Proactive Data Activism within the Framework of Communicative Action Theory: The Case of *DataKind*

Ilgar SEYIDOV[1]

Abstract

The current age is also known as the Big Data Age, which refers to the abundance of varied, complex data and information. While the digitalization process has provided online, non-physical social identities and communication platforms for people, the datafication process has offered the use of data in different areas of humanity effectively and salutarily. Since humanitarian issues such as famine, education, health, etc. are increasing dramatically worldwide, activists and organizations have started to benefit from the datafication process for social changes. In the last decade, data activism has become a phenomenon in the scope of digital activism practices. The current study focuses on this new trend of data activism and aims to find out how proactive data activism can be understood within communicative action theory. In this vein, case study analysis was selected as the research method to examine the data activist platform DataKind. Based on the theory, interactive structure, argumentation base, and implementation formation of DataKind were analyzed in a detailed way.

Keywords: Big data, Datafication, Data activism, Communicative Action Theory

Introduction

By the second half of the 20th Century, the development of technology, the Internet, and software systems have eased information processing and dissemination worldwide. In particular, following the 2000s, the easy access to online systems and the extensive use of computers and mobile phones have accelerated the formation and development of the

[1] Doç. Dr. Atılım Üniversitesi, İşletme Fakültesi, Halkla İlişkiler ve Reklamcılık Bölümü, Ankara, Türkiye
E-mail: ilgar.seyidov@atilim.edu.tr   ORCID: 0000-0001-8420-1413

Proactive Data Activism within the Framework of Communicative Action Theory: The Case of DataKind

digitalization process. The digital (online) world has become the essential area and source for various fields of humanity. The concept of *data* has become a buzzword. The abundance, scale, variety, and complexity characteristics rendered it being conceptualized as "big data".

Within the framework of computer terminology, "big data" refers to complex data sets which exceed the memory of standard and software processes (Lewis and Westlund, 2015, p.448). The technical definitions of big data are commonly based on three aspects, which are *volume* (the amount of data), *velocity* (the speed of data processing), and *variety* (multiple sources, formats, or structures for data use) (Moffat-Hannah, 2019, p.457). Facts, symbols, or signals, that are modified, and distributed, and interactive artifacts can be included in big data. To be more precise, big data understanding includes a wide range of topics such as an abundance of digital objects, user-generated online content, data results from datafication, signals of sensors, drones, and such devices, marks, metadata, etc. (Guiterrez, 2018, p.4).

Due to the massive capacity needed to store information, it is not easy to analyze all the data. The main challenges are abundance, variety, timeliness, dynamism, high rationality, and uncertainty. However, those challenges also provide opportunities for scholars to develop more sophisticated, wider-scale scientific models (Kitchin, 2014, p.2-7). On the other hand, big data is also a social and cultural phenomenon that allows the analysis of multiple data sets to determine forms within social, economic, and legal structures. In this way, big data is constructed and used as an effective way to tackle the complexity of the datafied social world through datafication (Breiter and Hepp, 2018, p.392).

Big data has provided new opportunities for individuals along with challenges. Citizens have become more aware of the possibility of data used for social change. This opportunity has led to the rise of new social practices through technology and data that has also paved the way for the formation of a new type of activism *data activism*. Data activism includes a series of practices based on the social and technological dimensions of human action. It aims either to resist massive data collection or pursue the exploitation of available data for social change (Milan and Gutierrez, 2015, p.120-122). As Gutierrez

(2018) stated, there are two main types of data activism: Reactive and Proactive data activism. While reactive data activism is related to resisting threats of massive data collection, proactive data activism is about the use of possibilities of datafication for civic engagement, advocacy, and campaigning. Data activism is a new concept and practice, which has become popular with the emergence of big data. In the literature, there are many studies analyzing data activism theoretically based on datafication and data politics (Milan and Gutierrez, 2015; Milan and van der Velden, 2016; Gutierrez, 2018; Kennedy, 2018; Lehtiniemi and Ruckenstein, 2019; Beraldo and Milan, 2019), and practically by focusing on particularly reactive data activism cases from a political participation perspective (Kenzi and Langlois, 2015; Kannengieber, 2019; Chenou and Capda-Masmela, 2019; Richer et al., 2020). However, no research or study has been reached, which analyzed proactive data activism from the perspective of Communication Action Theory by Habermas. As Gutierrez (2018) suggested, the analysis of data activism as a social meaning-building process based on a communicative action perspective is necessary. The current study aims to examine proactive data activism within the scope of the related theory of Habermas. In this context, this research can be the pioneer for further studies. The case study will be employed as a research method to analyze the case of *DataKind*. The research questions are as follows:

RQ1: How datafication has changed the practices of activism?

RQ2: How to understand data activism within Communicative Action Theory?

RQ3: What kind of communicative practices are conducted by the proactive data activist platform DataKind?

Based on the research questions, first, the research will focus on how the datafication process has changed the structure of activism in the digital age. Then, proactive data activism will be dealt with in the Communicative Action Theory of Habermas. Finally, the case of DataKind will be analyzed through case study analysis.

