

SMALL DATA AND SUSTAINABLE DEVELOPMENT: INDIVIDUALS AT THE CENTER OF DATA-DRIVEN SOCIETIES

Mamello Thinyane

United Nations University institute on Computing and Society, Macao SAR, China

Email: mamello@unu.edu

ABSTRACT

At the centre of data-driven societies are individuals and end-users who not only generate data, but also benefit from the outcomes of the data-driven development. Extensive work has been undertaken to understand and explore the challenges and potential impact of data, in particular Big Data, for the private as well as the public sectors. Similarly work has been undertaken within the domains of Personal Informatics and life-logging, which has investigated the role of data, and specifically personal physical activity and health data towards improving the wellbeing of individuals. In this research we investigate the engagement of individuals in the use of data towards the achievement of the sustainable development imperatives as articulated in the 2030 Agenda for Sustainable Development. The paper presents: the awareness levels of the participants with regards to the Sustainable Development Goals; their attitudes and perceptions around monitoring of social indicators; key considerations associated with data ownership, privacy and confidentiality of data, as well as sharing of data within the data ecosystem. The paper subsequently discusses how these findings could inform the implementation of small data tools to support the active engagement of individuals in data-driven societies.

Keywords— Small Data, Sustainable Development, Data Driven Development

1. INTRODUCTION

Data is increasingly pervasive and ubiquitous in the 21st century data-driven societies, and the use of data stands to increase and to permeate more societal domains. Advancements are being seen in the use of data, in particular Big Data, for supporting businesses and the corporate sector towards improved decision-making and effectiveness in areas such as Business Intelligence and Analytics (BI&A) [1] as well as management [2], [3]. Similarly the public sector use of data can be noted towards improved policy development [4], and service delivery, where “governments expect big data to enhance their ability to serve their citizens and address major national challenges involving the economy, health care, job creation, natural disasters, and terrorism” [5]. The increasing ubiquity of data in society is not only seen in the increased use in organizations but also in increased data use by individuals in areas such as life-logging [6], [7], associated with the

proliferation of activity trackers and mobile devices, as well as in the embedding of digital infrastructures and ‘everyware’ [8] within smart cities initiatives [9].

At the global level, there has been recognition of the role of data towards supporting the achievement of the global development imperatives as articulated in the 2030 Agenda for Sustainable Development. The United Nations Sustainable Development Goals (SDGs) resolution highlights, within Goal 17, technology as an explicit Means of Implementation towards the achievement of the goals [10]. Further, the role of the indicators data within the SDG programme has also been well articulated [11], [12]. The focus within the social indicators monitoring domain has largely been about using data for the purposes of Planning, Monitoring and Evaluation at the national and sub-national levels. Extensive research has been undertaken exploring the potential for data in this regard [12]–[14]. Social indicators monitoring in this context is typically driven by the work of the statistics community, in particular the National Statistics Offices, government departments, as well as multilateral organizations.

Beyond supporting the national-level and multi-lateral stakeholders, the use of data can be towards supporting individuals and community-level actors (e.g. Non-Governmental Organization, and Community-Based Organizations). There has been increasing efforts and research around the use of data for informing individual wellbeing goals and imperatives. The growing field of personal informatics, quantified-self, and lived informatics represent this interest and focus on data that is collected by individuals for the ultimate utility that accrues towards the individuals [15]. Within these domains work has been undertaken that explores data for supporting social sense-making [16], using personal data for improving patient-provider communication in the healthcare sector, and using quantified-self data for cancer rehabilitation.

Mortier *et al* have suggested that, with the growing amount of ubiquity and pervasiveness of data, there is a need to explicitly consider Human Data Interaction (HDI), which they define as being constituted of three key themes and domains: legibility, agency and negotiability [17]. *Legibility* regards ensuring the comprehensibility of data and the associated algorithms, so that the individuals are aware of their data and the implications of its use; *agency* in this context is about allowing individuals the freedom and capacity to act within the data ecosystems; and *negotiability* concerns the dynamic relationships that emanate from the individuals’ interaction with the data. These three themes

Paper accepted for presentation at “Challenges for a data-driven society” ITU Kaleidoscope academic conference, Nanjing, China, 27-29 November 2017, <http://itu.int/go/K-2017>.

provide a suitable initial framing of the key concerns for consideration in the role and engagement of individuals in data-driven societies.

In this research we investigate the engagement of individuals in the use of data towards the achievement of the sustainable development imperatives as articulated in the 2030 Agenda for Sustainable Development. Section 2 provides a theoretical articulation of dynamics of data for development and also provides a characterization of small data. This is followed in section 3 by a presentation of the survey that was undertaken to understand participants' attitudes and practice towards collection and sharing of social indicators data. Section 4 and 5 provides a discussion informed by the findings of the research, and section 6 concludes the paper.

