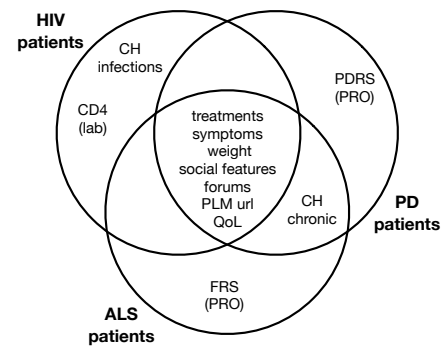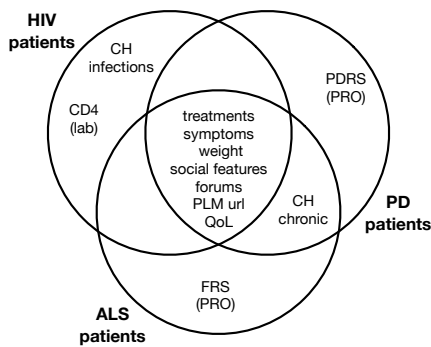
Post-GP architecture

condition_ID key determines the separation of the data and user access to content

Legend
CH: condition history
PRO: patient reported outcome

Interface view*

Patient data are aggregated and connected to patient data of all conditions
Patients interact with all patients

Codebase view

Patient profiles execute both shared and disease-specific code

*: - Labs and PRO tools could be used by
conditions other than HIV, ALS, PD;
- CHs data could be linked across different CH
categories (e.g. date of diagnosis)

Fig. 2 – Post-GP architecture of PLM.

4.5 DIGITAL SITUATIONS: FROM CONDITION COMMUNITIES TO A "COMMUNITY OF CONDITIONS"

The consequence of the GP was that existing patient networks were merged into one dynamic and ineffable network. Patients now belonged "fluidly" to multiple groups sharing one or several conditions in the context of a single, universal platform population. All forums were now open to all patients: the web application simply adapted navigation shortcuts according to a patient's conditions. More importantly, the system now calculated the many links to patients, reports (conditions, treatments, symptoms) or discussions that the platform generated and distributed across the web pages of the system – in the form of clickable scores, icons and conversation snippets, among others – by parsing through the entire database, taking into account the dynamic relationship between a patient and condition entities. Data structures were coordinating the system more than ever.

Formally, the data architecture was now highly comprehensive and better able to represent the diversity of patient experiences and contexts. However, the user experience had become generic, the data were somehow less informative, and the researchers felt that it was becoming increasingly difficult to better understand what patients were experiencing. Gains were challenged by the questions that the GP transition forced to the fore.

Something at the heart of the community was getting lost in the architectural translation. Patients from established legacy communities complained that they felt less oriented and able to navigate the website than they had previously. The Community Team, staff involved in patient communication (moderating forums, collecting user experience feedback), were concerned about the risk that patients from some conditions could be harassed by other members regarding the lifestyle stereotypically associated with these conditions (such as HIV).

Level 0 conditions were struggling to enroll and keep active the "sticky users"[2] who are dedicated to data collection and community life. An executive reflected:

"When we built GP, I think what we really learned was that what we meant by a condition was something very different from the medical definition of a condition. What we meant by a condition was something that a group of people identified with. "I have this condition; I am this kind of patient." And there's a difference, and I think we didn't understand that distinction, and there's a big sociological element to a condition."

In principle, patients' self-report data could be very valuable for research even with just a rigorous application of the now flexible core framework (conditions, treatments, symptoms). However, while symptoms and other forms of data were indispensable, they were not sufficient to explain what was going on in a patient's life. The core framework in itself was not enough to sustain desired levels of engagement – patients were not participating, and researchers were not getting good data. Level 0 condition communities were not collecting data in the way needed. The absence of custom tools and questionnaires and, most importantly, the community context that was implicitly constructed when websites were dedicated to one condition only was making it harder to attract and retain the critical mass of patients that would make social interaction on the website a self-renovating resource.

While the data architecture was flexible and ready to receive data from all kinds of patient experience situations, social functionalities had not evolved at pace. Communities were unbundled. The now piecemeal segmentation of the user base multiplied the trajectories for sociality. It made the framework for the generation of social links and interaction scalable and able to connect people across a broader range of medical phenomena, but along unpredictable and difficult-to-manage trajectories. Over the following months, the organization started taking measures to restructure its design process towards user-centeredness. User experience consultants were hired to help it think about how to better serve patients. Unable to create more community value and individual value, the GP threatened value creation growth also on the commercial and scientific research sides.

4.6 DREAMING A UNIVERSAL PATIENT EXPERIENCE INFRASTRUCTURE

The data collected through the new architecture now was much more comprehensive. And yet, the epistemological complexities unleashed by the architectural renovation were far-reaching. In an environment with multiple equally recognized conditions, researchers were unable to systematically determine which symptoms were attributable to which conditions. Working with correlations risked obscuring the associations that were the most valuable to discover – the "small patterns" (Floridi, 2012) that would make a breakthrough. In addition, while the system was now able to record the difference between present and past conditions – enabling it to record clinical history over a considerable time period – there was no clear way to automatically order present conditions by some kind of "priority" and adapt system behavior accordingly. As another informant, a research scientist, elaborated:

"Why is a heart transplant more interesting than chicken pox? There may be a case where chicken pox is more interesting than a heart transplant. But what is it? Is it just that it is worse? What is

---

[2] "Sticky users" is an actor's category.

the measure of worse? What's worse: stage four breast cancer or stage four pancreatic cancer? How do you measure that? Survival? Impact on life?"

Hitherto hidden, a certain lack of context of the data was now being revealed by the GP release. PLM staff were reflecting on questions that had never been so pressing before. They emerged as a result of the GP's shot for the moon – to become a universal technology. Multiple areas of the site were affected, from the organization of forums to the sign-up workflow (which was now generic and unable to get patients to link conditions to their profiles immediately). Expanding the range of medical phenomena and situations that the system could describe in its data architecture was increasing, in an exponential fashion, the range of questions that now required an answer. The same informant added:

"It wasn't until we added the GP that I suddenly realized that she had breast cancer. Obviously, knowing that completely changes your interpretation of the rest of the contextual information".

The poignant quote is emblematic of the complex and uncertain process of reconstructing patients' life contexts. Understanding patients, the numberless combinations of their life paths, and how to capture that through systems had been the overarching concerns in devising each system development iteration.