The Intersection of Datafication and Activism

Two main concepts of big data, which are *data* and *information* have different meanings. While data is related to processing, information is about interpretation. On the other hand, new epistemologies in Human and Social Sciences argue that data is not an objective construct, rather, it is completely constructed. Therefore, the term "raw data" is an idealization (Pellegrino, Söderberg, and Milan, 2019, p.91). During the development of civilization, various data factors have been defined, assessed, analyzed, and calculated by humans to navigate reality. Regardless of being qualitative or quantitative, these factors have given cultural eras their specificity. Different from the later age, data had never been abundant and complex in the past. In addition, in past, human observation was the main tool for interpreting the data. Today, networked non-human factors are used to generate machine-readable data (Uricchio, 2017, p.125).

There are three main characteristics of data, which are also called "3Vs" including *volume, velocity,* and *variety.* The key considerations of volume are scalability, distribution, and the ability to process. Velocity points to the granularity of the data system, decision making, and action-taking. The degree of information loss in clean-up, semantic integration, and versatility are the components of variety (Lycett, 2013, p.381). Based on those characteristics of data, "datafication" is coined to refer to digitalization, which mainly involves a growing number of media and software applications (Hepp and Hasebrink, 2018, p.22). There are three main concepts related to datafication, which are: dematerialization, liquefication, and density. While dematerialization is related to the skill of separating the informational aspect and its use from the context of the physical world, liquefication comes to mean the possibility of manipulation of dematerialized information in a given suitable infrastructure. On the other hand, density can be defined as the (re) combination of resources at a given time and place- as the result of the value formation process (Lycett, 2013, p.382).

Datafication is mainly about the rendering process of data aspects of the world. This process can vary from demographic data to behavioral metadata, which are automatically derived from smartphones. In this context, datafication is not only data collecting or

analysis but also providing such data back to its users (Kennedy, Poel,l and van Dijck, 2015). The datafication and digital world have provided two main advantages for activism: (a) low costs for the coordination of protests, and (b) a minimum physical requirement for protests. Those opportunities have changed social practices of activism such as information sharing,awareness-raising, and campaigning which have become more accessible and more effective online (Gutierrez, 2018, p.15).

The use of massive data collection has led to the formation of a new concept and type of activism that is called "data activism". Data activism research is mainly related to the processes and use of datafication. On the other hand, data activism doesn't focus only on the problematic process of datafication but also analyzes new shapes of civic and political engagement within datafication in order to propose more responsible data futures (Lehtiniemi and Ruckenstein, 2019, p.2). For Chenou and Cepeda-Masmela (2019, p.396-397), data activism is very useful and effective in the formation of alternative aspects of big data in society through the collection, organization, and diffusion of data concerning any social problem.

As for the question of whether activism re-shapes data or vice-versa, Beralde and Milan (2019) argued that data activism can be considered as the area where data re-mediate activism. According to their approach, if we consider communication as a process of the meaning-mathe king, the technological dimension of communication can play a key rothe le in the co-constitution of a field of action in terms of social movement formation. On the other hand, as a theoretical category, data activism includes a great variety of encounters with datafication: on the one hand, it focuses on claims of numbers as a source of tand ruth, on the other hand, it analyzes datafication as transmitting risks (Milan and van der Velden, 2016, p.61). Both aspects of datafication allow data activism to be formed through media practices, which include (a) routinized and creative social practices; (b) interactions with mobile phones, laptops, etc. (c) the perception of media objects and environment. Therefore, it is necessary to know how people consciously and actively seek to transform media technologies to change society (Kannengieber, 2019, p.3).

Proactive Data Activism within the Framework of Communicative Action Theory: The Case of DataKind

There are two sub-fields of data activism: reactive and proactive. While reactive data activism is related to resistance practices against problems or threats to civil and human rights derived from corporate and government, proactive data activism is about taking advantage of the possibility for social change and civic engagement offered by datafication (Milan and Gutierrez, 2015, p.122). To be more precise, the main concern of the reactive form of data activism is to resist threats of massive data collection through technological fixes, on contrary, proactive data activism in pursuit of using the possibilities of datafication to develop civic engagement, advocacy, and effective campaigning (Beraldo and Milan, 2019, p.4). For Gutierrez (2018, p.51), proactive data activism can be understood as self-organized or managed, more citizen-controlled and non-profit activism that hinged on the formation of an alternative communication platform for social change by utilizing the data infrastructure. In this vein, the Communicative Action Theory can project on analyzing different fields such as journalism, digital media, global studies, monitorial citizenship, etc. in terms of data activism understanding. In this vein, the current study is focused on proactive data activism and the digital platform- DataKind based on the communicative action approach of Habermas.

