Towards Informing Human-centric ICT Standardization for Data-driven Societies

Mamello Thinyane

United Nations University, Institute on Computing and Society (UNU-CS), Macau – SAR
E-mail: mamello@unu.edu

Abstract

Information and communication technology (ICT) standards play a crucial role towards harnessing technological developments and shaping the technology landscape. ICT standards development is largely driven by standards developing organizations or standards setting organizations that represent and are informed by the perspectives and interests of the ICT private sector and the public sector. Thus the shaping of the standards is mostly driven by the business, technical, and legal impetus, towards facilitating greater market influence, coordination, integration, interoperability, and legal conformity. The technocentric, business-focused perspective to ICT developments, and corollary to ICT standards development, is distinct from and can be orthogonal to the human-centric perspective that elevates the role and centrality of the human concerns over the technology and business concerns. Notwithstanding the crucial role of business and technology, it remains that at the centre of the 21st century data-driven societies are individuals and end-users who are the primary actants and agents within the technology and data ecosystems. This gives motivation for infusing the human perspectives and values into technology development as well as into ICT standards development.

This is more pronounced for cases where the business and technocentric interests are at odds and opposed to the human interests, such as, the need for increased datafication to support Big Data developments versus the need for individuals' privacy preservation. In this research, which is framed through a case study of personal health informatics in the context of sustainable development (i.e., sustainable development goal on "health and wellbeing" – SDG3) indicators monitoring, we have undertaken a survey to investigate the human-centric values and attitudes associated with the collection and use of personal data. From this inquiry, the paper highlights and surfaces: individuals attitudes and perceptions around monitoring of social indicators; key considerations associated with data ownership, privacy and confidentiality of data, as well as sharing of personal data within the data ecosystem. The paper then discusses how these findings could inform and be infused into the development of technology artefacts and standards, towards a realization of more human-centric data-informed societies.

Keywords: Data-driven societies, standardization, human-centric ICT.

1 Introduction

The ensuing fourth industrial revolution is poised to transform global societies in many fundamental ways through the proliferation of ubiquitous "every-ware" digital devices [1], the permeation of advanced technologies (e.g., Artificial Intelligence, Machine Learning, Robotics, Quantum computing, Neurocomputing) into various societal domains [2], increased datafication [3], and the societal evolution towards the "infosphere" [4]. Currently, advancements are 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) [5] and business management [6, 7]. Similarly the public sector use of data can be noted towards improved policy development [8], 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" [9]. At the global level, there has been recognition of the role of data towards supporting the achievement of the global development imperatives articulated in the 2030 United Nations (UN) agenda for sustainable development. The UN Sustainable Development Goals (SDGs) resolution highlights, within Goal 17, technology as an explicit means of implementation (MoI) towards the achievement of the goals [10].

Further, the critical role of the indicators data within the SDG programme has also been well articulated [11, 12]. The focus within the social indicators monitoring domain is largely about using data for the purposes of planning, monitoring and evaluation (PM&E) at the national and sub-national levels. Extensive research has been undertaken exploring the potential for data in this regard (e.g., [12–14]).

Beyond the use of data by the corporate and the public sector, there has been increasing efforts and research around the use of data by individuals towards informing their wellbeing goals and imperatives. The growing field of personal informatics represents this interest and focus on the use of data for the ultimate utility that accrues towards the individuals [15]. Thus the ubiquity of data in society is seen not only in the increased data use by organizations but also by individuals in areas such as quantified-self and life-logging – which are associated with the proliferation of activity trackers and mobile devices [16, 17].

The new data and technology advancements, which eventuate from the current trajectory of the global socio-technical systems, hold both great promise and great peril. Notwithstanding the critical role of technology in modern-day societies, it remains that at the centre of data-driven societies are individuals and end-users who not only generate data, but who are also the key actants in these data ecosystems. It is thus the role and responsibility of the global community to harness and shape these advancements for the global good, through the use of relevant instruments (e.g., legal frameworks and standards) that encapsulate the core human concerns and aspects of the "desired future". ICT standards development has to a large extent been driven by standards developing organizations (SDOs) or standards setting organizations (SSOs) that typically represent and are informed by the perspectives and interests of the ICT private sector and the public sector. Thus the shaping of the standards has been driven by the business, technical, and legal impetus, towards facilitating greater market influence, coordination, integration, interoperability, and legal conformity. This technocentric, business-focused perspectives to ICT developments, including to standardization efforts, are distinct and at times stand at odds to the more human-centric perspectives that elevate the role and centrality of the individual and human concerns over the technology and business concerns. It is this notion of human-centricism that is articulated in this research to suggest that technological developments, and the associated standardization processes, should be informed by the considerations that take into account not only the technical and economic considerations, but also more significantly the human concerns and developmental aspirations.