2. SMALL DATA AND DEVELOPMENT

While there is a tendency to consider data generally and broadly, different types (e.g. Big Data, open data, citizen generated data, small data, real-time data) of data stand to play varied roles within the 21st century data ecosystem. For example, the dynamics associated with the use of Big Data by governments are very different from the dynamics of open data, wherein the former could be associated with citizens disempowerment through increased surveillance and collection of citizens data, while the latter could be associated with empowerment of the citizenry through the increased openness and transparency of government towards the citizens. Similarly citizen generated data could play a more elevated role towards better describing grassroots social development phenomenon, while small data could lead to individuals being connected to more actionable and relevant insights.

Small data, which is of main focus in this research, is increasingly being conceptualized and defined in very distinct terms. Small data has been articulated: as the digital traces around an individual [18]; as data from an ethnographic and human-centric investigation of a social phenomenon [19]; and as an approach to analyzing data at the same unit of sampling [20]. In this research we adopt the characterization of small data for development as “an approach to data processing that focuses on the individual as the locus of data collection, analysis, and utilization towards increasing their capabilities and freedom to achieve their desired functioning” [21]. This definition focuses on the individual and their associated interactions within the data ecosystem, and from this perspective the emphasis in the consideration of data for development is about connecting individuals with the relevant data (including personal data, big data, open data, real-time data) towards their development and wellbeing.

The value proposition for data for development follows the traditional information value chain, wherein data that is converted to relevant information, informs decision making and has an impact on life (Figure 1). This basic information value chain however oversimplifies a complex and much

more nuanced process that typically unfolds at the interplay of individual's agency and the structural context. Using Sen's capability approach can help to unpack this process in a more generalized manner wherein the data that people have access to is simply a resource that potentially increases individuals' capabilities set, and therefore one that individuals can use to achieve their desired functionings [22]. This process can be further enunciated by identifying one of the mechanisms through which collected social indicators data is utilized, which is for facilitating individuals' reflection and critical awareness of their own life and circumstances [23].

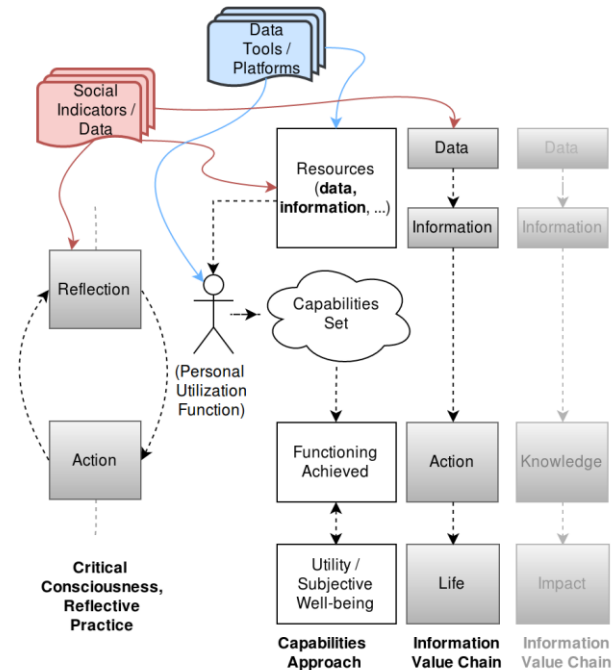


Figure 1 – Expounded information value chain

The notion of using data, in particular personal data, for reflection is also noted as an explicit phase in the Stage-based model of Personal Informatics Systems which consists of the stages of [15]: *preparation* – establishing motivations for tracking data and identifying which data is of interest; *collection* – the actual collection of the identified data; *integration* – processing of the data, which involves analysis, combination and transformation; *reflection* – when users engage with the data for the purposes of sense-making and meaning-making; and *action* – which is the stage at which individual actively chose a particular course of action informed by the reflection on the data collected.

The motivation, within the *preparation* phase in the Stage-Based Model of Personal Informatics Systems, for people to collect information about themselves is usually to advance self-knowledge, self-insight and to promote positive attitudes and behaviors [24]. Rooskby *et al* identify what they term “styles” of personal information tracking which are [6] : *Directive tracking* – wherein individuals record and track personal information towards a

goal achievement. For example, measuring and monitoring the number of steps taken towards reaching the goal of 10,000 steps a day; *Documentary tracking* – this is the recording of personal information for the purposes of documenting their lives, in a manner not different from journaling about one’s daily activities; *Diagnostic tracking* – this is where the recording of personal information is done with the goal of identifying links between various phenomena. For example, to diagnose the link between diet and the occurrence of stomach ailments; *Collecting rewards* – this is for cases where individuals undertake recording of personal information towards receiving a specific value reward; and *Fetishised tracking* – in this case the allure and the pull of the technology is the main motivation for individuals to measure and monitor their personal data.

