Journal of Information Systems and Technology Management – Jistem USP
Vol. 17, 2020, e202017008
ISSN online: 1807-1775
DOI: 10.4301/S1807-1775202017008 Original Article

THE POLITICS OF DATA PORTALS IN INTER- AND TRANSDISCIPLINARY RESEARCH

Maximiliano Vila Seoane[1] https://orcid.org/0000-0002-0134-7714
Anna-Katharina Hornidge[2,3] https://orcid.org/0000-0002-9599-4348

[1]School of Politics and Government, National University of San Martín, Buenos Aires, Argentina

[2]German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE), Germany

ABSTRACT

In this paper, we argue that the ongoing processes of datafication and dataism are constraining initiatives to construct open data portals contributing to inter- and transdisciplinary research. The former overvalues quantitative data, whereas the latter reinforces the belief that 'raw data' is neutral and apolitical, ignoring how data is processed. Based on the case study of an open data portal being developed at an inter- and transdisciplinary research institute, we argue that datafication and dataism are highly problematic trends, because they marginalize qualitative data employed in critical, constructivist, and other interpretive methods, thereby limiting the possibility of complementing and extending each other. Nonetheless, we also maintain that these trends are not technologically determined but are modifiable, based on the design of data portals. Accordingly, we conclude by offering suggestions for constructing data portals, such as opening up the design process and democratizing standards.

Keywords: Datafication, dataism, transdisciplinary research, data portal, standards

Manuscript first received: 2019-05-02. Manuscript accepted: 2020-09-01

Address for correspondence:
Maximiliano Vila Seoane. School of Politics and Government, National University of San Martín, Av. 25 de Mayo 1021, 1650 San Martín, Buenos Aires, Argentina. Email: mvila@unsam.edu.ar
Anna-Katharina Hornidge. German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE), Germany. Email: anna-katharina.hornidge@leibniz-zmt.de

CC BY

INTRODUCTION: DATAFICATION AND DATAISM

The ongoing data revolution has been led by an increasing process of datafication (Lycett, 2013; Mayer-Schönberger & Cukier, 2013), which includes the digitization of anything we can quantify such that it can be analyzed for various purposes, such as value creation or surveillance. This trend has affected many aspects of contemporary society, such as science, whose advancement has historically been heavily dependent on data. Yet, this growing process of data accumulation is not without its challenges, being the most evident one the increasing concentration of data in the hands of a few private corporations, like US-based Internet platforms. By contrast, the open data movement emerged by striving for open data, in opposition to the prevailing trend of data privatization (International Open Data Charter, 2015, p. 1). A growing list of organizations have started to build the infrastructure to achieve the movement's goals, such as open data portals, which are repositories to collect, systematize, and disseminate data free of charge from different disciplines and countries to the wider public.

Open data portals initiatives are particularly important to inter- and transdisciplinary research, because, in theory, researchers believe that the free access to diverse types of data will allow them to make creative new discoveries that have not thus far been possible (Science International, 2015). In practice, open data initiatives have been facing several challenges that undermine their potential to advance inter- and transdisciplinary research (Kim & Adler, 2015) that contributes to sustainable development. In particular, this paper investigates how the ongoing process of datafication, paradoxically, undermines the aim of inter- and transdisciplinary research that many research organizations strive for by reinforcing old and well-known disciplinary divisions (Redclift, 1998; Snow, 2012; Strang, 2009). This is an important hurdle to explore, since resource limitations and doubts with regard to data authorship are not that hard to address in the construction of open data portals once funds and clear rules are established. In contrast, datafication is an ongoing and pervasive process that might not be entirely under the control of just one organization.

However, the academic literature has not scrutinized the hurdles posed by the ongoing process of datafication, such as the overvaluation of quantitative over qualitative data. In this paper, we shall explore this question based on a qualitative assessment of a data portal initiative of an interdisciplinary development research institute in Germany. We selected this case because it concerns an organization that has been conducting inter- and transdisciplinary research projects for more than 15 years and has only recently begun digitizing all its data in an open data portal. Besides, we follow Van Dijck's (2014, p. 198) critical perspective on datafication, who asserts it goes together with dataism, an ideology that believes in the objectivity of data, questioning neither how nor by whom it has been obtained. Consequently, the thrill of data collection for the purpose of analysis might ignore, on the one hand, the processes of data creation, and might, on the other hand, undermine other research methodologies that are not so easily given to quantification. Indeed, the analyzed case study illustrates the specific challenges that datafication and dataism pose for the future of inter- and transdisciplinary research, reinforcing existing micro-politics between the different actors at the institute. We shall argue that, unless challenged, these processes will continue to favor certain disciplines as well as certain types of research, understanding, and thus 'knowing' that is more oriented toward quantitative data versus those that conduct qualitative ethnographic research, paving the way for what Carrigan (2015) has described as a 'methodological genocide'. Although this expression might sound hyperbolic, the trend does have quite worrisome implications for the ideal of a more holistic system- and process-focused understanding that inter- and transdisciplinary research aims for. This is because critical,

CC BY

constructivist, and other interpretive research approaches would seriously lose prestige and relevance in the process.