This history recapitulates the PLM staff's progressive discovery of how deeply the patient data linked medical, social and personal contexts with one another. Innovations in the system had repeatedly changed the understanding of who the patients were and what their health experience was. The data collected through the continuously evolving infrastructure progressively revealed a hitherto hidden web of "facts" – both in the statements that the data appeared to support at face value and in the emergence in the form of questions of other yet-to-be-addressed areas of understanding. Developing automated solutions for giving valuable information back to the patients became even more difficult in the face of all this situational uncertainty. How could they find out what information patients needed? What are the orders of facts that matter the most for the patients? There was no one universal standpoint, and there never would be. The researcher reflected:

"Suddenly you go 'Wow, this is not just about things we can even measure or things we can even know. It is about an incredibly wide variety of things.' So, there are things that we can measure and there are things that will be immeasurable in the system."

Sharing information would be persistently useful for patients only if it created value in multiple dimensions at once. Understanding the experience of living with a condition through the intermediation of an open, data-based space meant much more than dealing with the collection of 'purely medical' data. It required holistically embracing the multiple dimensions of human life. To conduct science for and with patients required rethinking what the platform should be about and concerned with. The platform needed not only to describe a variety of phenomena. It needed to embrace the variety of their meanings as well. Another informant, an executive, reflected:

"We should be able to think of conditions in multiple domains – they mean something medically, they mean something sociologically, they mean something from an importance standpoint, they mean something as a milestone in one's life. […] We confounded, we originally built communities

that overlapped with conditions, and thought they were about the condition. And then, when we built a community of conditions, we did not understand what those communities meant."

Learning as it opened new territory, the team continued its development efforts to build solutions that would close the gaps that were progressively highlighted. The underlying realization was that of the sheer complexity of the road ahead, of how difficult it would be to understand, through data, what kinds of patients PLM users were, what their experiences were like, and what their concerns and needs focused on. While the team had come a long way from its beginnings, its relationship with the users was no less bound to uncertainty.

5. DISCUSSION: DATA VALUE AND USE IN DATA-INTENSIVE INFRASTRUCTURES

Fig. 3 summarizes the dynamics of the Generalized Product (GP) project. As we have seen in detail, the GP was an unprecedented and complex undertaking, with dramatic consequences on the life of the online community gathered on PLM. Different kinds of value questions (pictured below, Fig. 3) were made to coalesce in the formulation of the GP requirements. As with previous releases, candidate solutions to the requirements were assembled in the GP release. No data-based value creation question could be addressed without integrating requests from competing value dimensions. Yet, the consequences of innovations implemented to address these value question 'entanglements' cannot be traced back to one specific value dimension, as along a single line of continuity. Instead, innovations driven by value creation questions are concatenated in a longer, stepwise and nonlinear trajectory of infrastructural evolution. The sequence of infrastructural innovations was punctuated by the emergence of unexpected consequences with implications for value creation.

The case material suggests that infrastructural innovation for multidimensional, data-based value creation is inherently problematic and contingent, and can have destructive consequences that compromise data-based value creation processes. This finding warns us about understanding infrastructure development for value creation as a largely linear trajectory. We observe a great diversity of individual, idiosyncratic data valuation and use practices that features only recalcitrantly in the broader analytical project built to underpin them and around processes of standardization, aggregation and comparison. In addition, the development of specific technology for this purpose, namely the calculative interactive environment that I have been describing, further exacerbated this friction. Explaining the dynamics underlying our observations will be the focus of the discussion.
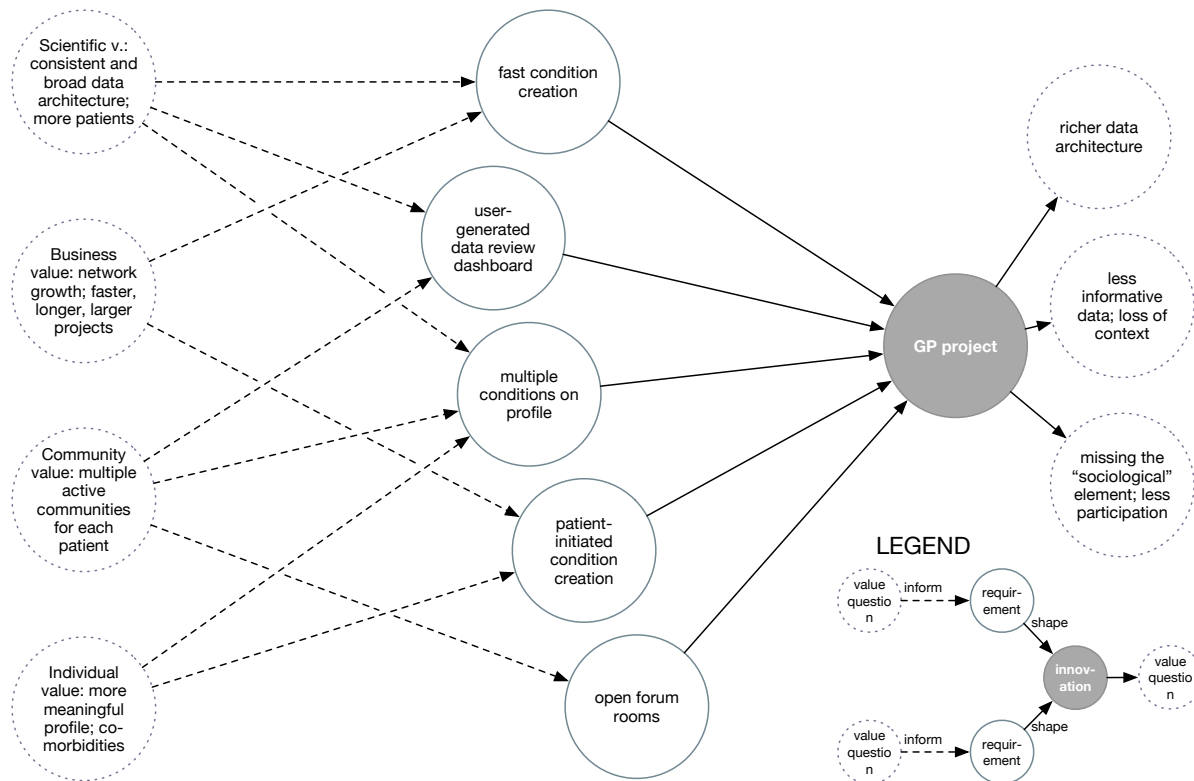
Fig. 3: Value entanglements of the GP project.