Individuals play varied and diverse roles within the data ecosystem such as data producers, collectors, curators, and consumers. Illustratively one can note examples of individuals playing the role of being data producers, not only in personal informatics domain but also through the myriad of mechanisms, such as using social media tools and also through what has in recent times been termed digital traces.

3. INDIVIDUALS ENGAGEMENT IN THE SDG DATA ECOSYSTEM

The methodological design of this research is framed around three lines of inquiry: to contribute better understanding of participants’ attitudes and values towards social indicators monitoring, their current practice towards social indicators monitoring, as well as the use of technology to support social indicators monitoring. This investigation is framed in the context of the Sustainable Development Goal 3, which aims to “ensure healthy lives and promote wellbeing for all at all ages”.

A survey instrument has been used in this research for quantitative data collection for: informing an understanding of the participants (i.e. in terms of basic demographics), for the high level identification of their values and attitudes towards monitoring and tracking of individually relevant social indicators data, and for understanding the current practice around monitoring of relevant social indicators. The survey is framed to inform a non-probabilistic descriptive understanding of these issues for the specific individual participants in the research, without making wider population group generalizations. This survey instrument was administered online and the recruitment of the participants was done through email invitations, social media channels and virtual snowballing techniques.

In total 53 people started the survey and 37 of those completed the survey. The bulk (58.2%) of the respondents are young people within the 25 – 34 age group, with the other big groups being 35 – 44 age group at 27.6%, and 45 – 54 age group at 9.8%. The participants are mostly educated and technology-savvy with 94.1% holding a

university degree; and with technology ownership at 96% and 88% for smart-phones and laptops/computers respectively.

The key findings from the survey are presented hereafter through the analytical lens of the Stage-Based model of Personal Informatics Systems to highlight and map the insights from the survey against the five phases of information flow within this model [15].

3.1. Awareness of SDGs

The Sustainable Development Goals form the backdrop against which the consideration of individuals’ contribution and participation in the data-driven society is considered in this research. The awareness and knowledge of the SDGs by the participants varies widely from 29.8% of the participants indicating not knowing the SDGs, and 25.5% indicating to have heard about the SDGs; to 6.4% who indicated that they know all the SDGs in detail (Figure 2).

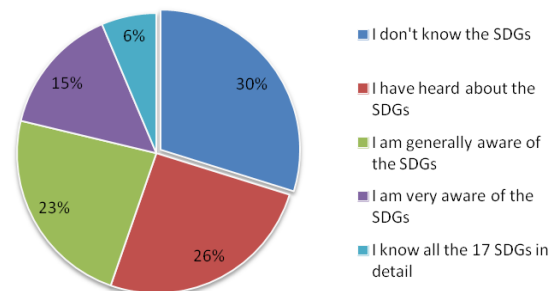


Figure 2 - Knowledge and awareness of the SDGs

Further evaluation was undertaken to understand the importance that the participants ascribe to the different developmental issues as articulated in the SDGs narratives. On a five star (1 – 5) rating of the importance of the various goals, the average rating across all the SDGs was 4.45. The lowest average rating at 4.06 was for SDG 14 (“*Conserve and sustainably use the oceans, seas and marine resources for sustainable development*”) and the highest average rating at 4.79 was for SDG2 (“*End hunger, achieve food security and improved nutrition to promote sustainable agriculture*”). For the rankings for SDG3 (“*ensure healthy lives and promote wellbeing for all ages*”), which is of particular focus in this research, the finding is that the majority (80.9%) of participants considered this a very important goal at 5 stars.

The importance of awareness as a step towards reaching the SDGs targets can be alluded from literature wherein a high correlation was found between the level of public awareness of the Millennium Development Goals (MDGs) and the likely achievement of the MDG indicators [25]. As such building and increasing public awareness around the SDGs remains an important initial step towards garnering public support and engagement in contributing data towards the monitoring of the SDGs.

3.2. Motivations and incentives for data use

During the *preparation stage*, individuals’ motivation for collecting and using data are considered. At this stage individuals also make decisions regarding not only the data that they intend to monitor, but also the associated tools that they use for the monitoring.

In economic theory and social psychology, motivations are typically considered to either be intrinsic or extrinsic [26]. In the case of intrinsic motivation, the benefits of undertaking an action accrue directly and immediately to the individual in a form of enjoyment-based or obligation-based satisfaction. Extrinsic motivations on the other hand involve an indirect reward, such as money. While the terms “motivation” and “incentive” are sometimes used interchangeably, a distinction is sometimes made where motivations are considered more intrinsic while incentives are more extrinsic to individuals [27]. There is an interplay between motivations and incentives towards influencing individuals decisions and actions, where phenomenon such a “crowding out” can occur due to extrinsic incentives eroding intrinsic motivations [26]. There has been research undertaken that explores incentivizing individual’s participation in data related activities, such as participating in online surveys [28], however there is a gap in literature on incentives for data contribution in the context of the sustainable development agenda.