The rest of the article proceeds as follows: First, we introduce key concepts related to technology and the relations between open data in inter- and transdisciplinary research. Second, we explain the methods we employed. Third, we examine the case study of an open data portal initiative in order to tease out the main drivers of datafication and dataism. Finally, we discuss the main challenges we identified and we provide suggestions to overcome them.

THE SOCIAL CONSTRUCTION OF TECHNOLOGY

How do open data portals contribute to inter- and transdisciplinary research? After defining inter- and transdisciplinary research, we will answer this question by introducing some concepts from Science and Technology Studies (STS), including boundary work, materialization through standardization, and a constructivist perspective on technology development, which will allow us to think about how to design open data portals to counterbalance the problems caused by datafication and dataism.

There are multiple definitions of inter- and transdisciplinary research (Klein, 2017). A shared feature is to try to surpass the silos of specialized disciplinary investigations by the interaction and integration of knowledge from different disciplines through the construction of models, conceptual frameworks, and other tools (Pohl, 2011). Yet, the degree of interdisciplinary research could vary considerably, for instance, in terms of the closeness between disciplines or whether interdisciplinary takes place at the theoretical or methodological level (Klein, 2017). Similarly, transdisciplinary research aims to integrate as much knowledge as possible and produce overarching frameworks that cover several disciplines (Klein, 2017). By contrast to interdisciplinary research, transdisciplinary research is a demand-driven approach aiming to solve complex societal problems (Jahn, Bergmann, & Keil, 2012; Klein, 2017), and is frequently understood as the final stage of a process integrating disciplinary, multidisciplinary, and interdisciplinary research practices (Pohl, 2011). In effect, it also includes other actors apart from scientists in the knowledge production process. Although these two research approaches are not without their hurdles (Jahn et al., 2012; Petts, Owens, & Bulkeley, 2008; Rosendahl, Zanella, Rist, & Weigelt, 2015), scientists believe that datafication can enable the gathering and systematizing of vast varieties of data sources produced by multiple actors involved in the research process, in turn advancing inter- and transdisciplinary research.

One of the main challenges of inter- and transdisciplinary research is how to enable researchers to cross the boundaries among disciplines. Boundary work, including reflections on boundary concepts, boundary objects, and boundary settings, has been identified as a way to turn this challenge into an opportunity for advanced knowledge production (Mollinga, 2010; Star & Griesemer, 1989; Star, 2010). While boundary concepts are understood very differently among different actor groups, they still allow communication across differences, which is essential to inter- and transdisciplinary research. Likewise, it paves the way for jointly working toward the development of boundary objects, which are tangible objects, such as databases, models, or innovation packages, which bring together different topics of expertise and encourage practice-oriented reflection among communities of practice (Doolin & McLeod, 2012; Mollinga, 2010; Star, 2010), such as researchers. For instance, Star and Griesemer (1989) describe the case of a repository of vertebrates composed of data gathered by varied actors, such as specialists and amateurs, who, despite their differing understandings of their

CC BY

objects of interest and aims for collecting the data, were still able to build a repository for research. The experience of gathering such a diverse dataset would have not been possible if the researchers had worked in a disciplinary and isolated way. Therefore, boundary objects seem to be crucial to break the remaining disciplinary silos that hinder inter- and transdisciplinary research.

If boundary objects are so important for inter- and transdisciplinary research, how can they be created? Star (2010) asserts that standardization processes play a central role in the dynamics of the creation of boundary objects, because they allow the coordination of different actors by creating rules in areas where there might not have been any before. This process involves the creation of standards, which are defined as "*any set of agreed-upon rules for the production of (textual and material) objects"* (Bowker & Star, 1996, p. 2), which are essential to maintain socio-technical systems in operation. For instance, the Internet would not have existed without the presence of an evolving set of standards that allowed disparate artifacts to communicate with each other. However, Star (2010) also warns that, at some point in time, there is a process of standardizing categories that are ill-defined, which leaves some of them out or undefined, potentially leading to the emergence of new boundary objects from those excluded categories. Furthermore, Bowker and Star (1996) stress that there is no rule regarding how to choose one standard over another; rather, they are the product of political decisions, taken in many cases after disputes, which lead to the establishment of one standard over the alternatives. Therefore, standards and the sorts of classifications that they imply cannot be interpreted as being 'objective' or the 'best' possible solution to the problem at hand. Rather, we must be careful to analyze the politics that influenced the formation of the standards and their influence on whether some elements are included and adequately represented or not (Bowker & Star, 1996, p. 4).

The development of any standard and, thus, of any boundary object, is consequently a substantially political process. This is emphasized in STS through an assessment of the socio-political entangledness of technology development (Nahuis & van Lente, 2008). As such, Winner (1980), for instance, outlines how the approximately two hundred low-hanging overpasses to Long Island in New York were designed by urban planners, architects, and municipal decision-makers such that buses, mainly used by social groups less welcome by them in Long Island, would be discouraged to use these parkways. Yet, the design of technology—and especially how it socially unfolds—is used and reinstrumentalized not solely by a few decision-makers, but instead involves various processes of continuous definition and redefinition by numerous groups. Along these lines, we employ a constructivist perspective to technology development, according to which there is a mutual shaping between technical artifacts and humans. Accordingly, technologies have interpretive flexibility, which refers to the various interpretations that social groups with particular interests might attach to them during their design phase (Bijker, Hughes, & Pinch, eds., 2012; Bijker, 2010). The changes in the dynamic interaction of these actors and the choices made in the process of technology design, construction, and use influence the technological outcome; thus, it cannot be understood as a determined process (Williams & Edge, 1996). This is an important insight for our analysis of open data portals, datafication, and dataism, because it suggests that there is no unique way to build a boundary object, nor do the effects of datafication necessarily converge into dataism. Instead, there are multiple ways of building a data portal, but the one that prevails will depend on the different actors that influence and participate in shaping the technical and social choices made during the elaboration process. Consequently, there is always some space for preventing the negative inherent biases that might arise in technology design.