While analysts are customarily envisioned as the data users of online community data, in a data-intensive infrastructure, users (here patients) are data users as well – the infrastructure itself is designed to compute and circulate the same data in order to support further user action and data generation. This "multi-versal" network of data-use recalls, in its own way, some well-known dilemmas of information infrastructure development: the unpredictability of its behavior and development, the limited scope for design and control (Hanseth & Lyytinen, 2010; Hanseth, Monteiro & Hatling, 1996; Star & Ruhleder, 1996; Tilson et al., 2010) and the complexity of coordination. Working from the example of the development of the Internet, Hanseth and Lyytinen (2010) have emphasized the nonlinear, emergent, irreversible and unpredictable evolutionary trajectory of information infrastructure. Because infrastructure projects lack user definition and use determination, designing a universally valuable technology means tackling the chicken-egg problem of achieving both user growth (bootstrapping, to enable network effects) and sustainable adoption (adapting to contextual diversity). The authors suggest addressing the bootstrapping problem by "generating early growth through simplicity and usefulness; [… and] the adaptability problem by promoting modular and generative designs" (2010:1). Similarly, Henfridsson and Bygstad (2013) pose "decentralized control and loosely coupled architecture" as conditions for the infrastructure to further develop its value creation capabilities.

The PLM case seems to suggest that the reality of data-intensive infrastructure development might make such design principles difficult to follow. In this specific context, concepts such as simplicity, usefulness, modularity and generativity can be taken to refer to the architecture and performance of data-intensive features, and their support to practices of data use. The GP was, in many respects, a project of this sort: it aimed to shift the data structures over to a loose coupling model (from condition numbers and siloes to flexible relationships of patient profiles

and medical entities) that would make data-intensive features (patient searching and filtering, data collection and aggregation) more flexible and able to decompose numberless diverse patient histories into a myriad of constellations of their basic elements – conditions, symptoms, treatments, etc. It was meant to improve the efficiency of communication and reduce redundancy – both essential to the construction of what amounts to a *calculating interactive environment*.

However, infrastructural innovations (including the GP) also disrupted value creation processes and contingent reaction. This was surprising and contrasts with the easy-to-make assumption that generic and flexible templates are the recipe for online community development, which seems to be backed by the successes of the likes of Facebook and Twitter. The modularization of the infrastructure codebase tested the ways in which adaptability can follow from reduction and simplification. The new, loosely coupled architecture, meant to be the most adaptable version of the underlying combinatorial system of data structures, was falling short of serving all kinds of data users: not only analysts, but also patients (especially from level 0 conditions, as we have seen). Let me identify some causes for this impasse.

5.1 THE FRAGILE CONSTITUTION OF (RE)USABLE DATA

First, it is crucial to understand that the disruptiveness of infrastructural innovations that we observe is linked to the semantic nature of digital data, and the extreme diversity of activities use and value creation. Infrastructural innovation can threaten the life of an online community when it alters the data-based resources and practices that its members use to communicate and share, as this can have unpredictable effects on the structure of social interaction and the context of data interpretation.[3] Actors shape their practices with data according to the moral commitments that are informed by situated notions of worth and order. Practices of distributed data collection and use constitute, through their enactment, a shared social order – the foundation for organized collaboration and communication. If, in the first place, a changed context of reporting fails to make sense to the data reporting users, they will be unable to share meaning as they intended until they find ways to restore the shared order that sustains cooperation and communication (Rawls & Mann, 2015). Information is recreated (Garfinkel, 2008) from the data by each actor and always within the frame of evaluative action, even if implicitly.

As we have seen, infrastructural innovations (for instance, change to data structures or data-intensive features) repeatedly alter the fragile equilibria of practices that develop to sustain mutually intelligible communication and intermediated cooperation. Consider, in PLM, the comprehensive symptom data collection model, which includes user-generated categories and allowed for the development of the conditions-as-symptoms workaround: if actors do not

---

[3] Indeed, the nature and identity of what is shared through symbolic objects is constituted in the very practices of communication (Garfinkel, 2008; Rawls & Mann, 2015). Meanings and language rules are always open to be refashioned according to the situation. Cooperatively efficient communication can necessitate informality, redundancy and change, as the situated actors require them to be able to achieve mutual understanding (Garfinkel, 1967; Star, 2010; Star & Ruhleder, 1996). Reduction, formalization and minimization test the stability of shared social practice.

share a set of uses and evaluative criteria with which to interpret and work with the data, data that are nonetheless shared onwards fail to be usable and valuable elsewhere.

Crucially, the "systematic" idiosyncrasy of individual practices (each compatible with the creation of individual value) challenged the creation of other forms of value from the data and through operations of aggregation, computation and analysis. The need to serve multiple value creation demands set development on a trajectory of standardization that sought "universal" data reusability, but one that repeatedly tore a vast fabric of continually developing commitments towards shared meaning and orders of worth. Often, the extent of this disruption is unintended and requires contingent countermeasures, as the effects reverberate across all situations in which data are generated and used. Effects, in all examples, are the collapse of shared communicative practice and the "isolation" of actors from one another and into one's own situated semantics.

This disruption can be both purposeful and unintentional, and introduced both via updates to the technological infrastructure and user-led changes in data-based practices (most often, the latter as a reaction to the former). We observed disruption caused by interventions, like in the GP, that inadvertently put at risk the patients' ability to use the system meaningfully from an *individual* or a *community* value perspective. But changes in value creation capability were also caused by changing user behavior resulting from local use innovation. For instance, early user-generated category features enabled those patient workarounds that, without targeted countermeasures (e.g. data review, dashboard), threatened the long-run sustainability of the infrastructure from a *business* and *research* perspective. Accordingly, countermeasures to patient workarounds included implementing processes to realign user-generated categories with all value creation dimensions. This included ensuring their creation in the appropriate group (e.g. conditions/symptoms) and in a language that could be understood by others. Reactive solutions addressing the consequences of designers' or users' innovations generated, in turn, further consequences for value creation. The infrastructure developed in a *stepwise* and *nonlinear* fashion: ramifications were multiple and multiple requirements at once coalesced in determining the value questions that successive innovations would try to address.

Understanding the way in which the data's symbolic nature and digital materiality underpins the ensuing complexity helps us to realize how difficult and inherently problematic it is to organize for data-based, multidimensional value creation. As we have seen, the anxious reflexivity following the implementation of the GP's data structures (which were supposed to be adaptable, universal) was associated with a sense of loss of purpose, of order and of value of the data and the practices underpinning their generation. This situation of broken cycles of knowing very much echoes the *infrastructural transcontextual syndrome* that Star and Ruhleder (1996) discussed in their seminal paper on the adoption of computing infrastructure technology. In their study, the diversity of users and use situations that the technology had to adapt to was creating structurally persistent problems of situated learning and organizing. But the difference in the empirical focus yields different double binds and learning issues here. In my study of data and data practices, as put in relation to the development and use of a data-intensive infrastructure, the ensuing "double levels of language in design and use" (Star & Ruhleder, 1996) are related to the inability of several participating actors to make use of the data and make sense of each other's practice. The data become too ambiguous to be made sense of in a trustworthy way. The impossibility of trusting one's own interpretation of the data and of the context of their generation provokes reflection on the relationship between data and user

life situations. "Inappropriate" use, as well as "inconsistent" data collection practices and workarounds, isolate data users from each other's contexts and produce data that is difficult to use for the production of multiple forms of value.