Mortier et al. have framed more formally the significance of the human concerns within the interaction dynamic between humans and data through the discursive construction of Human Data Interaction (HDI), which they define as being constituted of three key themes and domains: legibility, agency and negotiability [18]. *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 substantive 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 provide a suitable initial framing of the key concerns and considerations for the engagement of individuals in data-driven societies.

This research investigates the human-centric considerations in personal data collection and use, and suggests how these could inform the development of relevant technologies and standards. The research is framed through a case study on personal health informatics in the context of global sustainable development goal on "health and wellbeing" (i.e., SDG3). Section 2 discusses the small data perspective which has been adopted in this research towards understanding human-centric data valorisation. This is followed in Section 3 by a presentation of the personal health informatics case and the methodological design of the research. The findings from the research, highlighting the participants' attitudes and practice towards collection and sharing of social indicators data, are discussed in Section 4. Section 5 provides a discussion on how the findings could inform the development of technology artefacts and standards, and Section 6 concludes the paper.

2 Small Data and Human-centric Data Valorisation

While there is a tendency to consider data generally and broadly, different types (e.g., Big Data, open data, citizen generated data, and 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 a key data perspective in this research, has in recent years been conceptualized and defined in various very distinct terms. Small data has been articulated: as the digital traces around an individual [19]; as data from ethnographic and human-centric investigation of a social phenomenon [20]; and as an approach to analyzing data at the same unit of sampling [21]. In this research we adopt the characterization of small data 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" [22]. This definition focuses on the individual and their associated interactions within the data ecosystem. From this perspective, the emphasis in the consideration of the use of data is about connecting individuals with the relevant data (including personal data, big data, open data, real-time data) towards their wellbeing and sustainable livelihood. Thus framed more succinctly and in this case, small data is an approach towards a human-centric data valorisation.

The value proposition for data 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
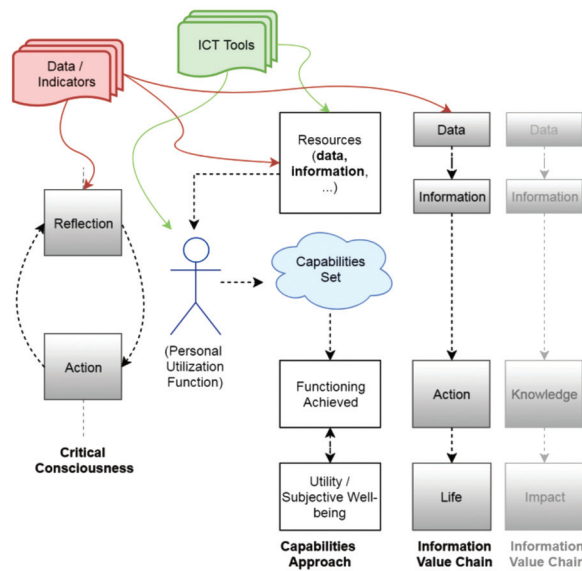

Figure 1   Expounded information value chain.

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 [23]. This process can be further enunciated by identifying one of the mechanisms through which data is utilized, which is for facilitating individuals' reflection and critical awareness of their own life and circumstances [24].

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 individuals actively chose a particular course of action informed by the reflection on the data collected.

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 the 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. Understanding the dynamics and the issues around individuals' engagement within the data ecosystem is essential towards informing the development of the supporting technologies and the associated standards. Some of these pertinent issues are explored and discussed in the next sections.

3 The Personal Heath Informatics Case

In recent years there has been an increase in the use of data for monitoring individuals' health and wellbeing, which has been fuelled in part by the developments in sensors technologies. The proliferation of mobile devices, smart wearable devices and health monitoring consumer devices that bundle these sensors means that more people are increasingly using data for personal health informatics [15]. Personal health data has a number of key characteristics that are relevant for the inquiry in this research – it is personal, private, and

confidential; and its primary utility is towards monitoring and informing the wellbeing of individuals. Notwithstanding the tight coupling of the utility of the data to individuals, there are cases of the use of this data beyond personal and individual reasons. For example, the data can be used, under relevant transformations, towards informing the monitoring of health indicators for a community. In the context of the SDGs there is increasingly a general interest in exploring the use of citizen-generated data, as well the personal data from Big Data digital traces, towards informing the monitoring of SDG indicators through proxy indicators. Thus the use of personal data needs to be considered from the context of the wider data ecosystem, with an understanding of the various data flows and dependencies within the data ecosystem. This is the context within which the investigation in this research has been formulated.