Exploring Proactive Data Activism within Communicative Communication Theory

BasTheteraction between people, information, and technology can be considered as the basis of data activism. This special relationship emerges and reshapes tactics, identities, and modes, of organizing (Milan and van der Velden, 2016, p.61-62). As data activism initiatives or practices are mainly based on interaction, the lens of the communicative action theory can be the most appropriate approach for the analysis. In addition to communicative action, Habermas underlined also three ways of action regulating relations between actor and the world. Those action ways are: *teleological action* (as the decision is the central concept of this action, it refers to the choice of proper means to realize the goal); *normatively-regulated action* (this action is related to members of a social group, whose action is intended to reach common goals); and *dramaturgical action* (this action refers to interaction of the participants to constitute public for another, before

whom they present themselves) (Rochberg-Halton, 1989, p.340). Based on those actions, Habermas underlined that every speech can be contested on three grounds through validity claims for non-coerced or communicative action, which are: (1) the social world, (2) the subjective world, and (3) the objective world. While in the social world, the hearer can contest normative rightness of the utterance, in the subjective world the hearer can contest the subjective truthfulness. On the other hand, the hearer can deny certain existential presuppositions in the objective world (Chriss, 1995, p.554). The social, subjective, and objective worlds are constructed through intersubjectivity. Therefore, the key concept is interpretation. In such a way, interaction is formed between the agents who are capable of speech and action (Flecha et al., 2001, p.116). Similar to communicative action, data activism also hinges on the integration of the social, subjective, and objective realms. Moreover, data activism involves aim-based elements (bounding actors and facts in the objective world); norm-regulated factors, which occur in the social world (campaigning), and dramaturgical action (subjective worlds of campaign leaders). In this respect, Habermasian perspective can help us to analyze proactive data activism as a social meaning-building process, which provides an alternative communication platform (Gutierrez, 2018, p.51).

Within the scope of formal pragmatic reconstruction, Habermas argued that all communication includes a mode of action, which is called communicative action (Dahlberg, 2005). Fundamentally, communicative action is based on the concept of interaction, which is formed between (at least) two subjects that are capable of speech and action to establish and maintain interpersonal relations. To this end, the participants try to reach mutual understanding and find a way of agreement (Habermas, 1987, p.86). In this vein, communicative action is related to intersubjective redemption of validity claims rather than instrumental or strategic action (Dahlberg, 2005, p.5).

Language is the central point of the theory, which was identified as the main tool for conducting all communication actions. In teleological action, language plays the role of one another means among the others. In normatively-regulated action, it leads to conveying of values, contributes to the formation of consensus, and establishment of interpersonal relations. In dramaturgical action, language helps dramatization to

formalize in terms of the expression of experiences. However, only in communicative action, language is the means of understanding without any purpose. In this way, language can be considered as the basis of mutual understanding and consensus formation (Flecha et al., 2001, p.117).

In communicative action, the consensus should be based on the validity claims of the participants. The consensus can be formed if the participant relies on a shared definition of the situation and thereupon acts (Habermas, 1979, p.4). Consensus is the *telos* of communication in communicative action theory, which constitutes the base of his deliberative politics. Talk-based politics is aimed for the telos of communication. On the other hand, consensus can be considered the most criticized concept in Habermas's theory. It is also the most modified concept by Habermas over the years (Jezierska, 2019). Contrary to purposive-rational action, communicative action aims at observing intersubjective valid norms linking to reciprocal expectations. The validity basis of speech/communication is presupposed in communicative action. These validity claims are universal (truth, rightness and truthfulness) that participants raise and recognize reciprocally, which also makes consensus possible (Habermas, 1979, p.118)

In communicative action, those universal validity claims form the basis of the argumentation process. In other words, arguments must be constructed based on those universal claims. The argumentation is related to the "type of speech, in which participants thematize contested validity claims and attempt to vindicate or criticize them through arguments". An argument includes reasons or grounds that are intertwined in the validity claim of expression (Habermas, 1987, p.18). In argumentation, a competitive communication situation is formed where opponents and proponents attempt to convince one another and reach a consensus-based on arguments (Habermas, 1990, p.160). In a broad sense, the argumentation process allows the participants to express their ideas, suggestions or defend their opinions upon proper rationality (van Eemeren, 2001, p.11). For Houtlosser (2001: 40), the construction of rationality in the argumentation process should be based on three main elements: clarity, sincerity, and truthfulness, which are connected with the validity claims of Habermas.

For the possibility of formation of the communicative action, Habermas pointed to differences between the system and "lifeworld" (*Lebenswelt*). The realm system is related to market, bureaucracy, and economy, which refers to the reproduction of the power and money. The lifeworld is the integration area dominated by language. Such integration factors as family, education, etc. are involved in this area. Therefore, the lifeworld is the optimal realm for the formation of communicative action (Sandberg, 2012, p.352). For Salter (2003, p.123), to comprehend the difference between the system and lifeworld clearly, it should be better to focus on the types of rationality upon which they are based. While lifeworld is based on communication rationality, the system is related to instrumental rationality. Similar to lifeworld, the public sphere is also reproduced through communicative action. The public sphere is also a social phenomenon that includes action, actor, association, and collection. It can be defined as "a network for communicating information and points of view" (Habermas, 1996, p.360). In Habermasian normative understanding, the public sphere can be understood as "social space generated communicative action, which should be also protected by systematic imperatives" (Salter, 2003, p.124). Habermas didn't define the public sphere as a specific or bounded public. Instead he mentioned the order of complex networks of multiple and overlapping publics (Dahlberg, 2005, p.6). To this end, interaction or communicative action is formed in the sphere of social institutions (e.g. family, mass media) mediated by language and governed by social rules. Therefore, it is the socio-cultural lifeworld (Keane, 1975, p.88).