This research investigated the extent to which the participants would be incentivized to record and share their data, both personal health data and information that they had access to, such as water and air quality data (Figure 3).

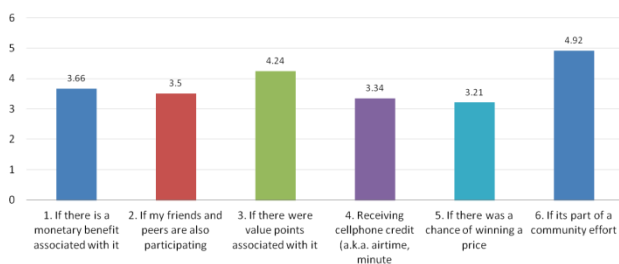


Figure 3 - Motivations and incentives for data collection and information sharing

The motivations and incentives for recording and sharing of social indicators data are observed to be both intrinsic and extrinsic. The majority of the respondents highlighted the intrinsic socially framed motivation of undertaking the monitoring and sharing “*if it’s part of a community effort*”. The extrinsic motivations of “*monetary benefit*”, “*cellphone credit*” and “*a chance of winning a price*” are observed to be the least influential for motivating the participants towards monitoring and sharing of data.

3.3. Data collection and monitoring

From the *preparation stage*, once individuals have established the motivations for monitoring and collecting data, and having identified the relevant tools, they progress to undertake the actual collection and recording of the data.

From the participants in the survey the use of fitness / activity trackers is observed at 18.75% for *daily use*, 6.25% for *weekly use*, 2.08% for *monthly use*, 10.42% for *seldom use*, and 62.5% for *never used*. While the use of fitness / activity trackers is specifically for tracking personal health metrics, increasingly individuals are also making use of the smart phone apps and smart watches for monitoring and tracking of personal health metrics. From the survey 87.5% of the people indicated that they never used smart watches, while 10.42% and 2.08% use their smart watches daily and weekly respectively. The use of smart phones is high as expected at 95.83% of individual claiming daily use, 2.08% for seldom use and 2.08% for having never used a smart phone before. This captures the general use of smart phones by the participants and not just specifically for self tracking and monitoring.

3.4. Data utility, sharing, and social sense-making

The conversion of monitored and collected individual data into developmental action is decomposed by Li *et al* into two distinct processes of *integration* and *reflection* [15]. Integration primarily consists of processing and manipulating the data in order to feed into the next process of reflection. Reflection as a cognitive technique for meaning and sense making has been studied and expounded on in various fields including education [29], psychology, and human computer interaction [30]. In the field of Personal Informatics, recent work has explored supporting reflection and behavior change through sharing of personal data [31], [32], and through social sense-making [16].

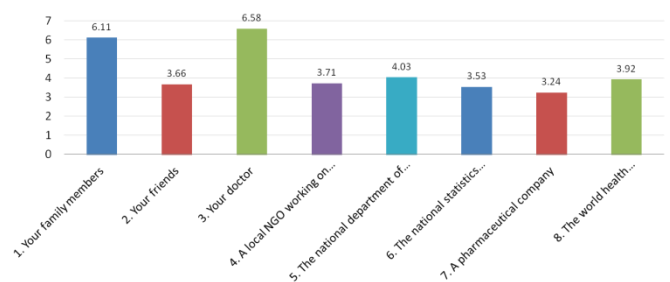


Figure 4 – Sharing of personal health information with different stakeholders

In the context of the sustainable development data ecosystem or that of future data-driven societies, the sharing of personal data needs to be considered not only within individuals’ personal social circles but also with other stakeholders within the wider data ecosystem. As such this research explored the participants attitudes

towards sharing of their personal data within the sustainable development data ecosystem, exploring both the willingness of the participants to share their data with specific stakeholders, as well as the factors that would inform their willingness to share (or not to share) their personal data.

Using a continuous scale from 1 to 7 at “low willingness to share” and “high willingness to share” respectively, the participants are most (mean 6.58) willing to share their personal health data with their doctors, and least (mean 3.24) willing to share their data with pharmaceutical companies (Figure 4).

Further analysis was undertaken to understand how the participants’ attitudes towards sharing their personal data correlates across the different stakeholders. For this analysis a Spearman correlation matrix was derived and subsequently agglomerative hierarchical clustering (complete linkages method), using the Euclidian distance between the correlation scores, undertaken to understand the main clusters for the different stakeholders (Figure 5).