The methodological design of this research is framed around the following lines of inquiry with regards to individuals' use of personal health data – to understand: the reasons and motivations for data collection and monitoring; the current practice around personal health informatics; and the attitudes and values around data sharing, and social sense-making. Towards this end, a semi-structured survey instrument was developed and has been used for data collection from virtual participants. The survey consisted of 14 questions: two closed questions on demographics, five closed questions on current personal health informatics practice, two closed questions on sustainable development goals, three closed questions on data sharing, and one open-ended question to discover individuals' current practice. The survey was framed to inform a non-probabilistic descriptive understanding of these issues for the specific individual participants in the survey, without seeking to make wider population group generalizations. This survey instrument was administered online and the recruitment of the participants was done through email invitations, social media, virtual snowballing techniques, and paid participant recruitment through the survey portal.

4 Research Findings

In total 1,115 people started the survey and 981 of those completed the survey. The findings discussed are only for the complete responses. This paper discusses the findings specifically for the questions addressing the current practice of personal health informatics, and the attitudes and values around data ownership and sharing. 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].
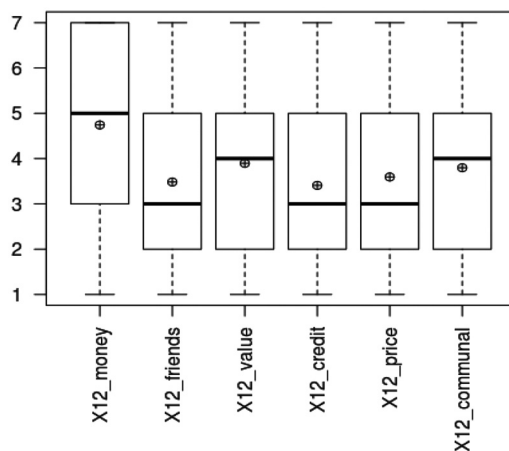
4.1 Motivations and Incentives for Data Use

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 behaviours [25]. Rooskby et al. identify what they term "styles" of personal information tracking which are [16]: *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.

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 activities, such as participating in online surveys [28].

Figure 2 Motivations and incentives for data collection and information sharing.

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

As notable in Figure 2, the motivations and incentives for recording and sharing of social indicators data are observed to be both intrinsic and extrinsic, with the extrinsic "*monetary benefit*" *(X12_money)* being the highest scored incentive. The respondents also ranked highly the intrinsic socially framed motivation of undertaking the monitoring and sharing "*if it's part of a community effort*" *(X12_communal)* as well as the other extrinsic motivations of "*if there were value points associated with it*"*(X12_value).*

4.2 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 20% for *daily use*, 7% for *weekly use*, 2% for *monthly use*, 4% for *seldom use*, and 68% 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 85% of the people indicated that they never used smart watches, while 11% and 3% use their

smart watches daily and weekly respectively. The use of smart phones is high as expected at 81% of individual claiming daily use, 2% for seldom use and 13% 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.

4.3 Data Utility, Sharing, and Social Sense-making

The conversion of monitored and collected individual data into 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 behaviour change through sharing of personal data [31, 32], and through social sense-making [33].

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 between 1 and 7 (with labelled end-points of *"low willingness to share"* and *"high willingness to share"* respectively), the participants are most (i.e., mean 5.99) willing to share their personal health data with their doctors (i.e., X8_Doctor), followed by the willingness to share health and wellbeing data with their families. The stakeholder that participants are least (mean 3.38) willing to share their data with are pharmaceutical companies (i.e. X8_PHARMA) (Figure 3).