In his theory, Habermas (1987) emphasized the socialization role of communicative action in the coordination of purpose-oriented activities on different issues beyond the function of agreement. This role allows the participants to integrate their subjective interests for mutual understanding based on rationality. Communicative discussions can be formed only through validity claims (Sandberg, 2012, p.353-354). Based on this rationality, communicative action aims for three general functions: (1) to convey the "information", (2) to build social relations with others, and (3) to express subjective (individual) emotions and opinions (Edgar, 2006, p.22). Therefore, communicative

rationality provides an unbounded, integrative and consensus-oriented deliberative platform (Habermas, 1984).

In the framework of communicative action theory, as a social meaning-building process, proactive data activism will be examined based on selected three main components- *interaction*, *argumentation,* and *consensus*. By focusing on the case of DataKind, the case study method will be employed to analyze the interactive structure, argumentation base and mutual understanding/consensus formation.

Methodology

In the study, qualitative research framework was selected as the base for methodology. The main reason is to understand the digital platform based on the theoretical approach of Habermas in a more detailed way. On the other hand, descriptive case study analysis was selected as the research technique to analyze the case of DataKind based on the communicative action theory. The case study is one of the most-preferred qualitative research methods to examine any case in a detailed way. The case is the object of the study and it can be analyzed as single or multiple. A case study analysis allows the researcher to tackle one or a few instances of the selected phenomenon comprehensively. In this qualitative analysis, the cases are considered as configurational context, or "path-dependent entities". "Thick description" and "process-tracing" are the main focus, rather than "variable-centered" which dominates in quantitative/positivist research (Given, 2008: 22-68).

A case can be an individual, group (e.g. family, class), institution (e.g. school, factory), event, or large-scale community (e.g. town, industry, profession). Along with single cases, different cases can be also analyzed with a multiple case approach. It depends on the aim of the study and researcher (Gilham, 2000). Case study methods allow the researcher to examine different functional aspects of the selected case. Therefore, the case study should not be considered a data-gathering technique. It is much more a methodological approach that includes various data-collections techniques. There are several designs of case study analysis. Exploratory, explanatory, and descriptive types are commonly used by scholars. Different from others, in descriptive case studies, a

researcher should determine if the unit of analysis is compatible with the theoretical framework (Berg, 2001: 225-230). In this respect, Yin (1994: 21) identified five fundamental components for descriptive case study analysis as follows:

- Research question
- Theoretical propositions
- Unity of the analysis
- The logical connection between theory and practice
- Categorizing of the findings

Although the above-stated approach is the most-preferred and used system, the design and approach can be varied relevant to the functions and characteristics of the case (Hancock and Algozzine, 2006: 31).

Data Collection

In this study, the digital platform DataKind was selected within purposive sampling due to its relevance for this research. The selected case was analyzed through descriptive case study analysis. The information about the platform's infrastructure and activities was obtained by mainly focusing on its website. Although the researcher attempted to do also an in-depth interview with the team, any person from the platform didn't respond to the request. Therefore, the main data was collected from the platform's website. Based on the research question of how the case can be analyzed within the framework of communicative action theory, *interaction, argumentation,* and *consensus/mutual understanding* are identified as the main criteria for the analysis in addition to the general structure of the study's unit. By focusing on the mentioned four categories, the case will be examined in the next section in a detailed way.

Findings: The Case of *DataKind*

The current study aims to find out how to understand data activism from the communicative action approach. In this context, the digital platform DataKind was selected as the object of research due to its focus on the use of data for activist practices. Based on the communicative action theory, four main categories were identified for the

research findings that are: *General overview, interactive structure, argumentation process, and consensus-based implementation*

General Overview

The digital platform DataKind (www.datakind.org) has been continuing its activities since 2011. It has become a global network that includes five chapters located in the USA (Washington and San Francisco), UK, India (Bangalore), and Singapore. DataKind is defined as an effective platform that attempts to bring data science and its forms for the service of humanity. In this vein, high-impact organizations are encouraged to work in cooperation with various data scientists and social change-makers to struggle with humanitarian challenges in effective way. The objective of this platform is to help organizations develop their evidence-oriented decision-making, increase efficiency and enhance their data literacy. In this respect, DataKind provides an effective platform for data scientists, leading social change organizations to work together and cooperate on the latest analytics and advanced algorithms to increase social impact. In doing so, several programs from daily/weekly events to multi-month projects are designed and conducted to connect social organizations with the related data science team relevant to the needs, which can be any kind of critical humanitarian issue such as education, poverty, health, human rights, the environment, cities, etc[2]. The slogan of the platform addresses the importance of data for social changes worldwide, as below:

> *"...use data to not only make better decisions about what kind of movie we want to see but what kind of world we want to see..."*

There are three main components of the DataKind platform: *DataCorps, DataDives,* and *Community Events*. DataCorps was designed for long-term projects in which organizations determine their needs and discuss the possibilities, and thereby they can be matched with the relevant team that can transform the requirements into data science problems and solve them with advanced analytics. The DataCorps teams consist of top

---

[2] General Information about DataKind. Available at: https://www.datakind.org/about (accessed 21 September 2020)

data scientists and project managers in the sector under the control of a Data Ambassador and Project Manager. Each DataCorps project is conducted by different teams with different skillsets, which can change according to the project type. DataDives was designed for marathon-style events in which mission-driven organizations work together with the related team members who are volunteer data scientists, program developers, and designers to process data for purposive works[3]. The DataDives teams consist of up to 15-20 volunteers with a variety of skills and are controlled by a Data Ambassador and Project manager. Any kind of mission-driven organization including nonprofits, social enterprises, or governmental agencies can apply to cooperate with DataDive teams. Community events are designed to be effective spaces for data and social sector experts to connect and collaborate for the current and future works. Social sessions and workshops are mostly conducted within those events.

Interactive Structure

As frequently stated on its website and activities, the main objective of DataKind is to bring together data scientists, volunteers, and social change activists in a common platform to connect and collaborate for efficient solutions. To this end, different interactive activities are conducted for the participants. DataDives sessions and Community Events are the best examples to focus on.

Within the scope of DataDives, one or two-day work sessions are carried out to allow volunteers to analyze, visualize and present data for the organizations. These sessions are considered as effective platforms in which volunteers become much more skillful and establish new connections to find quick solutions for the problems. They act as a sort of discussion place for the participants.

Under the category of community events, three main sub-categories were mentioned, which are social mixers, educational workshops, and project accelerators. In social mixers, cross-sector collaboration is aimed to be achieved through one-to-one

---

[3] DataDive Section. Available at: https://www.datakind.org/datadive (accessed 21 September 2020)

communication, and conversation activities. In educational workshops, the events are conducted to allow an expert speaker or DataCorps team to present their works as real examples of data science in action. Under the sub-category project accelerators, the brainstorming sessions are conducted to bring the data science experts and organizations together in order to provide advice and consulting for the development of the projects.

There are also several meetup groups which are located in the above-mentioned regions and conduct various interactive sessions between data science experts and the related organizations. There are 3,784 members of DataKind DC group, 5,001 members of DataKind UK, 3,741 members of DataKind NYC, 2,987 members of DataKind SF Bay Area, 4,4490 members of DataKind SG, and 2,349 members of DataKind Bangalore.

Argumentation Process

In this research category, it will focus on how the platform DataKind provides the argumentation process for collaboration and thereby the development of projects. In this respect, the guidelines of the collaboration for the projects of DataCorps and DataDives were mainly examined. It was commonly shown that three main questions are asked within the argumentation process of project development, which are *who, how,* and *what*.

In the scope of DataCorps projects[4], the first question who is responded to data scientists and project managers from the sector, it was stated that "any mission-driven organization can apply to work with the DataCorps team". For this, the representative or a focal person should be identified by the organization to contact and manage the relations with the DataCorps team. Projects are led by the selected Data Ambassador and Project manager, who is responsible for conducting the mutual communication and development process of the project. The whole team includes one project manager, one data ambassador, two data experts, one partner representative, one project champion, and two data specialists. As to the question of how five consecutive phases are determined. The first step is problem exploration which aims to explore what is possible and start the collaboration by making a team. The following step is called data discovery the DataCorps team identifies

---

[4] DataCorps section. Available at: https://www.datakind.org/datacorps (accessed 21 September 2020)

internal and external data sources relevant to the organization's needs. Within the third step- prototyping, the team works on solutions to the problem. Following this, the ultimate needs of the organization are met through adjustments made by the team relevant to the feedback of the organization. In the last phase, the team shares the completed version of the output. Consequently, based on the question of what, the argumentation process for arguing the importance of the collaboration and convincing the applicant organization is completed with the emphasis on the production of data science solutions that lead to action and transform an organization's work identifying the people in need.

In DataDives projects, the responsible team can consist of between 10 and 20 volunteers with specific skills and members of mission-driven organizations. The ream set-up is similar to DataCorps projects. Different from DataCorps, the whole team includes one data ambassador, 5-10 DataDivers, and 1-2 partner representatives for DataDives projects. In the scope of the question of how, three steps were determined that are *exploration, the big event,* and *handoff*. Exploration is clarified as identifying of the key data questions in collaboration with the partner organization. The big event refers to the set-up of a team and effective collaboration through the assignment of the related volunteers for the project. The last step is handoff, which is related to understanding a truly difficult problem through data and presenting the data science solutions to the organizations. In the framework of the question of "what gets produced", it was stated that organizations will have an ability to use data to advance their missions.