Clustering of Data Sharing entities/individuals

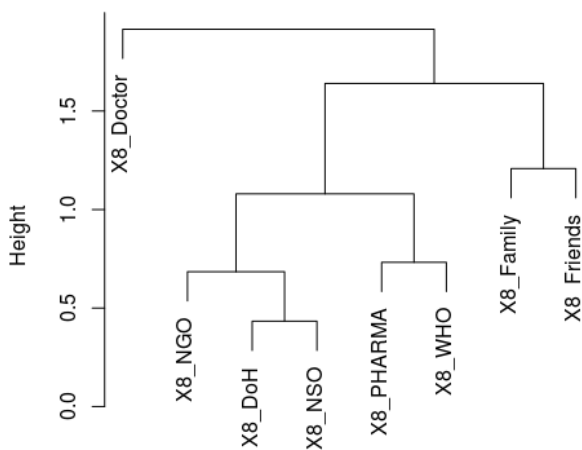


Figure 5 - Data sharing entities clustering

From this analysis three primary clusters of stakeholders are noted (cutting the dendrogram in Figure 5 at the height of 1.5) and these are: *Cluster 1* - individual’s doctor; *Cluster 2* – NGO working on health issues, national Department of Health, National Statistics Department, a pharmaceutical company, and the World Health Organization; and *Cluster 3* - family members and friends. Clear characterization emanates from these clusters, based on the relationship between the stakeholders and the individual, and the nature of the utility that accrues to the individual, as follows:

- Cluster 1 is a stakeholder that is able to use the shared personal health data towards the provisioning of an immediate health service, wherein the data can be used for health monitoring

or to inform diagnosis of medical ailments. This therefore represents a direct (and future) benefit to the individual.

- Cluster 2 are organizational entities within the wider health sector with clear sub-clusters of governmental, non-governmental, and international/multinational organizations. The benefits that accrue to the individual from sharing data with these entities are indirect and generally not immediate.
- Cluster 3 are entities with a high social proximity to the individual, where the sharing of the personal health data could be more towards the associated social benefits, such as sense-making [16], and social support [32].

The findings from the survey are that these initial clusters of stakeholders not only highlight the need for differentiated data sharing arrangements with entities within the data ecosystem, but also point to the willingness of the participants to consider sharing their data across the ecosystem.

The advent of social media has meant that individuals are increasingly used to sharing their data. However a lot of the voluntary and active sharing of data is typically in the context of the social networks that the individuals have. Currently a lot of individuals’ data is collected, without their full awareness and complicity, from individuals’ digital traces and from tracking of individuals online through surveillance. Solove suggest a taxonomy that identifies four basic activities around which violation of individuals’ privacy violation can occur, and these are [33]: *information collection* – in which activities such as surveillance and interrogation can be employed (by data holders) to gather information about individuals (the data subjects); *information processing* – through the processing of the data involving aggregation and analysis; *information dissemination* – encapsulates activities such as breach of confidentiality, disclosure, exploration, blackmail and distortion, which would contribute towards violating individuals’ privacy; and lastly *invasion* – which is not about individuals’ information but rather about violating privacy associated with individuals personhood. The contention and opposition to the practice of mass collection of individuals’ data is growing, and increasingly there is push back from civil society to have increased privacy and confidentiality of their data, to have control over who collects the data, what data is collected, and how the data is used (i.e. increased data *legibility* [17]).

As such, beyond just understanding the participants’ attitudes towards sharing data with specific stakeholders, this research also sought to investigate the factors that affect the willingness of participants to share data, based on 10 pre-selected factors and an evaluation using a continuous scale of between 1 (for *low influence*) and 7 (for *high influence*).

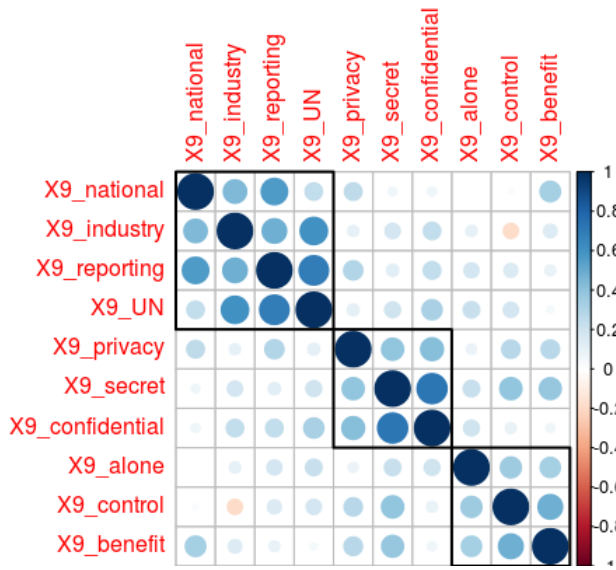


Figure 6 - Correlation matrix of influence factors

The factor that scored the highest (mean 6.16) on influencing individuals to share (or not to share) their personal health information is *the privacy of the information*, with the lowest being the individuals’ *right to be left alone*. A subsequent clustering of these factors was undertaken, using hierarchical clustering based on the Euclidian distance off the Spearman correlation matrix of the factors (Figure 6), and using the complete linkages agglomeration method to find similar factor clusters (Figure 7).