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 4).
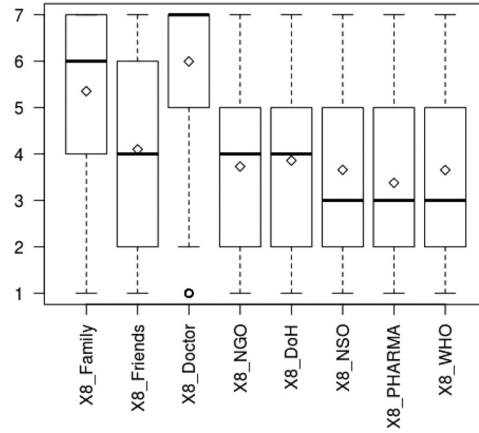
Figure 3   Sharing of personal health information with different stakeholders.
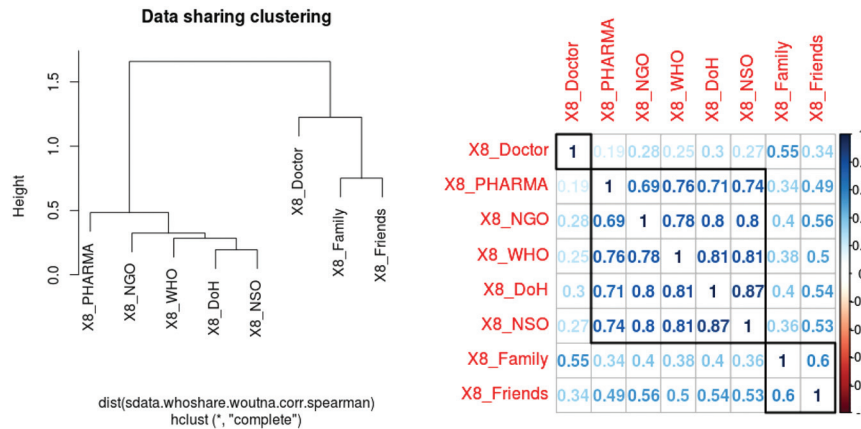
Figure 4   Data sharing entities clustering.

The correlation analysis and the subsequent clustering identified three clusters of entities and stakeholders for data sharing with individuals. Table 1 lists the entities within each of the three clusters that have emanated from the cluster analysis.

The first cluster consists of a single stakeholder, the individuals' doctor or medical provider (i.e., X10_Doctor). While not established explicitly in this research, it is evident that this cluster represents the stakeholder who is able to provide a health service for the individuals with the benefit accruing directly to the individual. The sharing of health and wellbeing data with this specific

Table 1   Data sharing entities clusters

Cluster	Entities (*mean*)
Cluster 1	● X8_Doctor – Your doctor *(5.99)*
Cluster 2	● X8_PHARMA – A pharmaceutical company *(3.38)*
	● X8_NGO – A local NGO working on health issues *(3.73)*
	● X8_WHO – The World Health Organization *(3.65)*
	● X8_DoH – The Department of Health *(3.86)*
	● X8_NSO – The National Statistics department *(3.66)*
Cluster 3	● X8_FAMILY – Your family members *(5.36)*
	● X8_FRIENDS – Your friends *(4.102)*

stakeholder therefore contributes to the direct enjoyment and achievement of the individuals' health and wellbeing goals.

The second cluster from the analysis comprises of pharmaceutical companies, NGOs working on health related issues, the World Health Organization, national department of health, as well as the national statistics department. The stakeholders in this cluster are characterized by being organizations within the wider health sector including governmental, non-governmental as well as international/multinational organizations. As far as sharing data with these organizations are concerned, it can be noted that the benefit that accrues to the individuals would mostly be indirect and not immediate. For example, sharing data with the national statistics department, could inform the reporting on national health and wellbeing and subsequently the development of relevant health policies, which would in time be of benefit to the individual. In the case of sharing data with the local NGO working on health related issues, there are possible scenarios where the resultant benefit could accrue more directly to the individuals.

The third cluster identified from the analysis above comprise of entities who share a high social proximity and relational coupling with the individual. The sharing of data with these stakeholders is typically for the purposes of sense-making [33] and social-support [34].