Consensus-Based Implementation

Under this category, it focused on three different projects of DataCorps to understand the infrastructure of implementation based on consensus and collaboration that was derived from the argumentation process. This category is also necessary to form a general opinion on the implementation methods of proactive data activism practices as the outcomes of collaboration.

In the project, which was entitled "Sanergy" and implemented in Nairobi, Kenya, two main objectives were identified for the aim of using time-series prediction to develop access to safe sanitation. While the first aim is to enable updates for forecasts in response

to changing network plans, the second one is to design a field research to measure usage at a small number of the selected locations. Based on those objectives, the mission of Sanergy is defined to provide safe sanitation to people who are deprived of access to sanitation facilities in dense urban areas. In this vein, the daily data on residential waste are used to foresee future collection volume. In addition, a field survey was employed within a specific sampling of FLT (Fresh Life Toilets) locations to inform both the numbers of users and estimates of residential usage. Consequently, the DataCorps team set a forecasting model to estimate the amount of solid waste that will be collected on weekly and monthly bases across the FLT network[5].

In the project, which was titled "The Welcome Centre" carried out in Huddersfield, UK, two main objectives were identified to identify food bank dependency early. The first objective is to establish a machine-learning model to determine and analyze which of the Welcome Centre's clients are presumably in need of the food bank's support[6]. The second objective is to adapt this machine-learning model into the system of The Welcome Centre. The Welcome Centre (TWC) is a food bank located in Huddersfield, UK. The mission of this center is to provide support such as food, toiletries, or household packs to people in crisis. In collaboration with TWC, the DataCorps team developed a machine learning model using the existing data which was available after three years of TWC activity. The data included the specific issues, personal characteristics, and historic patterns of referrals of the clients who were registered in TWC.

In the project "Using machine learning to understand what drives student success in Dallas County", which was carried out in Texas, USA, the DataCorps team worked with The Commit Partnership (Commit) to analyze the academic success performances of students on a periodical basis The Commit Partnership is a platform including more than 200 partners (e.g. public and private schools, colleges, and universities, foundations, businesses, and nonprofits) founded in 2012 to deal with the educational problems of

---

[5] DataKind Projects. Av ailable at: https://www.datakind.org/projects/using-time-series-forecasting-to-improve-access-to-safe-sanitation (accessed 21 September 2020)

[6] DataKind Projects. Available at: https://www.datakind.org/projects/identifying-food-bank-dependency-early (accessed 21 September 2020)

Dallas County. Those problems were mainly seen as low-levelled childhood education, improvement of educators and the need to increase completion rates of colleges. Although public education data is difficult to access in Texas, the DataCorps team conducted an analysis on aggregate school and district data levels to access deidentified student-level records. The team prepared re-usable scripts for data extract, database management, and descriptive visualizations that ease the data to be accessed, analyzed, modelled and stored in a reproducible way[7].

Conclusion

Data activism can be understood as the collection, organization, and diffusion of data to cope with any specific social problem (Chenou and Cepeda-Masmela, 2019). Although data activism is a new concept and understanding for social movements, in particular, digital activism literature, it has been studied mostly theoretically and concerned with political participation. No study has been done that analyzes data activism from the perspective of communicative action theory. Since data activism can be considered as the alternative-digital communication platform based on data use and therefore, as Gutierrez (2018) stated, it is also "a social-meaning building process". The theory of Communicative Action of Habermas can help us to understand this phenomenon from a different perspective. Starting from this point of view, the current study aimed to analyze proactive data activism within the possibilities of social and civic engagement that can be formed through datafication. In this context, three main research questions were identified for the detailed analysis.

Based on the first research question, it was revealed that, in addition to providing cheaper costs and non-physical organizing, the datafication process offered an abundance of data and information to use for the good of social changes. Since the datafication is not only a

---

[7] DataKind Projects. Available at: https://www.datakind.org/blog/using-machine-learning-to-understand-what-drives-student-success-in-dallas-county (accessed 21 September 2020)

data collection or analysis but also feeding such data back to the users (Kennedy, Poell, and van Dijck, 2015), data activism practices have become an effective tool to cope with various social problems by utilizing the data infrastructure. As Beraldo and Milan (2019) stated, datafication re-mediates activism. It has re-shaped strategies, and forms of organizing (Milan and van der Velden, 2016).

In the scope of the interrelated consecutive research questions concerning the connection between data activism and communicative action theory, the proactive data activist platform DataKind was analyzed based on the main components of theory interaction, argumentation process, and consensus. Basically, interaction is the core concept of the communicative action theory that is formed at least between two subjects which are capable of speech and action to establish and maintain interpersonal relations. In such way, the participants seek to reach mutual understanding and agreement based on validity claims that form the argumentation process (Habermas, 1987). In the argumentation process, a competitive communication situation is created for opponents and proponents to convince one another and reach a consensus (Habermas, 1990). To put it differently, this process allows the participants to express their ideas, suggestions or defend their opinions (van Eemeren, 2001). Therefore, in communicative action, consensus can be made through the validity claims (truth, rightness, and truthfulness) of the participants. The formation of consensus is hinged on the reliability of the shared definition of the situation (Habermas, 1979). Above all, Habermas addressed the realm- lifeworld for the formation of communicative action. Lifeworld is an integration area, which includes family, education, etc. and is dominated by language based on communication rationality (Sandberg, 2012). Similar to lifeworld, the public sphere involves action, actor, and association and is reproduced through communicative action (Salter, 2003). Simply, the public sphere can be understood as "a network for communicating information and points of view" (Habermas, 1996). Within the framework of the Habermasian perspective, since the digital world is also a realm of network and interlaced relations, it can be also considered as a public sphere or a part of lifeworld that makes the formation of communicative action possible. DataKind has been one of the platforms of this digital world since 2011. There are three main components of DataKind which are: DataCorps