CC BY

METHODS AND CASE SELECTION

In order to unveil the tensions between the aims of conducting inter- and transdisciplinary research and the practice of constructing open data portals, we selected an interdisciplinary development research institute in Germany as a case study for a number of theoretical and empirical reasons. First, it is a well-known international development research institute that aims to cross several boundaries in terms of disciplines, of knowledge production between Global 'North' and 'South,' and of boundaries between 'science' and 'development cooperation.' Second, the institute is part of a German University, which has a strong mandate for capacity development via research by training PhD students, who are mainly part of interdisciplinary projects. Third, the main donor of the institute is the German Ministry of Education and Research, which is part of a broader move in favor of open data, opposing the privatization of research by publishing houses. Fourth, in 2013, the institute started constructing an open data portal to gather data on all its projects. Therefore, we hypothesized that if the institute's researchers are highly socialized in the importance of inter- and transdisciplinary research, the challenges they face will be illustrative for other more recently launched initiatives planning to tread similar paths. Finally, both the authors were researchers at the institute during the creation of the open data portal, which gave us access to inside information on the process of designing and implementing the portal.

With respect to our empirical sources, we first gathered all available data about the institute's portal (i.e. text, images, and notes from all the portal's sections and internal presentations on the project). Second, we conducted twelve in-depth interviews with researchers from different departments, many of whom had participated in the process of designing and implementing the portal. This allowed us to examine both the human actors and the process of constructing the artifact that aims to function as a boundary object. Finally, we inductively analyzed the data with the aim of detecting themes that illustrate the different perceptions about the project and its hurdles, contrasting the perspectives of those who actively participated in its construction.

THE CONSTRUCTION OF AN OPEN DATA PORTAL

The German development research institute we investigated grounds its research projects in the concepts of inter- and transdisciplinarity, intended to foster collaborative knowledge production between disciplines, practitioners, and policy makers. This is an important objective for the institute's large-scale development research projects, which produce substantial research in a wide range of topical fields, which, not long ago, had not been systematically centralized and arranged for easy future retrieval. Hence, the construction of a data portal became a solution to update the institute's organizational tools and practices for knowledge management as well as to better display its research output. Researchers also believed that the data portal would help link the data dispersed out in the 'wild,' facilitating less fieldwork-intensive research projects. Furthermore, they believed it would help collect long-term heterogeneous data gathered through various methods, at different scales (i.e. personal, societal, and environmental), and in different regions, paving the way for the development of new methods of conducting inter- and transdisciplinary research. Among the rhetoric for advancing the portal, we found the usual argument that open access to research data will improve not only research but also development policies more broadly. In effect, researchers at the institute frequently argue that making data open will facilitate resource-scarce researchers in the 'global South' by lowering the costs to access data. Finally, the pervasive argument in favor of making research data open is

CC BY

that this will increase researchers' impact, because others will not only cite their papers but also their data sources, increasing the researchers' citations, which are extremely important in contemporary assessments. Due to this, the institute follows an open access policy to the data published on the portal, while respecting data privacy and authorship.

In sum, the construction of a data portal was understood as an extra step toward advancing inter- and transdisciplinary development research and thus a tool to capture real-life complexity; therefore, it can be understood as a boundary object. However, we shall argue that its construction does not necessarily contribute to this goal, because of the specific micro-politics that exist between the different actors at the institute contributing to the portal and the ongoing datafication and dataism processes underlying its construction. These two processes have led to the introduction of standards in line with quantitative data analysis, neglecting qualitative research, which undermines the aims of inter- and transdisciplinary research.

DISCIPLINES, THE PORTAL, AND DATAFICATION

In this section, we argue that the micro-politics of the different disciplines operating at the studied institute, shaped by the implicit hierarchy of disciplines within development research, is one important explanation for the prevalence of datafication in the construction of the data portal.

At the institute, researchers work in three disciplinary departments: Social Sciences, Economics, and Natural Sciences. At the same time, for several projects, the departments strive to work in tandem in an inter- and transdisciplinary way in order to tackle complex societal problems related to, for example, food, health, and migration, among others. However, these departments do not have the same influence; in effect, the Economics and Natural Sciences departments have been able to secure more financial and human resources for projects than the Social Sciences department. Hence, the project of constructing a data portal is taking place with an asymmetric distribution of resources. In the broader context of development research, this is not surprising. Indeed, the disciplines that attract more attention and funds are those considered more 'valuable' for policy makers, like economists and natural scientists, who tend to enjoy more prestige than social scientists, particularly more than those who criticize and ask inconvenient questions about development that people in higher echelons of power might prefer to sideline.