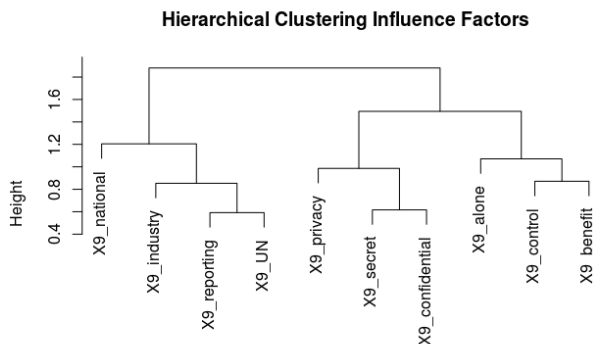


Figure 7 - Hierarchical clustering of the factors

Three key categories emanate from the clustering exercise undertaken, and these have been labeled: 1) the *intrinsic value based factors* – which reflects internally held individuals beliefs about the values and attributes of the data; 2) *intrinsic personal factors* – which are factors centered around the individual and their interaction with their data; and 3) *extrinsic factors* – these are factors that are associated with externally-oriented use of the individuals’ data. Table 1 lists these categories and the associated mean scores for the factors and across the categories.

Table 1 - Influence factor clusters and means

Key	Description	Mean
Intrinsic value based		5.64
X9_confidential	The confidentiality of the information	5.66
X9_secret	The secrecy of the information	5.11
X9_privacy	The privacy of the information	6.16
Intrinsic personal		4.7
X9_benefit	That the information is used for my direct benefit and wellbeing	5
X9_control	The need to control access to personal information	5.26
X9_alone	My right to be left alone	3.84
Extrinsic		4.48
X9_UN	Assisting the UN to compare quality of health across different countries	4
X9_reporting	Assisting government to report on quality of health in the country	4.34
X9_industry	Contributing to improving the health industry through better medicines	4.84
X9_national	That the information is used for national health policies	4.74

On average the factors that the participants point to as having the highest influence on their willingness to share their personal health information are those associated with the privacy, confidentiality and secrecy aspects of the personal data. These three concepts are of course tightly coupled and represent varied conceptualizations of the notion of privacy [34]. The factors associated with the external use (e.g. contribution to the national or UN-level social indicators efforts) of the individuals data in general are of lower influence on the participants towards them sharing their data.

The two broad findings, on the clustered differentiation of the entities within the data ecosystem that the participants are happy to share their data with, as well as on the clusters of factors that influence individuals to collect and share their personal data, give further insights on the strategies for engagement of individuals within the sustainable development data ecosystem.

4. DISCUSSION

Mortier *et al* have suggested that legibility, agency and negotiability are three of the core themes for consideration in Human Data Interaction [17]. These three themes are associated with amplifying the capability of individuals to engage meaningfully and actively with their data. This research has engaged with these concepts from the perspective of small data to support the individuals’ participation within the sustainable development data ecosystem. The formulation of small data for development in section 2, which is framed around the Capabilities Approach, emphasizes individuals’ freedom and agency to use data to achieve the development ends that they desire. Recognition is given to the fact that data, including small data, has the potential both to restrict and to expand individuals’ substantive freedoms towards their development and wellbeing, as such the individuals participation and engagement within the 21st century data-driven societies needs to be critically considered from this

perspective of empowerment and expansion of substantive freedoms.

In the use of data towards development and wellbeing, individuals employ both reflection [31] and social sense-making [16] as techniques towards converting the data into actionable insights. Beyond sharing data within their immediate social circle, there are opportunities for increased effectiveness (e.g. for better understanding of various social phenomena) within the sustainable development ecosystem from sharing of data widely with other stakeholders. The findings from the participants in this research identify three clear groups of stakeholders for consideration in data sharing. The first is associated with entities that are able to provide immediate benefit and service to the participants (in this case, the doctors), the second represents the close social relationships, and the last broadly represents organizational entities. These clusters not only highlight the need for differentiated data sharing strategies, policies, and systems, but also provide an initial encapsulation of the different requirements for data sharing.