The three clusters of the entities for data sharing show a statistically significant difference of means (alluded to in Figure 5) with Wilcoxon rank sum test (a.k.a., Mann-Whitney test) results as shown in Table 2. From this we can conclude that the individuals' attitudes and perceptions regarding data sharing with the different stakeholders within the data ecosystem vary significantly per group of stakeholders. Furthermore, sharing data with stakeholders that provide direct benefit and that support the direct enjoyment of the health and wellbeing goals (i.e., cluster 1) has the highest ranking, followed by sharing data with friends and family (i.e., cluster 3) typically towards social sense-making
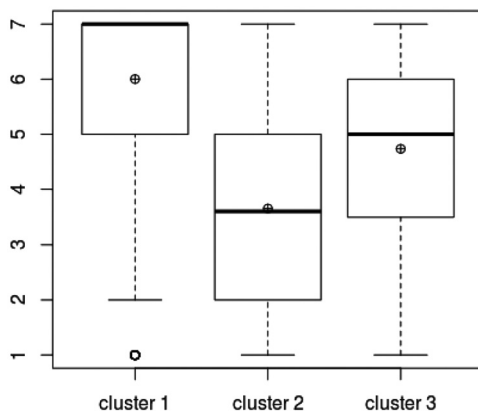
Figure 5 Data sharing clusters means.

Table 2 Wilcoxon tests on the clusters

	Cluster 1	Cluster2
Cluster 3	V = 257640, p-value < 2.2e-16	V = 56800, p-value < 2.2e-16
Cluster 2	V = 354620, p-value < 2.2e-16	

and social support. Lastly, sharing data with cluster 2 stakeholders has the lowest ranking – this finding has implications for further consideration on engaging individuals and community level actors in the monitoring of personal social indicators for official purposes (e.g., planning monitoring and evaluation).

4.4 Factors Affecting Data Sharing

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)*. 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 consent, from individuals' digital traces and from tracking of individuals online. Solove has suggested a taxonomy that identifies four basic activities (i.e., information collection, information processing, information dissemination, and invasion) around which violation of individuals' privacy can occur [35]. Further, he suggests
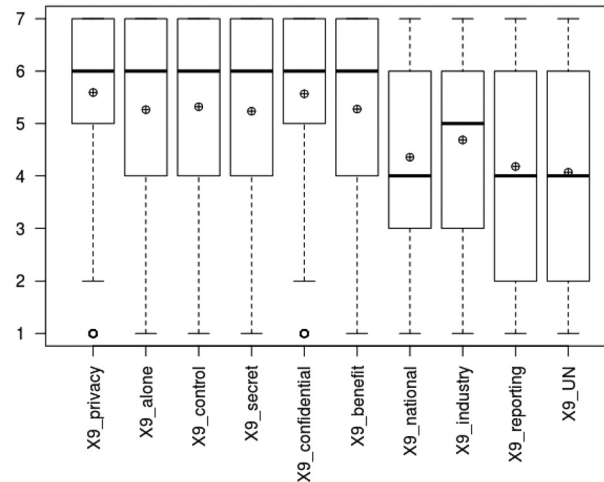
Figure 6   Boxplot for factors influencing willingness to share data.

a bottom-up perspective to thinking about privacy, taking into consideration the contextual factors and the dynamic nature of privacy, instead of defining privacy in singular, universal and abstract terms [36]. He notes the following six conceptualizations of privacy:

1. The right to be left alone
2. The ability to shield oneself from unwanted access by others
3. Secrecy and the concealment of certain aspects from others
4. Control over personal information
5. Personhood and the protection of one's personality, individuality and dignity
6. Control over one's intimate relationships or aspects of life

These six concepts of privacy have informed five of the ten factors explored in this section, with some concepts being merged and rephrased with emphasis on the informational and data related aspect of the concepts – for example concepts 4 ("*control over personal information*") and 5 ("*personhood and protection of one's personality, individuality and dignity*") have been merged into the "*need to control access to personal information*" (i.e., X9_control) option in the survey.

The overall choice of factor options for this question was informed by the following considerations: factors associated with privacy, factors associated with personal benefit from sharing data, and factor associated with external
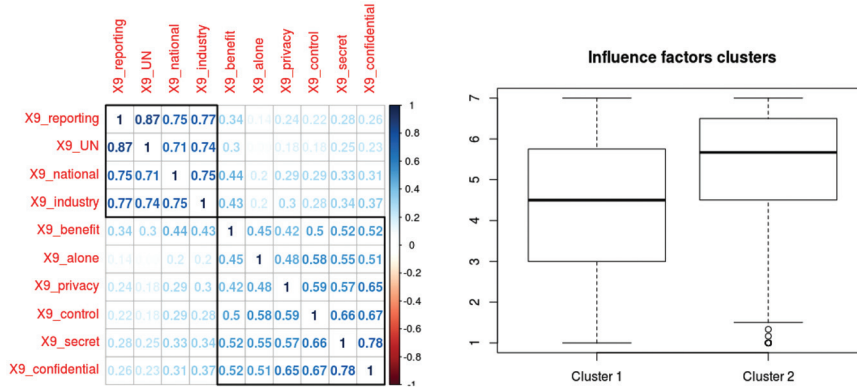
Figure 7 Correlation matrix of influence factors with the boxplot of the clusters.