In this context, Table 1 outlines the main differences between the three departments' disciplinary specializations, data sources used, and the data analysis techniques preferred, which help to explain the influences that shaped the design of the data portal[1]. Starting with the Natural Sciences department, the researchers mainly gather data from different types of sensors and experiments. This allows them to find causal connections that explain why a given problem is occurring, or to model a process of interest. These researchers mainly collect data on biodiversity, health, geographic maps, and renewable energies, among other key research areas of the institute. Second, the researchers in the Economics department primarily work with quantitative data revealing different characteristics of the economic decisions and processes that they investigate. In general, their projects aim to propose new policies addressing pressing issues by showing causal connections between phenomena through econometric methods. Hence, in this department, many researchers accept methodological individualism, which assumes that individuals take rational decisions to achieve their goals, and consider these principles as

1 We constructed the table by grouping the gathered empirical material on each department based on the research paradigms identified by Guba & Lincoln (1994). Despite the importance of these categories to outline main researchers' worldviews, it is important not to forget that from an STS perspective, these categories are a construction, and a simplification of the practice that all scientists – even the most positivists ones–follow.

CC BY

universal, sidelining societal or cultural variables (Gilpin, 2001). In this sense, their research practice resembles that of natural scientists, because they strive toward universality in their findings, mainly employing quantitative methods to achieve their goals, though they mostly work with variables that directly involve humans. In terms of philosophies of science, the researchers in these departments fall either in the positivist or postpositivist camps—the former believes that we can map and understand reality out there as it is, whereas the latter maintains that our knowledge of reality can only ever be imperfect and probabilistic (Guba & Lincoln, 1994).

In contrast, the Social Sciences department mainly hosts social scientists who work as per constructivist and critical philosophies of science, who use ethnographic and interpretive methods that emphasize the multiple ways in which the social is constructed and the influence of values in shaping the research process (Guba & Lincoln, 1994). Therefore, most researchers put a lot of emphasis on diverse context-specific and historical conditions that have molded their objects and subjects of study. Accordingly, these researchers mainly employ different qualitative methods, such as archive research, participant observation, and in-depth interviews. Therefore, interview transcripts, lived experiences, and highly detailed pieces of contextual knowledge are the key data sources for these practitioners. Although we outline a division between the qualitative and quantitative approaches, this does not mean that scientists in the Natural Sciences or Economics departments do not use qualitative data, nor does it mean that social scientists do not make models or conduct surveys for statistical analysis. In fact, many try to cross the methodological divides, precisely due to the interdisciplinary goals of the institution. However, the prevailing philosophies of science that divide the disciplines still considerably shape the priority given to these different sources as outlined in Table 1.

Table 1. Philosophies of science by department at the institute and their different disciplines, data sources, and data analysis methods employed.

Philosophy of science	Constructivism, critical theory	Positivism, postpositivism
Departments	Mainly Social Science	Economics and Natural Sciences, though a minority of social scientists uses them as well
Disciplines	Anthropology, Sociology, Human Geography, Political Science, and Area Studies	Economics, Geography, Agriculture, Hydrology, Agronomy, Soil Sciences, and other natural sciences
Data sources	Qualitative, ethnographic data: Archives, interviews, participant observation, and others	Surveys, geographic information, experiments, interviews
Data analysis	Interpretive approaches based on rich contextual historic-political knowledge	Statistical techniques to find correlations, predictions, different types of computer models and simulations, and GIS analysis

Among these actors, natural scientists have taken the lead and are in charge of the construction of the data portal. One of the researchers leading the work on the portal explains their aims as follows:

> "A lot of the examples always pointed out that there is not enough memory in the organization, it does not matter whether for quantitative or qualitative data. Thus, we had the idea of creating at least a neutral database. However, it does not matter how this database looks as long as we have neutral data, with measurements of whatever you have done in the past—temperature, soil moisture, etc. It is neutral. Then comes the interpretation of it [...]. The same data can be interpreted in different ways" (Natural science researcher, 06/11/2015, personal communication).

CC BY

This quote is illustrative of datafication and dataism for two reasons. First, the cited examples are all easily quantifiable variables related to climate change, which so far prevail in the portal. Second, the interviewee believes it is possible to construct a 'neutral' database. This reveals the importance attached to quantitative, 'objective' data, which aligns with the postpositivist belief that our knowledge of reality might not be perfectly known, but we can get closer to it through 'objective' and varied data. Accordingly, these observations indicate that the portal focuses on quantitative data in line with the positivist and postpositivist philosophies of science that prevail in the Economics and Natural Sciences departments at the institute. Indeed, the project heads decided to use GeoNetwork, sophisticated open source software that standardizes the storage, management, and sharing of geographic data, enabling communication between different portals and geographic data sources. Hence, GeoNetwork excels at archiving geographic data, revealing how some quantitative variables change through space, such as the amount of rain or the distribution of populations on certain parameters. From the point of view of the project leaders, selecting this software was an excellent decision, because it eliminated the time needed to code a portal from scratch and gave them access to GeoNetwork's superior functionalities for the data they focus on, such as visually appealing maps and easily accessible quantitative data and metadata. In addition, other organizations use GeoNetwork, thereby allowing for compatibility and exchange among data portals, which is an important objective for the open data movement—to create an ecosystem of portals that challenge the increasing privatization of data.