Beyond the identification of the participants' attitudes towards sharing personal data with specific stakeholders, the investigation of the factors that affect participants' willingness to share their data reveals the high importance of data privacy, secrecy and confidentiality. This suggests a need for explicit data security mechanisms within the systems that are implemented for processing of personal data. While the factors that are associated with external use of personal data had a low influence on individuals collecting and sharing of their data, it still remains that at a mean of 4.48 (out of 7) these factors are still taken into consideration at some level by the individuals. The three clusters that emanate from this investigation also provide an encapsulation of concerns to inform the development of relevant system policies or functionalities.

Research has been undertaken to explore the role and effectiveness of incentives for influencing individuals towards data sharing, in the context of participating in web surveys [28]. The findings from the participants in this research, with regards to factors that would motivate their collection and sharing of data is that both intrinsic motivations and extrinsic incentives would be considered and relevant, however the indication of the highest incentive as "*if it's part of a community effort*" puts even greater emphasis on the potential of intrinsic obligation-based motivations. Linking this motivation with the observations around social sense-making also suggests the potential of collective (e.g. at a community level, or social grouping level) engagement in data for sustainable development initiatives. This would thereby be towards not only encouraging social sense-making, and providing incentives, but also expanding the collective capabilities of the individuals [35].

5. INFORMING THE IMPLEMENTATION OF SMALL DATA TOOLS

This research is undertaken in the context of a larger exploration of the role of Information and Communication Technologies to support individuals and community based organizations towards the achievement of the sustainable development goals. The findings from this research stand to inform the development of the associated small data tools towards supporting individuals' use of data for their health and wellbeing (SDG3) as follows:

- The tools need to allow for a targeted, differentiated and secure sharing of data with specific individuals and stakeholders.
- There is a requirement for data provenance preservation, associated with expression of concern by the participants to have control over their data, and also to maintain the confidentiality of the data.
- The general high level of use of smart-phones presents an opportunity for varied kinds of data (e.g. recorded, derived, and observed) to be processed, taking into consideration the privacy concerns.
- While it might be necessary to incorporate mechanisms for extrinsic rewarding within the tools for user contribution of data, there is also a potential to facilitate community building and collective engaging between the users of the tools.

6. CONCLUSION

The ubiquity and pervasiveness of data is growing and the use of data to drive efficiencies, achievement of goals, and improved decision making is increasingly permeating all societal domains. This data revolution is characteristic of the 21st century data-driven society and it presents numerous opportunities and risks not only to individuals, but also to organizations and governments. Extensive research has been undertaken that explores these opportunities and challenges, in particular within the context of the 2030 Agenda for Sustainable Development goals.

This paper has explored the engagement of individuals within the sustainable data ecosystem, by investigating through a non-probabilistic survey and presenting through a descriptive analysis: the awareness levels of the participants with regards to the Sustainable Development Goals; their attitudes and perceptions around monitoring of social indicators; key considerations associated with data ownership, privacy and confidentiality of data, as well as sharing of data within the data ecosystem. While the results from the survey are specific to the participants and cannot be immediately generalized to wider populations (due to the non-probabilistic sampling), the findings highlight important considerations that not only stand to inform the development and implementation of further small data solutions in this research, but also contribute to the general discussions around data ownership, data sharing, data provenance, and incentives and motivations for sustainable development data.

The role of individuals in data-driven societies is paramount and the necessity to support individuals' active involvement and participation in the associated data ecosystem is critical. This paper has presented research that's part of an ongoing effort towards ensuring the benefits of the data revolution accrue to all, without leaving anyone behind.