(long-term projects), DataDives (short-term events), and Community events (various activities). Holistically, DataKind attempts to provide a multi-dimensional platform for data scientists and leading social change organizations to work together on the latest analytics and advanced algorithms to increase social impact. The main objective of this platform is to help organizations enhance their decision-making, improve data literacy and increase efficiency. In doing so, DataKind aims to form a network and collaboration among data scientists, volunteers, and social change actors through various interactive activities such as DataDives sessions (one or two-day sessions) and community events (conversation activities, one-to-one communication, brain-storming sessions, workshops). The argumentation process of building the relations and collaboration between the participants is formed based on three questions: who, how, and what. The questions or phases- who and what are defined similarly addressing participants and the outcomes in both DataCorps and DataDives projects; however, the question how involves different arguments. While in DataCorps, five steps are identified which are problem exploration, data discovery, prototyping, refinement, and solution, in DataDives three steps should be followed, which are exploration, the big event, and handoff. As a result of this argumentation process, the related teams and activities are determined to be implemented based on the mutual understanding and agreement of the participants. In this context, three projects of DataCorps have analyzed in the category of consensus-based implementation. It is thought this can also allow us to understand how the collaboration resulted in proactive data activism practices, and what is achieved. The findings showed that in the project- "Sanergy", first the sanitation issue was explored, the field study was conducted for the required data and the solutions were evaluated for refinement. Consequently, a forecasting model was developed by the DataCorps team to estimate the amount of solid waste based on the weekly and monthly collection bases. As to the other project- "The Welcome Centre", two main objectives were identified- to build a machine learning model and to employ this model in the center. Similar to the implementation steps followed in Sanergy, the team developed a machine learning model using the existing data to determine and analyze which clients of the Centre are most probably in need of food support. In the project related to educational problems in Dallas County, the

Proactive Data Activism within the Framework of Communicative Action Theory: The Case of DataKind

DataCorps worked in effective collaboration with the Commit Partnership to analyze the academic success of the students on a periodical base. The DataCorps team analysed aggregate school and district data levels to access deidentified student-level records. Finally, they developed reusable data scripts for data extraction, database management, and descriptive visualizations.

In this research, (proactive) data activism was analyzed within the framework of Communicative Action Theory. The findings revealed that as an example of proactive data activism, DataKind provides an integrative and agreement-oriented deliberative platform for the development and implementation of projects. In addition, the platform is useful and effective for information exchange, and social relations building among the users. Since no comprehensive research has been reached during the literature review that focuses on the connection between data activism practices and the theory of Habermas, it is expected that this study will contribute to both digital activism and communication literature. Furthermore, the theoretical background and the research findings can be directive in future studies, which can analyze different cases from different regions.

REFERENCES

Beraldo, D., & Milan, S. (2019). From data politics to the contentious politics of data. *Big Data & Society*, 1-11. DOI: 10.1177/2053951719885967

Berg, L. B. (2001). *Qualitative research methods for the social sciences*. Boston: Allyn-Bacon.

Breiter, A., & Hepp, A. (2018). The complexity of datafication: Putting digital traces in context. In A. Hepp, A. Breiter and U. Hasebrink (eds.), *Communicative figurations* (pp. 387-407). Bremen: Palgrave Macmillan.

Chenou, J., & Cepeda-Masmela, C. (2019). #NiUnamenos: Dta activism from the South. *Television & New Media, 20* (4), 396-411. https://doi.org/10.1177%2F1527476419828995

Chriss, J. J. (1995). Habermas, Goffman, and communicative action: Iplications for professional practice. *American Sociological Review*, *60* (4), 545-565

Dahlberg, L. (2005). The Habermasian public sphere: Taking difference seriously? *Theor Soc, 34*, 111-136. https://doi.org/10.1007/s11186-005-0155-z

DataKind Platform. (2020). Information about the platform. Retrieved from: https://www.datakind.org/ (accessed date: 15 September 2020).