CHALLENGES

At first sight, the trend toward datafication does not seem very problematic; indeed, it is mainstream, and even researchers of the digital humanities claim that an increase in quantitative data strengthens the potential interpretations of phenomena under study (Heuser & Le-Khac, 2011). However, this perspective is highly problematic if we contrast it with critical and humanistic understandings of the role of data in societies. For instance, Van Dijck (2014) argues that datafication goes hand in hand with a belief in dataism—the assumption that quantitative data is trustworthy or, as one of the pioneers of the studied institute's portal described it, neutral. However, as Gitelman and Jackson (2013, p. 3) point out, there are key features of data that need to be taken into account. First, data does not grow like apples—its creators generate it somehow using some device along with a process, which requires imagination and data engineering skills, and it is not without its limitations and implicit assumptions regarding what the data can 'see' (Baack, 2015; van Dijck & Poell, 2013). Therefore, data is neither neutral nor objective but guided by the specific ontologies (what exists?) and epistemologies (how do we produce knowledge?) selected for data collection (Wendt, 1999). Second, data is not 'objective,' because its generation always requires contextualization and framing, which might become invisible once established. This is evident in the abstract nature of data (both mathematical and linguistic symbols) that require some sort of means of expression, storage, and dissemination. Accordingly, it is inaccurate to assume that 'raw data' will move us closer to the 'truth' through a neutral database. Finally, supporters of open data tend to repeat that everything is quantifiable, but we need to bear in mind that although this might be the case, by no means does this imply that the quantifications will make sense. Only objects and processes that can be divided and combined without qualitative changes can produce meaningful quantifications (Sayer, 2010, p. 119), such as the measurement of weight or temperature. In contrast, many qualitative variables are context dependent, such as attitudes, beliefs, meanings, etc., which can be neither geographically fixed nor

CC BY

socially essentialized, and they are also not universal. Therefore, analysts have to be extremely careful when comparing such data with mathematical techniques, which if carelessly used would only generate meaningless or utterly incorrect results.

Despite these caveats, the initiators of the studied portal believe that qualitative data is technically not a problem, mirroring the widespread belief that it can be handled in the same way as quantitative data so as to become open and replicable, requiring only some changes in the practices of qualitative researchers (Parry & Mauthner, 2004). Besides, researchers working with different methods often express contempt toward one another, as seen in the observation of Boyd and Crawford (2012, p. 667) that there is the '*mistaken belief that qualitative researchers are in the business of interpreting stories and quantitative researchers are in the business of producing facts*'. This observation reveals an underlying hierarchy, wherein researchers employing quantitative data are at the top of what is considered 'good science', because they contribute to its 'advancement' by making their work 'open' and 'replicable'. In contrast, qualitative researchers are much lower down in the hierarchy, in the worst cases perceived as mere storytellers, who contribute little – if at all – to such collective projects.

Not surprisingly, researchers employing constructivist and critical approaches tend to disagree with such caricatures, underlining that they aim to 'understand' complex change processes rather than predominantly 'map' their aspects in a positivist or postpositivist manner. Indeed, in qualitative research sources such as interviews, there is a large amount of contextual data needed to read a transcription. After all, the main instrument of observation is the researcher, who uses field notes, videos, audio, etc., as important means to store and complement the experiences that they have lived through in the situations that they observed and the previous contextual knowledge that they brought to the field. This explains why they might be useless for readers who do not have such background or who know very little of the specific theme under analysis, of the situation in which the interview took place, and of the process of producing the transcription. This obviously contrasts with the reading of quantitative data, whose interpretation requires little contextual knowledge.

This brings us to the central point—the methods of data collection are different in the different disciplines, and the registries play different roles that vary from discipline to discipline and with the respective philosophies of science inspiring their research. Consequently, datafication and its association with dataism and the replicability of research are not equally important for all disciplines. On the one hand, researchers consider qualitative data as useful memory notes of the situation they investigated. Yet, they are aware that it does not serve to codify all the knowledge they captured, which explains why these types of data are hard for outsiders to decode. On the other hand, in the more mathematically oriented disciplines, data sets—though very much processed—are the products on which the posterior analysis is mainly based. Therefore, opening data sources for replication might not be as useful in the interpretive social sciences as in others[2]. Field research, whether in the social or natural sciences, moves into the background. Therefore, unless different standards of the processes are sought that consider these differences in terms of situated knowledge and tacit skills, the aim of quantitatively oriented researchers making data open and replicable seems even harder to attain for those working with critical, constructivist, and interpretive approaches.

Despite these caveats, many researchers of the institute under study believe they are eclectic enough to work with different research approaches and types of data, as if there were no opposing philosophies of science inspiring their practices. However, such illusions are soon shattered when

2 It is also important to bear in mind that replicating the collection of data is rarely possible and only in completely controlled environments, if at all. Indeed, even natural scientists acknowledge that there is a 'reproducibility crisis' of scientific research (Baker, 2016).