REFERENCES

- [1] C. Hsinchun, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Q.*, vol. 36, no. 4, 2012.
- [2] A. McAfee, E. Brynjolfsson, and T. Davenport, "Big data The management revolution," *Harv. Bus. Rev.*, vol. 90, no. 10, pp. 61–67, 2012.
- [3] C. Loebbecke and A. Picot, "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda," *J. Strateg. Inf. Syst.*, vol. 24, no. 3, pp. 149–157, 2015.
- [4] J. Höchtl, P. Parycek, and R. Schöllhammer, "Big data in the policy cycle: Policy decision making in the digital era," *J. Organ. Comput. Electron. Commer.*, vol. 26, no. 1–2, pp. 147–169, Apr. 2016.
- [5] K. Gang-Hoon, T. Silvana, and C. Ji-Hyong, "Big-data applications in the government sector," *Commun. ACM*, vol. 57, no. 3, pp. 78–85, 2014.
- [6] J. Rooksby *et al.*, "Personal tracking as lived informatics," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2014, pp. 1163–1172.
- [7] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "LifeLogging: Personal Big Data," *Found. Trends? Inf. Retr.*, vol. 8, no. 1, pp. 1–125, 2014.
- [8] A. Greenfield, *Everyware : the dawning age of ubiquitous computing*. New Riders, 2006.
- [9] R. Kitchin, "The real-time city? Big data and smart urbanism," *GeoJournal*, vol. 79, no. 1, pp. 1–14, Feb. 2014.
- [10] General Assembly, United Nations, "Transforming our world: The 2030 agenda for sustainable development," <https://sustainabledevelopment.un.org/content/documents/7891Transforming%20Our%20World.pdf>, no. 1, pp. 1–5, 2015.
- [11] E. Letouzé, *Thoughts on Big Data and the SDGs*. United Nations Sustainable Development, 2012.
- [12] *A World that Counts: Mobilising the Data Revolution for Sustainable Development*. Economic Commission for Latin America and the Caribbean (ECLAC), 2014.
- [13] E. Letouzé, *Big Data for Development: Challenges & Opportunities*. United Nations Global Pulse, 2012.
- [14] SDSN, *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development*. Sustainable Development Solutions Network, New York, USA, 2015.
- [15] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, 2010, p. 557.
- [16] A. Puussaar, A. Clear, and P. Wright, "Better together: Reciprocal sharing and social sensemaking of data," in *Workshop on "Quantified Data and Social Relationships" - CHI2017*, 2017.
- [17] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, and J. Crowcroft, "Human-Data Interaction: The Human Face of the Data-Driven Society," *SSRN Electron. J.*, Oct. 2014.
- [18] D. Estrin, "Small data, where n = me," *Commun. ACM*, vol. 57, no. 4, pp. 32–34, Apr. 2014.
- [19] M. Lindstrom, *Small Data: The Tiny Clues That Uncover Huge Trends*. St. Martin's Press, 2016.
- [20] M. Best, "Small Data and Sustainable Development," in *International Conference on Communication/Culture and Sustainable Development Goals : Challenges for a new generation*, 2015.
- [21] M. Thinyane, "Investigating an Architectural Framework for Small Data Platforms," in *Data for societal challenges - 17th European Conference on Digital Government (ECDG 2017)*, 2017, pp. 220–227.
- [22] A. Sen, *Development as freedom*. Oxford University Press, 1999.
- [23] P. Freire, *Pedagogy of hope : Reliving Pedagogy of the oppressed*. Bloomsbury Academic, 2014.
- [24] S. Stals, M. Smyth, and O. Mival, "Sharing and exploring Quantified-Self Data on In-Place experiences and emotions," in *ACM CHI 2017 - Quantified Data and Social Relationships*, 2017.
- [25] P. Rankin and E. Ngwenya, *Linking Awareness and Likely Achievement of Global Millennium Development Goals: Evidence and Implications for Advocacy from 46 Countries*, vol. 7, no. 1. Common Ground Publishing, 2011.
- [26] N. Teles, "Motivations and Incentives: from the 'crowding-out effect' to 'peer-production'." 2007.
- [27] D. M. Kreps, "Intrinsic Motivation and Extrinsic Incentives," *Am. Econ. Rev.*, vol. 87, pp. 359–364.
- [28] C. Cobanoglu and N. Cobanoglu, "The effect of incentives in web surveys: application and ethical considerations," *Int. J. Mark. Res.*, vol. 45, no. 4, pp. 475–488, 2003.
- [29] J. Dewey, "Experience and education," in *The Educational Forum*, 1986, vol. 50, no. 3, pp. 241–252.
- [30] L. Hallnas and J. Redstrom, "Slow Technology? Designing for Reflection," *Pers. Ubiquitous Comput.*, vol. 5, no. 3, pp. 201–212, Aug. 2001.
- [31] R. Fleck and D. Harrison, "Shared PI: Sharing personal data to support reflection and behaviour change," in *Workshop on " Beyond Personal Informatics: Designing for Experiences of Data" - CHI2015*, 2015.
- [32] B. Ploderer, W. Reitberger, H. Oinas-Kukkonen, and J. van Gemert-Pijnen, "Social interaction and reflection for behaviour change," *Pers. Ubiquitous Comput.*, vol. 18, no. 7, pp. 1667–1676, Oct. 2014.
- [33] D. J. Solove, "A Taxonomy of Privacy," *Univ. PA. Law Rev.*, vol. 154, no. 3, p. 477, Jan. 2006.
- [34] D. J. Solove, "Conceptualizing Privacy," *Calif. Law Rev.*, vol. 90, no. 4, p. 1087, Jul. 2002.
- [35] S. S. Ibrahim, "From Individual to Collective Capabilities: The Capability Approach as a Conceptual Framework for Self-help," *J. Hum. Dev.*, vol. 7, no. 3, pp. 397–416, Nov. 2006.