Idiosyncratic, divergent data practices might well be associated, prima facie, with individual value creation, and work at cross-purposes with the creation of other forms of value (e.g. business, scientific, community). However, to the extent that some analytics features can deliver individual value to the community member as well, the more limited aggregatability and comparability of idiosyncratic data seems to potentially undermine this form of value creation. Choices made for the opportunity of individual value creation can undermine other opportunities of the same kind. Importantly, in the move that this study makes from the "context of computing" to the "context of data practice", the technology of the data-intensive infrastructure is not the center of the observed infrastructural transcontextual syndrome. Yet, it remains an exacerbating force through the *structuring* and *pacemaking* role it takes in respect to the life of the distributed community. This is what I elaborate upon in the next section.

5.2 DATA-INTENSIVENESS AND USE COMPLEXITY

Several poignant quotes showed us the reflexive recollection that ensued when the project of distributed shared practice faltered. This recollection was radical to the point of putting into question what exactly the data-intensive infrastructure creates. Informants touched on the limits of 'datafication' (*"there are things that will be immeasurable in the system"*), their consequence on the possibility of sharing evaluation criteria (*"what is the measure of worse?"*) and the inexorable distance that exhaustive, universal, commensuration processes of data collection, aggregation and computation impose between the experiencer (the patient user) and the observer (the data re-users) (*"what we really learned was that what we meant by a condition was something very different from the medical definition of a condition"*). Crucially, these reflections touched on questions of value, asking how to know what is more valuable for the patient user. The informants backtracked to root elements of a shared social order – the criteria of worth: "*We should be able to think of conditions in multiple domains […] we originally built communities that overlapped with conditions, and thought they were about the condition. And then when we built a community of conditions, we did not understand what those communities meant.*" If a data-intensive infrastructure can be used to increase the convergence between different actors by (almost paradoxically) perfecting the isolation of situated data practices from one another through an increasingly complex and computationally flexible architecture of data collection and computation, the PLM experience seems to touch the limits of how much divergence data practices can develop from one another. The criss-crossed, multi-sited semantics of digital data intersect in unpredictable ways to make it incredibly difficult and complex to organize for universal and distributed data use that is simultaneously universal and plural.

Crucially, the technology of the data-intensive infrastructure contributes to the complexity of organizing for data reuse at the same time as it is deployed to reduce it. Patterns of data use and value creation are made complex and unpredictable also because, in an intricate data circulation structure, standardized data are used for the construction of the community itself and the *structuring* and *pacemaking* of its life – through techniques of aggregation and computation (Kallinikos & Tempini, 2014; Tempini, 2015). A whole set of computational and visualization features are aimed at making data movable and meaningful to different end-users. Data use is

the community's underlying foundation and pre-requisite for enabling social interaction (Fig. 4 tries to capture the organizing activity in PLM based on data circulation, evaluation and use).
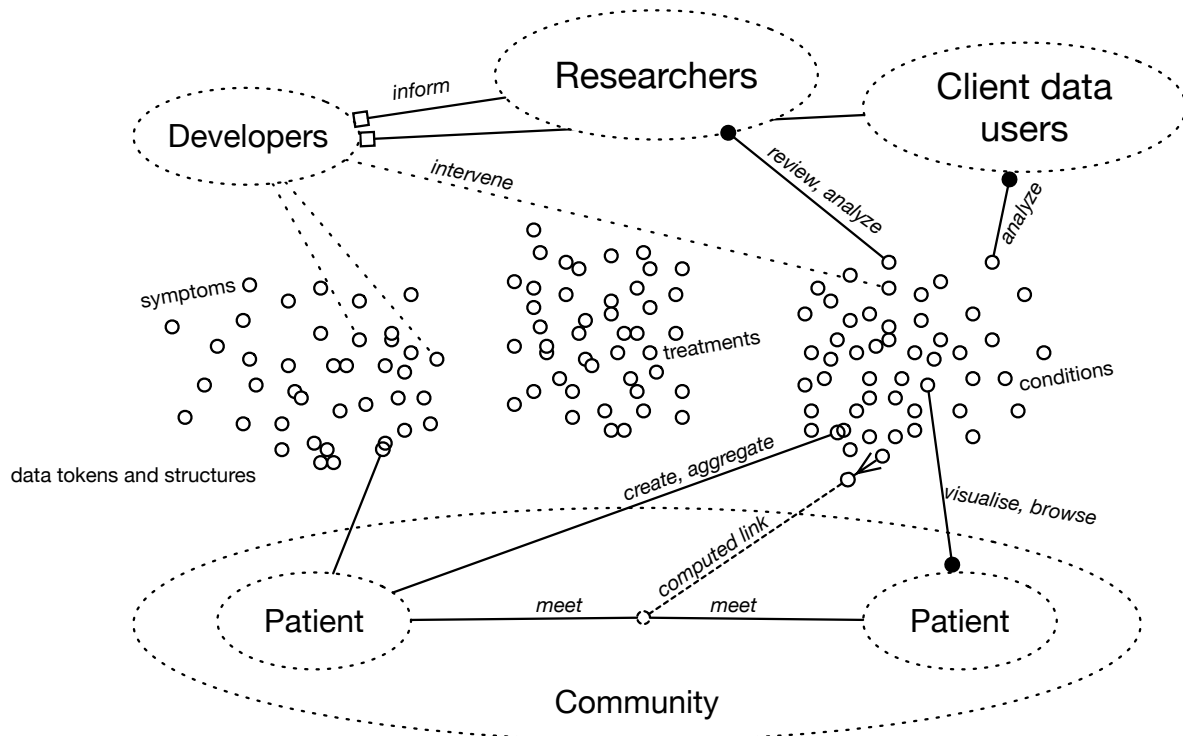
Data-intensive organizing in PLM

Fig. 4: A summary of data-intensive organizing in PLM. Patients request category creation, aggregate their experience data and visualize and browse data aggregates (personal profile and report pages). Business partners analyze data and reports and provide feedback. Researchers review data input and requests, analyze data, and formulate requirements. Developers implement infrastructural innovations (in collaboration with other teams, including Community). Automated computation facilitates patient networking.