CC BY

dissent arises from the more interpretivist researchers. This was what happened at the observed institute, where tensions emerged between researchers employing qualitative versus quantitative methodologies. These tensions are not new but linked to old and ongoing debates between different cultures of constructing knowledge (Snow, 2012), and in particular, between different research methods (Guba & Lincoln, 1994) that still remain a challenge for inter- and transdisciplinary research (Petts et al., 2008; Phoenix et al., 2013; Redclift, 1998). From our point of view, it seems that the trend toward increasing datafication and dataism, unless challenged, will reinforce such tensions and arguments about the incommensurability between different philosophies of science. Although inter- and transdisciplinary research consciously tries to cross barriers between disciplines, in conducting research, it is important not to forget or neglect the differences that separate them.

THE STRUGGLE FOR STANDARDS

An essential element of the construction of any boundary object is a common tendency in communities of practice of tailoring such artifacts to their own specific use (Star & Griesemer, 1989). In principle, the institute's open data portal conforms to this tendency, because its designers built the portal for researchers from all departments and with the clear aim to cross disciplinary divides among the social sciences, economics, and natural sciences, by storing and retrieving data from diverse regions and topics of interest. Eventually, this would expose researchers to data gathered by other disciplines, employing different methods, and increasing the chances of making creative connections among the data stored in the portal, which would in turn contribute to inter- and transdisciplinary research. However, during our fieldwork, we detected that those disciplines employing largely quantitative approaches took the lead in building the database, which resulted in the portal being in line with the problematic processes of datafication and dataism. In effect, we found imbalances in the 'voices' of the different disciplines in the design process of the portal, which has reinforced some standards over others and shaped decisions on the emerging boundary object in the direction of datafication and dataism and not others.

The events unfolded in the following way: During the initial phases of the data portal project, the senior members of the Social Sciences department showed a lack of interest in the project. Consequently, they had no representation during the initial decision-making processes, and most of the decisions were taken by representatives from the other departments. Only after a change in management did the Social Sciences department appoint a representative to shape the portal more in line with their methodological needs; however, by then, the portal pioneers had already taken the decision to use the GeoNetwork platform as the backbone of the project, thereby favoring the quantitative methods employed by the other two departments. Therefore, there was an imbalanced engagement in the original design of the portal, which resulted in a pro-quantitative bias with regard to the standards to be included, sidelining more interpretive, qualitative, and contextual ones. Second, the selection of GeoNetwork for the open data portal interface favored archiving geographic quantitative data, which offers useful standards for quantitative data. Nevertheless, the institute's portal initiators have not—partly because of a lack of interest of the qualitative Social Scientists—adapted the GeoNetwork platform for the needs of storing and disseminating qualitative data. This decision further led to opposition from social scientists employing critical and interpretive methods who found no way to fit their data into the available quantitative standards in the portal. In opposition, quantitative-oriented researchers neglected the problem, arguing that everything is data; thus, technically, it was not a

CC BY

problem to digitize, store, and release the transcripts of interviews and other qualitative sources that researchers in the Social Sciences department gather through qualitative methods. Not surprisingly, neglecting the specificities of qualitative data led to the opposition of social scientists, who had avoided participating in the project or simply gave no useful input for its implementation.

DISCUSSION

The results show considerable disagreements among those deciding the features of the data portal. In this section, we propose two explanations for these differences, which shall pave the way for potential solutions in the design of other open data portals.

First, we have to take into account that those who pioneered the portal and those who at first refused to participate come from distinct social worlds (Star & Griesemer, 1989). On the one hand, the Natural Sciences and Economics departments' managers have a tendency to employ rather positivist and postpositivist research methods that assume the superiority of the quantitative approach to science or simply regard it as the data collecting source best suited for data portal integration. Therefore, it is not difficult to hypothesize that they are not very aware of the subtleties of critical, constructivist, and interpretive qualitative research. Consequently, the portal pioneers studied here were biased toward choosing open source code for the data portal, tailored for geographic and other types of quantitative data, mirroring the mainstream trend in their scientific disciplines. On the other hand, the initial decision of the Social Sciences department was to stay out of the project altogether, because of their conviction that qualitative data, especially personalized data, cannot, for reasons of research ethics as well as due to the limitations of this data being analyzed by a third person, be handled in the same way. Evidently, this decision reduced the possibility to foster multiple forms of scientific understanding, which is the core aim of interdisciplinarity. Therefore, in the first phase of the portal's design, the gap between the mindsets of researchers employing different methods led to a standardization process that limited the capacity of the open data portal to work as a boundary object, because the quantitative approach prevailed.

Second, along the lines of Star (2010), standards often lead to anything that does not fit into a particular category being collectively put into an 'other' category, theoretically leading to those interested in these excluded dimensions to create new boundary objects that include them. Yet, after deciding to get involved in the process, the social scientists did have the opportunity to intervene and change the design of the portal by adding new dimensions to the portal, but they did not do so. We can explain this by the fact that their department's representative specialized in quantitative approaches that are employed by a minority of social science researchers rather than in interpretive ones. Therefore, a complete representation of the department's methodological diversity was missing, mainly that of researchers trained in qualitative, ethnographic approaches, who refused to intervene for two reasons: A) They believed they had very little chance of bargaining with their more quantitatively oriented colleagues from more influential departments. Hence, these researchers, although preferring not to be entirely out of the project, participated only to a limited extent, thereby delegitimizing the initiative. B) They lacked the technical skills to understand the process of creation and operation of data portals, and of the underlying data storage standards, which might have prevented them from realizing that by participating in the design process, they would have had far more opportunities to change the data portal than they might have believed.