DataKind Platform. (2020). DataCorps section. Retrieved from https://www.datakind.org/datacorps (accessed date: 21 September 2020)

DataKind Platform. (2020). DataKind projects. Retrieved from: https://www.datakind.org/projects/using-time-series-forecasting-to-improve-access-to-safe-sanitation (accessed date 21 September 2020)

DataKind Platform. (2020). DataKind projects. Retrieved from: https://www.datakind.org/projects/identifying-food-bank-dependency-early (accessed date 22 September 2020)

DataKind Platform. (2020). DataKind projects. Retrieved from: https://www.datakind.org/blog/using-machine-learning-to-understand-what-drives-student-success-in-dallas-county (accessed date 23 September 2020)

DataDive Projects. (2020). DataDive section. Retrieved from: https://www.datakind.org/datadive (accessed date 21 September 2020)

Edgar, A. (2006). *Habermas: The key concepts*. New York: Routledge

Flecha, R., Gomez, J., & Puigvert, L. (2001). The theory of communicative action. *Contemporary Sociological Theory*, 250, 109-127.

Gilhem, B. (2000). *Case study research methods*. New York: Continuum.

Given, M. L. (2008). The sage encyclopedia of qualitative research methods. Thousand Oaks: Sage.

Gutierrez, M. (2018). *Data activism and social change*. Madrid: Palgrave Macmillan

Habermas, J. (1979). *Communication and the evolution of society*. Boston: Beacon Press.

Habermas, J. (1987). *The theory of communicative action-Volume II*. Boston: Beacon Press

Habermas, J. (1990). *Moral consciousness and communicative action*. Cambridge: MIT Press.

Habermas, J. (1996). *Between facts and norms*. Cambridge: MIT Press.

Hancock, R. D., & Algozzine, B. (2006). *Doing case study research*. New York: Teachers' College Press.

Hepp, A., & Hasebrink, U. (2018). Researching transforming communications in times of deep medialization: A figurational approach. In A. Hepp, A. Breiter and U. Hasebrink (eds.), *Communicative figurations* (pp. 15-51). Bremen: Palgrave Macmillan.

Houtlosser, P. (2001). Points of View. In H.F. van Eemeren (ed.), *Crucial concepts in argumentation theory*, (pp.27-51). Amsterdam: Amsterdam University Press.

Jezierska, K. (2019). With Habermas against Habermas. Deliberation without consensus. *Journal of Public Deliberation*, *15* (1), 1-31. DOI: 10.16997/jdd.326

Kannengieber, S. (2019). Reflecting and acting on datafication- crypto-parties as an example of reactive data activism. Convergence, 1-14. DOI: 10.1177/1354856519893357

Kennedy, H., Poell, T., & van Dijck, J. (2015). Data and agency. *Big Data & Society*, 1-7. DOI: 10.1177/2053951715621569

Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis : Journal for Contemporary Philosophy,* (1), 18-30. DOI: 10.1177/2053951715621569

Keane, J. (1975). On tools and language: Habermas on work and interaction. *New German Critique*, 6, 82-100.

Kitchin, R. (2014). Big data, new epistemologies, and paradigm shift. *Big Data & Society*, 1-12. DOI: 10.1177/2053951714528481

Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *Big Data & Society*, 1-12. DOI: 10.1177/2053951718821146

Lewis, S. C., & Westlund, O. (2015). Big data and journalism. *Digital Journalism, 3*(3), 447-466. https://doi.org/10.1 080/21670811.2014.976418

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

Milan, S., & Gutierrez, M. (2015). Citizen's media meets big data: The emergence of data activism. *Mediaciones*, 14, 120-132.

Milan, S., & van der Velden, L. (2016). The alternative epistemology of data activism. *Digital Culture and Society, 2* (2), 57-74. DOI 10.14361/dcs-2016-0205

Moffat-Hannah, K. (2019). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology, 23* (4), 453-470. tps://doi.org/10.1177/1362480618763582

Pellegrino, G., Söderberg, J., & Milan, S. (2019). Datafication from below: Epistemology, ambivalences, challenges. *Tecnoscienza, 10* (1), 89-113.

Renzi, A., & Langlois, G. (2015). Data activism. In G. Langlois, J. Redden and G. Elmer (eds.), *Compromised data: from social media to big data* (pp.202-226). New York: Bloomsbury.

Richer, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in cape town, South Africa. *The Canadian Geography*, 1-15. DOI: 10.1111/cag.12608

Rochberg-Halton, E. (1989). Jurgen Habermas's theory of communicative etherealization. *Symbolic Interaction, 12* (2), 333-360.

Salter, L. (2003). Democracies, new social movement, and the internet: A Habermasian analysis. In M. M. Coughey & M. A. Ayers (eds.), *Cyberactivisim* (pp. 117-145). New York: Routledge.

Sandberg, F. (2012). A Habermasian analysis of a process of recognition of prior learning for healthcare assistants. *Adult Education Quarterly, 62* (4), 351-370. DOI: 10.1177/0741713611415835

Uricchio, W. (2017). Data culture and the ambivalence of algorithm. In M. T. Schafer and K. van Es (eds.). *The datafied society* (pp.125-139). Amsterdam: Amsterdam University Press.

Van Eemeren H, F. (2001). *Crucial concepts in argumentation theory*. Amsterdam: Amsterdam University Press.

Yin, K. R. (2003). *Case study research: design and methods*. Thousand Oaks: Sage.