As such, the data structures need to meet both these computational requirements and the semantic requirements of the different kinds of community members engaged in practices of data-based value creation and communication. The increasing efficiency and rationalization of the infrastructure of data circulation and calculated interaction that PLM developed over time could eventually deprive the community of reporting conventions, reference points and resources (e.g. the redundancy of duplicates, the context implicit in the user experience in siloed communities) that individual users used to learn and analyze data (see Garfinkel, 2008). This optimization of computable data structures is based on a model of communication efficiency that can act as a straightjacket for people's ability to understand each other (Rawls, 2011). It led to an underestimation of the poorly understood, implicit components of the collaborative environment that attracted patients to the network. These disregarded components then had to be reckoned with in the aftermath of an innovation roll-out, where eventually the context of data generation and use stopped making sense to any of the involved parties and further undermined value creation processes.

New condition communities created in a sequence of minimal research and a few clicks, as the powerful GP features allowed, lacked the basis for sustainable, multidimensional value creation across all relevant dimensions (individual, community, business, scientific). The universal

data-intensive infrastructure had been decoupled from existing understandings of community: as the informant put it, it was now a "community of conditions".

In a data-intensive infrastructure where every user is data consumer and producer at once, the underlying calculating, interactive technology further exacerbated these challenges of *prosumer plurality*, whereby data are produced under incredibly diverse circumstances and motivations and yet must remain amenable to computational analysis to find and draw equations between records originating in everyday human activity. The technology aimed to support community value creation and social interaction through both an interactive computational environment that required standardized, computable data, and the unboundedness of unstructured, redundant, adaptable virtual spaces and features of communication and self-representation. But its short loops of data collection, computation and reuse, whereby newly generated data are immediately computed and recirculated across the network, eventually further highlighted the ensuing gulf between diverging practices of data generation and use, and the gaps in value creation capability therefore emerging.

5.3 INFRASTRUCTURAL INNOVATION AS A CYCLE OF SOCIAL ORDER DISRUPTIONS

The progression of this discussion, which has reflected on the vagaries of data circulation, use, reuse and data-intensive infrastructure development through their relationship to observed outcomes (and conflicts) of value creation, leads me to try and pull these elements further together to build a conclusive point. The aim is to capture how the activity of infrastructural development and innovation for value creation could be so easily frustrated and lead to unexpected outcomes, and to provide a model of data-intensive infrastructural innovation.

Through the discussion, I have argued how the repeated updating and expansion of the web-based data-intensive infrastructure, exercised with a view to gradually improve and refine data practices across the network, repeatedly ignited *cycles of value creation disruption*. When an innovation disrupts shared practice, actors need to resituate data use processes in a way that is valuable according to any of the value dimensions they have stake in. Some dimensions of value creation could be enabled while others are hampered or shifted. I have also linked the unpredictable value creation disruptions to the symbolic nature of digital data. Data are continuously circulated, computed and analyzed to be reused multiple times through a complex fabric of situations. The consequences of infrastructural innovation are far-reaching and yet nonlinear.

Against this backdrop, it becomes possible to see how the development of a data-intensive infrastructure is better understood as a series of multipolar, multistage processes, rather than as a regular alternation between developers and an indistinct group of users. While technological innovation has often been understood as a bipolar cycle between technology standardization and use (e.g. Boland, 2016; Bowker & Star, 1999; Star, 2010; Vargo & Lusch, 2004; Feenberg, 2010), I suggest that a data-intensive infrastructure should be understood as multipolar, and that we need to break down the customarily monolithic phase of use in multiple situations. In Fig. 5, I try to capture this working concept of infrastructural innovation in the most economic, minimally viable model that yet fits the empirical case I present. It necessitates at least two different kinds of data-use situations to follow technology standardization activities. These are *data use and generation* and *data reuse and analysis*.

The inception of an innovation's cycle follows earlier models. Standardization (of templates, analytics features and data structures) implemented by infrastructural innovations is aimed at extending and re-articulating data structures and functionalities, in expectation of certain value creation outcomes to be distributed across different actors. These changes are distributed to the user network to shape generation and use of new data. As made familiar by the broad standardization literature, users accommodate the innovation in myriad different situations. In PLM, we have seen how, other than simply opening new spaces for experimentation and tinkering, the innovation can also disrupt existing data practice and data-based value creation. In this second phase, use innovations are often developed that were not envisioned by the developers (e.g. the symptoms-conditions workaround) and which can re-articulate meanings and shift data-based value creation.

But, due to the data's circulation, computation and reuse (again and elsewhere) through the data-intensive infrastructure and in the form of web analytics or compiled datasets, the combined effects of infrastructural and use innovations in turn reverberate to a second kind of situation: that of data reuse by other actors. Broken reporting conventions and unexpected workarounds can shape these users' data reuse in ways that were not expected and are sometimes unsustainable. The multipolar loop I have just described is a composite loop of data standardization, generation and use, and reuse, where at least two different kinds of data use situations (and obviously, many more individual instances) shape standardization and are shaped by it, and one shapes the other. Data-based value creation processes are repeatedly disrupted by the effects of infrastructural innovation, and can eventually require the recovery of value creation capability when effects are disabling rather than enabling. The empirical material documented how the observation of an innovation's impact, from the perspective of the infrastructure developers, entailed understanding its unfolding through both data use and data reuse situations, and required focusing on the enablement and sustainment of multidimensional value creation as the core of the reflection. This two-fold observation was a starting point to inform the formulation of the requirements to be developed in the following iteration. This model should make it easier to understand why consequences of infrastructural innovations were often unpredictable. The infrastructure development proposition, which I here summarize as "*A in S1 intervenes to change infrastructure I, to change data-generation behavior of B in S2, to enable data use behaviors by both C in S3 and B in S2 of data about B in S2*", is open to multiple interferences. The picture is more complex and rich of interdependencies than a bipartite user-developer relationship would allow us to understand.