CC BY

These causes led to an unfortunate trend in the database's building process, with the process of datafication, rooted in the belief of dataism, prevailing in the first year of the portal's operation. Due to this trend, the interfaces and standards that qualitative researchers required were relegated to the background, mirroring the fears and rejection that social scientists have with regard to the rise of datafication and dataism (Carrigan, 2015; van Dijck, 2014). The data portal at its current stage of development continues to marginalize interpretive and qualitative methods of data collection and analysis and thus widens the chasm between disciplinary approaches at the institute, which, in the long run, weakens the institute's potential for inter- and transdisciplinary research.

Nonetheless, these challenges indicate that at least two types of policies could be taken to mitigate the challenges identified: opening up the design process and democratizing standards.

First, the creation of an open data portal should not be seen as just another side activity that researchers do in their 'free time', because this reduces the amount of resources that they dedicate to such projects. Not surprisingly, researchers might fall into the trap that the available technology is 'objective' and the construction of a portal just another technical job, undermining the importance of the design process that fundamentally shapes the outcome. Given the trend in science toward datafication and dataism, if the design process is not inclusive of the perspectives of all disciplines, it will probably reinforce the prevailing biases favoring quantitative approaches. For these reasons, we share the recommendation in the literature for social scientists to practice an upfront engagement in the initial stages of research and development projects in general (Lowe, Phillipson, & Lee, 2008; Redclift, 1998), and we extend it to the design of boundary objects in particular, such as data portals. In this way, social scientists will avoid the strategy of non-participation that we saw at the studied institute. Otherwise, researchers employing other methods will lose the chance to change these portals' designs, which, once finalized, will make the values and standards of those believing in the objectivity of quantitative data harder to change in the resulting technology.

Second, participation in the design process is not enough unless it paves the way to modifying the standards employed in the technology under design to encourage truly inter- and transdisciplinary research. STS insights teach us that we ought not to treat standards lightly, because their logic is to homogenize; thus, the fewer standards built into portals, the less diversity of data types and standards will the portals be able to handle. Accordingly, if researchers involved in the construction of these portals do not question standards, they might end up reinforcing quantitative research methods, disregarding the methods of other disciplines. In this line, it is important to bear in mind the relevance of dialogue among disciplines in order to learn on different disciplinary data needs and worldviews (O'Rourke & Crowley, 2013; Phoenix et al., 2013). With this in mind, the participatory design process needs to be open enough to elaborate diverse types of standards that different disciplines and users might need, increasing the chances that each type of actor can upload their data based on its specific research practices.

CONCLUSION

Despite the promises of open data portals to contribute to the future of inter- and transdisciplinary research, based on our investigations of the construction of an open data portal by development researchers who were socialized to these research techniques, we found two main hurdles hindering the achievement of this goal. First, the uncritical expansion of standards favoring datafication and dataism processes – the tendency to consider quantitative data as less personalized and more objective

CC BY

than other sources – is a serious risk for the holistic understanding of change processes that is the core goal of inter- and transdisciplinary research. Indeed, these processes favor the methodological and philosophical understanding of the science of the dominant disciplines at the institute, paradoxically neglecting the more critical, constructivist, and interpretive approaches of the Social Sciences department.

Second, we have argued that the expansion of standards in line with datafication and dataism processes is based on preexisting biases, dissents, and hierarchies among disciplines. These preconceptions tend to favor those working with positivist and postpositivist philosophies of science, while there is also great reluctance on the part of social science scholars working with other approaches to engage constructively in data portal design processes for assuring a more advanced interdisciplinary inclusiveness of data portals. If unquestioned, these asymmetries lead to the unbalanced design of technical artifacts, such as the data portal, favoring and reinforcing the negative aspects of datafication and dataism.

Finally, in order to address the challenges identified, we have suggested opening up the design process so that all partners involved in the creation of data portals for inter- and transdisciplinary research have a voice concerning their desired functionalities. In this way, the portals have more chance of materializing standards considering both quantitative and qualitative research approaches. Although these suggestions are not necessarily new, we believe that under the contemporary trend towards datafication it is of the utmost importance to take them seriously into account in order to improve the ongoing designs of new data portals to address contemporary sustainable development challenges.

REFERENCES

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 1–11. https://doi.org/10.1177/2053951715594634

Baker, M. (2016). Is there a reproducibility crisis? *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Bijker, W. E. (2010). How is technology made?-that is the question! *Cambridge Journal of Economics*, *34*(1), 63–76. https://doi.org/10.1093/cje/bep068

Bijker, W. E., Hughes, T. P., & Pinch, T., eds. (2012). *The social construction of technological systems* (anniversary edition). Cambridge, Massachusetts, USA: The MIT Press.