At any rate, this working model I present follows earlier ones in depicting development activity as requiring continual adjustment and, as such, as cyclical if seen at a certain resolution. This remains, as I anticipated, an economic model. Recalling the argument about infrastructural evolution as nonlinear, the reader should not be confused by the only apparent logical (and not geometrical) linearity of the circular shape. The effects of innovation, in different situations, on the four dimensions of value creation could not be given pictorial representation without making the figure illegible, yet they are arguably not cyclical. The presence of different requirement vectors converging to inform successive development efforts should further dispel the doubt regarding the linearity of this model.
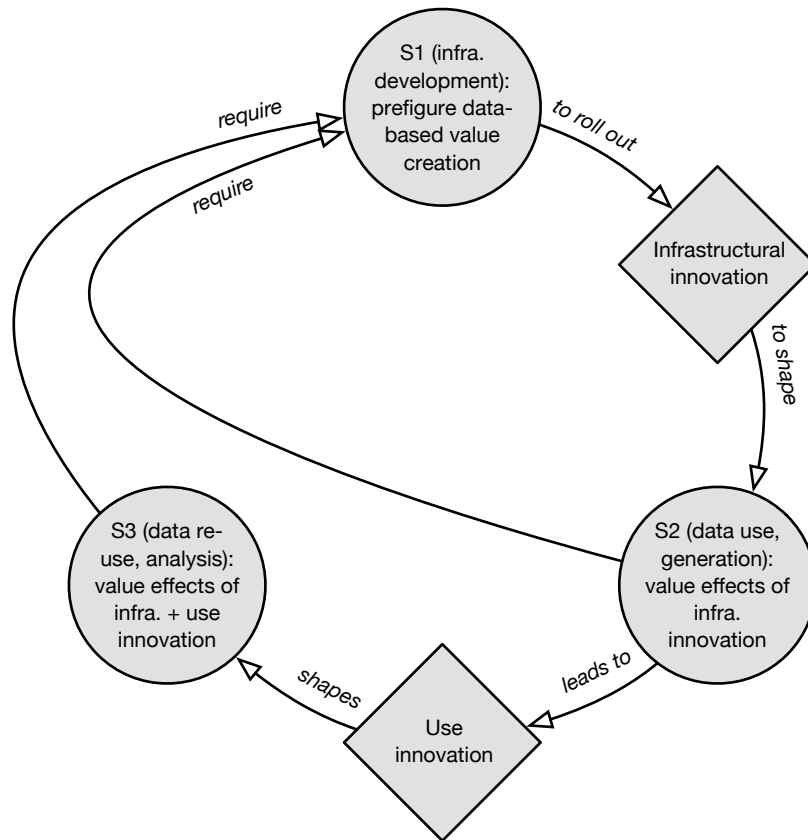
Fig. 5: Conducted from the developers' situation (S1), the development of data structures, features and templates leads to the roll-out of infrastructural innovation (extending and re-articulating data structures and relationships) to shape the (patient) user situation (S2) of data use and new data generation. Eventually, it can upset data practices and value creation in S2. The accommodation of the technology (re-articulating value creation processes) leads to use innovations that combine with infrastructural innovations to shape other users' (researchers, analysts, customers, other patients) data uses in their respective situations (S3), eventually upsetting, in turn, data practices and value creation intersections in S3, and requiring further accommodation and intervention through the formulation of new requirements and reactive change in development plans.

6. CONCLUSIONS

In this paper, I have presented a study of a data-intensive infrastructure. Notably, what is shared in this kind of infrastructure are first and foremost digital data, rather than applications, tools or other resources to operate on them. The findings that I have presented are the result of an original combination of an empirical focus on data-based organizing and a theoretical framing centered on relating infrastructure evolution to data-based value creation, understood as multidimensional and situated. The findings demonstrate that we cannot understand value and valuation if we do not attend specific practices: different kinds of value creation require different sets of engagements with data. Instead of taking notions of data and value for granted, we need to understand how they are constituted, and interrelated, in practice and in situ (Dussauge, Helgesson & Lee, 2015; Leonelli, 2016; Stark, 2009).

In the discussion, I have tried to demonstrate that infrastructural innovations introduced to a complex fabric of data use and reuse processes often have unpredictable consequences, as innovations have different effects in different situations. This complexity, I argued, is connected

to the symbolic nature of digital data (Kallinikos, Ekbia & Nardi, 2015) and the social practices that construct the whole of their semantics – social order, moral commitments, objects, relations.

The implications of this argument are that the development of the infrastructure needs to overcome unexpected contingencies and trade-offs, sometimes critically important ones, and that the creation of different kinds of value, for different actors and from data circulating in myriad of different situations, is not assured – it unfolds along nonlinear and sometimes regressive trajectories. Different actors are simultaneously participating in the infrastructure with different interests and moral commitments to different (shared) dimensions of value. A development strategy that emphasizes increasingly nimble configurations and recombinations of data structures and code can shift over time to a kind of "reactive community development" that is oriented towards the reduction of the uncertainty generated by the data infrastructure's increasing complexity (cfr. Mathiassen & Stage, 1990) and which can cause cycles of value creation disruption. To conclude with a summative representation of these dynamics, I have offered a model of the multi-situational value consequences of data-centered infrastructural innovations. These insights might be useful to practitioners who are able to resituate them in the context of their projects and manage to keep from frictions ensuing in the convergence of an extreme diversity of users and reporting practices, and computational approaches to data reuse that are inevitably based on reduction.

With its empirical focus and theoretical framing, the paper aims to make an original contribution to the literature on (data-intensive) infrastructures. As I argued in the introduction and discussion sections, some of the issues summarized above can be traced to precursors such as problems of adaptability, flexibility and adoption (Hanseth & Lyytinen, 2010; Henfridsson & Bygstad, 2013; Star & Ruhleder, 1996; Tilson et al., 2010). Hanseth and Lyytinen (2010), for instance, have emphasized infrastructure evolution as a nonlinear, emergent process with unpredictable outcomes. This literature has opened a space to ask questions on the distributed creation of multiple kinds of value. These questions indeed yield very complementary results. While one can liken the evolution of information infrastructures to that of complex adaptive systems from a perspective focused on the sharing and distribution of capabilities (Hanseth & Lyytinen, 2010), a second perspective focused on the interdependencies and convergences between multiple, incommensurable and enacted notions of worth in complex contemporary organizing can draw the same comparison (Stark, 2009:22). This paper tries to make a contribution towards the convergence between these perspectives. More work is, of course, needed to flesh out the compatibilities between these perspectives, which might be linked to the issue of generativity. But I have largely exhausted the space available here (and probably the sympathy of the reader with it). This will be a concern for future work.

Still, by deploying a multiple value creation framework to interpret shifting data practices as the observer follows the data across situations, the paper uncovers the dynamics implied in translating complex data structures across an immense, distributed, criss-crossed fabric of interdependent situations, with the resulting theoretical contribution that I just summarized above. In this way, this study expands from the more common focus, in information infrastructure studies, on the sharing of capabilities and resources, the adoption of computing technology, and the openness of design. It encourages information infrastructure scholars to study problems of data valuation, use and reuse. It also provides findings that integrate with similar, recent work on value in online communities (Barrett, Oborn & Orlikowski, 2016), while

showing nonlinear, contingent dynamics that warn against eventual interpretations of development for value creation as a kind of linear accretion trajectory. Some convergences between value dimensions can be destructive of value creation capabilities, I have suggested. The porousness and boundlessness of the social media organization means that challenges, complexities and uncertainties that are found outside the firm can be found inside it. This makes it dramatically difficult to create value from data, for all the parties involved, at the same time. Avoiding the assumption that the exploitation of data-based value creation processes is linear and unilateral, the paper empirically demonstrates how, despite the network dependency on user participation, issues of control and management of the infrastructure are key (Andrejevic, 2015).

Of course, there are limitations to the conclusions one can draw from a single-site qualitative study. It does not, of course, identify infallible regularities. The strength of studying limited and not mainstream cases is, at the same time, a weakness. Here, it is not easy to tell what specific, innovative forms of infrastructure development will be repeated in other examples, and it can take us towards discussing specific issues that have little relevance elsewhere. Nonetheless, I suggest that the multidimensional framework of value creation that I have developed here is likely to be applicable to other SNS projects, where data generated by members are central to the value creation activity of both network members and other organizational actors (developers, analysts, clients). Notable examples, such as Facebook, Twitter and LinkedIn, are all calculated interactive environments replete with data-intensive features and supported by business models centered on data reuse. They are obvious candidates. Even if the specific set of value dimensions of this framework (namely individual, community, business and scientific value) were not perfectly applicable, at least its multidimensional structure of value dimensions as defined through concepts of practice might be a valuable starting point for new analyses.

As I pointed out in the methodology, the empirical narrative has been constructed through the evidence, accounts and institutional memory of the developing organization. The perspectives of patient users have been constructed through indirect inference. It is also left to future research to continue this history by studying how PLM reacted to these learning episodes. The narrative tells how the company embarked on a revision of its user experience strategy, but its reaction is yet to be covered. Still, I have aimed to explore a new terrain through an original theoretical approach. The hope is that other studies will engage with the ideas I presented here, and test them to help refine an understanding of data-based value creation dynamics in a time of unprecedented circulation of, and experimentation with, digital data. This is one of the tasks ahead, nonetheless, if we want to continue and understand the social relevance of pervasive digitalization, data analytics and big data.

REFERENCES

Aaltonen, A., & Tempini, N. (2014). Everything counts in large amounts: a critical realist case study on data-based production. *Journal of Information Technology*, *29*(1), 97–110. https://doi.org/10.1057/jit.2013.29

Aaltonen, A., & Lanzara, G. F. (2015). Building Governance Capability in Online Social Production: Insights from Wikipedia. *Organization Studies* 36(12) 1649-1673.

Andrejevic, M. (2015). Personal Data: Blind Spot of the "Affective Law of Value"? *The Information Society*, *31*(1), 5–12. https://doi.org/10.1080/01972243.2015.977625

Ankeny, R. A., & Leonelli, S. (2015). Valuing Data in Postgenomic Biology: How Data Donation and Curation Practices Challenge the Scientific Publication System. In *Postgenomics* (pp. 126–149). London: Duke University Press.

Avgerou, C. (2013). Social Mechanisms for Causal Explanation in Social Theory Based IS Research. *Journal of the Association for Information Systems* 14(8) 399–419.

Arnott-Smith, C., & Wicks, P. (2008). PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy. *AMIA Annual Symposium Proceedings*, *2008*, 682–686.

Barrett, M., Oborn, E., & Orlikowski, W. (2016). Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. *Information Systems Research*, *27*(4), 704–723. https://doi.org/10.1287/isre.2016.0648

Bateson, G. (1972). *Steps to an Ecology of Mind*. London: University of Chicago Press.

Benkler, Y. (2007). *The wealth of networks: How social production transforms markets and freedom*, Yale University Press, New Haven & London.

Boland, D. (2016). The Concept of Boundary Objects and the Reshaping of Research in Management and Organization Studies. In G. C. Bowker, S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond. Working with Leigh Star* (pp. 229–237). Cambridge, MA: MIT Press.

Bolin, G. (2009). Symbolic Production and Value in Media Industries. *Journal of Cultural Economy* 2(3) 345–361.

Boltanski, L., & Chiapello, E. (2005). The New Spirit of Capitalism. *International Journal of Politics, Culture, and Society*, *18*(3–4), 161–188. https://doi.org/10.1007/s10767-006-9006-9

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton, NJ: Princeton University Press.

Bowker, G. C. (2013). Data Flakes: An Afterword to "Raw Data" Is an Oxymoron. *"Raw data" is an oxymoron*, L. Gitelman (ed.), 167–172.

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*, MIT Press, London.

boyd, d. m., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13(1) 210–230.

Cheney-Lippold, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society* 28(6) 164–181.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*, Sage, London.

Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. *MIS Quarterly Executive* 9(4) 243–259.

van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*, Oxford University Press, New York, NY.

Dussauge, I., Helgesson, C.-F., & Lee, F. (Eds.). (2015). *Value Practices in the Life Sciences and Medicine* (First Edition). Oxford, United Kingdom: Oxford University Press.

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Technical Report (Working Paper).

Ekbia, H. R. (2016). Digital inclusion and social exclusion: The political economy of value in a networked world. *The Information Society*, *32*(3), 165–175. https://doi.org/10.1080/01972243.2016.1153009

Ekbia, H., & Nardi, B. (2014). Heteromation and its (dis)contents: The invisible division of labor between humans and machines. *First Monday*, *19*(6). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/5331

Ekbia, H., & Nardi, B. (2015). The Political Economy of Computing: The Elephant in the HCI Room. *Interactions*, 22(6), 46–49. http://doi.org/10.1145/2832117

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge Collaboration in Online Communities. *Organization Science* 22(5) 1224–1239.

Feenberg, A. (2010). *Between Reason and Experience: Essays in Technology and Modernity*, MIT Press, Cambridge, MA.

Feenberg, A. L., Licht, J. M., Kane, K. P., Moran, K., & Smith, R. A. (1996). The online patient meeting. *Journal of the Neurological Sciences*Amyotrophic Lateral Sclerosis/Motor Neurone Disease 139, Supplement 129–131.

Floridi, L. (2012). Big Data and Their Epistemological Challenge. *Philosophy & Technology* 25(4) 435–437.

Foucault, M. (1989). *Archaeology of knowledge*, Routledge, London.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Garfinkel, H. (2008). *Toward a sociological theory of information / Harold Garfinkel ; edited and introduced by Anne Warfield Rawls.* Boulder, Colo.: Paradigm Publishers.

Gerlitz, C., & Helmond, A. (2013). The Like economy: Social buttons and the data-intensive web. *New Media & Society* 15(8) 1348–1365.

Gillespie, T. (2012). The Relevance of Algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.). Cambridge, MA: MIT Press.

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: the case of building internet. *Journal of Information Technology*, *25*(1), 1–19. https://doi.org/10.1057/jit.2009.19

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing Information Infrastructure: The Tension Between Standardization and Flexibility. *Science, Technology & Human Values*, *21*(4), 407–426. https://doi.org/10.1177/016224399602100402

Henfridsson, O., & Bygstad, B. (2013). The generative mechanisms of digital infrastructure evolution. *MIS Quarterly*, (1), 1–50.