Bowker, G. C., & Star, S. L. (1996). How things (actor-net)work: classification, magic and ubiquity of standards. *Philosophia*, *25*(3-4), 195–220.

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Carrigan, M. (2015). Emma Uprichard: Most big data is social data – the analytics need serious interrogation. Retrieved January 10, 2016, from The Impact Blog. The London School of Economics and Political Science website: http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-data-science-emma-uprichard/

Doolin, B., & McLeod, L. (2012). Sociomateriality and boundary objects in information systems development. *European Journal of Information Systems*, *21*(5), 570–586. https://doi.org/10.1057/ejis.2012.20

CC BY

Gilpin, R. (2001). *Global Political Economy. Understanding the international economic order*. New Jersey, USA: Princeton University Press.

Gitelman, L., & Jackson, V. (2013). Introduction. In *" Raw Data " Is an Oxymoron* (pp. 1–14). Cambridge, Massachusetts, USA: The MIT Press.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln, *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: SAGE.

Heuser, R., & Le-Khac, L. (2011). Learning to read data: bringing out the humanistic in the digital humanities. *Victorian Studies*, *54*(1), 79–86. https://doi.org/10.2979/victorianstudies.54.1.79

International Open Data Charter. (2015). *International Open Data Charter* (p. 7).

Jahn, T., Bergmann, M., & Keil, F. (2012). Transdisciplinarity: between mainstreaming and marginalization. *Ecological Economics*, *79*(C), 1–10.

Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, *35*(4), 408–418. https://doi.org/10.1016/j.ijinfomgt.2015.04.007

Klein, J. T. (2017). Typologies of interdisciplinarity: the boundary work of definition. In R. Frodeman, J. T. Klein, C. Mitcham, & J. B. Holbrook (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 21–34). Oxford University Press.

Lowe, P., Phillipson, J., & Lee, R. P. (2008). Socio-technical innovation for sustainable food chains: roles for social science. *Trends in Food Science & Technology*, *19*(5), 226–233. https://doi.org/10.1016/j.tifs.2007.11.005

Lycett, M. (2013). "Datafication": making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386. https://doi.org/10.1057/ejis.2013.10

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work and think*. New York, USA: Houghton Mifflin Hartcourt Publishing Company.

Mollinga, P. P. (2010). Boundary work and the complexity of natural resources management. *Crop Science*, *50*(1), 1–9. https://doi.org/doi:10.2135/cropsci2009.10.0570

Nahuis, R., & van Lente, H. (2008). Where are the politics? Perspectives on democracy and technology. *Science, Technology & Human Values*, *33*(5), 559–581. https://doi.org/10.1177/0162243907306700

O'Rourke, M., & Crowley, S. J. (2013). *Synthese*, *190*(11), 1937–1954. https://doi.org/10.1007/s11229-012-0175-y

Parry, O., & Mauthner, N. S. (2004). Whose data are they anyway? practical, legal and ethical issues in archiving qualitative research data. *Sociology*, *38*(1), 139–152. https://doi.org/10.1177/0038038504039366

Petts, J., Owens, S., & Bulkeley, H. (2008). Crossing boundaries: interdisciplinarity in the context of urban environments. *Geoforum*, *39*(2), 598–601. https://doi.org/10.1016/j.geoforum.2006.02.008

Phoenix, C., Osborne, N., Redshaw, C., Moran, R., Stahl-Timmins, W., Depledge, M., … Wheeler, B. W. (2013). Paradigmatic Approaches to Studying Environment and Human Health. *Environmental Science and Policy*, *25*(11), 218–228. https://doi.org/10.1016/j.envsci.2012.10.015

Pohl, C. (2011). What is progress in transdisciplinary research? *Futures*, *43*(6), 618–626. https://doi.org/10.1016/j.futures.2011.03.001

Redclift, M. (1998). Interdisciplinary research on the global environment. *Global Environmental Change*, *8*(3), 177–182.

Rosendahl, J., Zanella, M., Rist, S., & Weigelt, J. (2015). Scientists' situated knowledge: strong objectivity in transdisciplinarity. *Futures*, *65*, 17–27. https://doi.org/10.1016/j.futures.2014.10.011

Sayer, A. (2010). *Method in social science. A realist approach.* (Second). Abingdon, Oxon, UK: Routledge.

Science International. (2015). *Open data in a big data world. An international accord.* (p. 15) [International accord]. Paris, France: International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership (IAP).

Snow, C. P. (2012). *The two cultures*. In *Canto Classics*. Cambridge, UK: Cambridge University Press.

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., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: amateurs and professionals in Berkeley's Museum of vertebrate zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001

Strang, V. (2009). Integrating the social and natural sciences in environmental research: a discussion paper. *Environment, Development and Sustainability*, *11*(1), 1–18. https://doi.org/10.1007/s10668-007-9095-2

Van Dijck, J. (2014). Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.

Van Dijck, J., & Poell, T. (2013). Understanding Social media logic. *Media and Communication*, *1*(1), 2–14. https://doi.org/10.17645/mac.v1i1.70

Wendt, A. (1999). *Social Theory of International Politics*. Cambridge, UK: Cambridge university press.

Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy*, *25*, 865–899. https://doi.org/10.1016/0048-7333(96)00885-2

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

CC BY