Heaton, L., & Proulx, S. (2015). Paradoxical Empowerment: Immaterial Labor Translated in a Web of Affective Connections. *The Information Society*, *31*(1), 28–35. https://doi.org/10.1080/01972243.2015.977629

Helgesson, Claes, C.-F., & Krafve, L. J. (2015). Data transfer, values, and the holding together of clinical registry networks. In I. Dussauge, C.-F. Helgesson, & F. Lee (Eds.), *Value Practices in the Life Sciences and Medicine* (First Edition, pp. 224–246). Oxford, United Kingdom: Oxford University Press.

Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*, Crown Business, New York.

Jin, D. Y., & Feenberg, A. (2015). Commodity and Community in Social Networking: Marx and the Monetization of User-Generated Content. *The Information Society* 31(1) 52–60.

Josefsson, U., & Hanseth, O. (2000). Patient's Use of Medical Information on the Internet: Opportunities and Challenges. *Proceedings of IRIS 23.*

Kallinikos, J., & Tempini, N. (2014). Patient Data as Medical Facts: Social Media Practices as a Foundation for Medical Knowledge Creation. *Information Systems Research*, *25*(4), 817–833. https://doi.org/10.1287/isre.2014.0544

Kallinikos, J., Ekbia, H., & Nardi, B. (2015). Regimes of Information and the Paradox of Embeddedness: An Introduction. *The Information Society* 31(2) 101–105.

Kane, G. C. (2015). Enterprise Social Media: Current Capabilities and Future Possibilities. *MIS Quarterly Executive* 14(1) 1–16.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, *111*(29), 10779.

Leonardi, P. M. (2014). Social Media, Knowledge Sharing, and Innovation: Toward a Theory of Communication Visibility. *Information Systems Research* 25(4) 796–816.

Leonelli, S. (2016). Data-Centric Biology: A Philosophical Study. Chicago, IL: University of Chicago Press.

Leonelli, S. (2014). Data Interpretation in the Digital Age. *Perspectives on Science.*

Majchrzak, A., Faraj, S., Kane, G. C., & Azad, B. (2013). The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing. *Journal of Computer-Mediated Communication* 19(1) 38–55.

Manyika, J., Chui, M., Bughin, J., Brown, B., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity* (pp. 156–156). McKinsey Global Institute. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Big+data+:+The+next+frontier+for+innovation+,+competition+,+and+productivity#0

Marton, A. (2013). Purposive Selection and the Quality of Qualitative IS Research. *Thirty Fourth International Conference on Information Systems* ICIS.

Mathiassen, L., & Stage, J. (1990). The principle of limited reduction in software design. *Information Technology & People*, *6*(2/3), 171–185. https://doi.org/10.1108/EUM0000000003550

Mingers, J. (2004). Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and Organization* 14(2) 87–103.

Negri, A. (1999). Value and affect. *Boundary 2*, *26*(2), 77–88.

Pollock, N., & Williams, R. (2009). *Software and Organizations: the biography of the enterprise-wide system or how SAP conquered the world*, Routledge, London.

Rajan, K. S., & Leonelli, S. (2013). Introduction: Biomedical Trans-actions, Postgenomics, and Knowledge/Value. *Public Culture* 25(3 71) 463–475.

Rawls, A. W. (2011). Wittgenstein, Durkheim, Garfinkel and Winch: Constitutive Orders of Sensemaking. *Journal for the Theory of Social Behaviour*, *41*(4), 396–418. https://doi.org/10.1111/j.1468-5914.2011.00471.x

Rawls, A. W., & Mann, D. (2015). Getting Information Systems to Interact: The Social Fact Character of "Object" Clarity as a Factor in Designing Information Systems. *The Information Society*, *31*(2), 175–192. https://doi.org/10.1080/01972243.2015.998106

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

Rigi, J., & Prey, R. (2015). Value, Rent, and the Political Economy of Social Media. *The Information Society* 31(5) 392–406.

Sayer, A. (2000). *Realism and Social Science*, Sage, London.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*, Penguin Press, London.

Shirky, C. (2010). *Cognitive Surplus*, Penguin Press, London.

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology & Human Values*, *35*(5), 601–617. https://doi.org/10.1177/0162243910377624

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Stark, D. (2000). *For a sociology of worth* (Working Paper) (p. 11). Center on Organizational Innovation, Columbia University and Santa Fe Institute. Retrieved from http://www.coi.columbia.edu/pdf/stark_fsw.pdf

Stark, D. (2009). *The sense of dissonance: accounts of worth in economic life*. Princeton: Princeton University Press.

Tempini, N. (2014). *Governing Social Media: Organising information production and sociality through open, distributed and data-based systems* (Doctoral Dissertation). London School of Economics and Political Science, London. Retrieved from http://etheses.lse.ac.uk/id/eprint/1026

Tempini, N. (2015). Governing PatientsLikeMe: information production and research through an open, distributed and data-based social media network. *The Information Society*, *31*(2), 193–211. https://doi.org/10.1080/01972243.2015.998108

Tilson, D., Lyytinen, K., & Sorensen, C. (2010). Research Commentary--Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research*, *21*(4), 748–759. https://doi.org/10.1287/isre.1100.0318

Treem, J. W., & Leonardi, P. M. (2012). Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Communication Yearbook* 36 143–189.

Turner, M. R., Wicks, P., Brownstein, C. A., Massagli, M. P., Toronjo, M., Talbot, K., & Al-Chalabi, A. (2011). Concordance between site of onset and limb dominance in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* 82(8) 853–854.

Van Maanen, J. (1979). The Fact of Fiction in Organizational Ethnography. *Administrative Science Quarterly* 24(4) 539–550.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, *68*(January), 1–17.

Wicks, P. (2007). Excessive yawning is common in the bulbar-onset form of ALS. *Acta Psychiatrica Scandinavica* 116(1) 76–76.

Wicks, P., & Frost, J. (2008). ALS patients request more information about cognitive symptoms. *European Journal of Neurology* 15(5) 497–500.

Wicks, P., & MacPhee, G. J. (2009). Pathological gambling amongst Parkinson's disease and ALS patients in an online community (PatientsLikeMe.com). *Movement Disorders* 24(7) 1085–1088.

Wicks, P., Massagli, M. P., Frost, J., Brownstein, C., Okun, S., Vaughan, T. E., Bradley, R., & Heywood, J. (2010). Sharing Health Data for Better Outcomes on PatientsLikeMe. *Journal of Medical Internet Research*, *12*(2), e19. http://doi.org/http://dx.doi.org/10.2196%2Fjmir.1549.

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(5) 411–414.

Wynn, D., & Williams, C. K. (2012). Principles for Conducting Critical Realist Case Study Research in Information Systems. *MIS Quarterly* 36(3) 787–810.

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*. New Haven: Yale University Press.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1) 75–89.