Table 3 Clusters of factors affecting data sharing

Cluster	Factors
Cluster 1	• X9_reporting: Assisting the government to report on quality of health in the country • X9_UN: Assisting the United Nations to compare quality of health across different countries • X9_national: That the information is used for national health policies • X9_industry: Contributing to improving the health industry through better medicines
Cluster 2	• X9_benefit: That the information is used for my direct benefit and wellbeing • X9_alone: My right to be left alone • X9_privacy: The privacy of the information • X9_control: The need to control access to personal information • X9_secret: The secrecy of the information • X9_confidential: The confidentiality of the information

use within the wider data ecosystem – these options evaluated to a strong internal consistency with Cronbach's index at 0.88.

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 of personal data, in its varied conceptualizations as articulated by Solove [36]. 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 data ecosystem.

5 Informing the Formulation of Artefacts and Standards

Mortier et al. have suggested that legibility, agency and negotiability are three of the core themes for consideration in Human Data Interaction [18]. 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 through the adoption of the small data perspective to support the individuals' participation within the data ecosystem. The formulation of small data for human-centric data valorisation in Section 2, which is framed around the Capabilities Approach, emphasizes individuals' freedom and agency to use data to achieve the life goals and ends that they desire. Recognition is given to the fact that 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 for wellbeing, individuals employ both reflection [31] and social sense-making [33] as techniques towards converting the data into actionable insights. Beyond sharing data within their immediate social circles, there are opportunities for increased individual and collective utility from sharing data widely with other stakeholders. The findings from the participants in this research identify three distinct 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 broadly represents external organizational entities, and the last represents the close social relationships. 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 importance of

personal data privacy. 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 these factors are still taken into consideration at some level by the individuals. The two clusters that have been identified in this investigation also encapsulate the human-centric concerns that can 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, for example, 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 motivation as an extrinsic monetary motivation, suggest an economic valuation mechanism that individuals employ to determine their willingness to collect and share data. 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 [34].

Examples of the ways through which the findings from this research could inform the development of technology artefacts and the standardization processes are as follows:

- The technology tools need to allow for a targeted, differentiated and secure sharing of data with specific individuals and stakeholders. This would suggest that in consideration of data encryption and authentication (for a certain set of use-cases), it is necessary to employ protocols and technologies that support this requirements.
- 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 privacy concerns from the individuals are clear yet complex and varied – thus it is not just about ensuring the confidentiality of the data, it is also about enhancing users' agency so that they remain in control of their data. It is as well about the users right to opt-out (as expressed through the notion of *"X9_alone: My right to be left alone"* in Table 3).

- 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.

There is current standardization work that is actively exploring and engaging with the human-centric issues highlighted in this research. One such standard is the ISO/IEC 29100:2011 "Information technology – security techniques – privacy framework". This work which is undertaken by the International Organizations for Standardization (ISO) aims at formulating a privacy framework towards shaping the processing of personally identifiable information (PII) and stipulating guidelines in line with existing and known principles for ICT privacy. This framework has application to individuals and organizations involved with the processing of PII – thus there is opportunity to embedded the human-centric considerations around data, such as has been surfaced and highlighted in this research, towards the formulation of this framework.

Within the standardization work of the International Telecommunication Union (ITU), some of the work that would be relevant for the findings in this research includes the work of the Focus Group on Data Processing and Management to support IoT and Smart Cities and Communities (FG-DPM). The focus group has as its focus the development of standards to meet the requirements of the future smart cities and communities (SC&C) for "comprehensive data processing and management frameworks and guidelines which incorporate reasonable measures to achieve layered, data-centric paradigm to support SC&C". Various working groups within the focus groups are tasked with addressing the very concerns that have been raised and explored in this research. For example, the working group 3 is exploring "Data sharing, Interoperability and Blockchain", and the working group 4 is working on "Security, Privacy and Trust including Governance". As highlighted earlier in the findings, these are the issues which can be considered from two perspectives, from the business and technology-centric perspective or from the human-centric perspectives where the formulation of the standards would seek to explicitly incorporate and infuse the values and perspectives that advance some of the core human aspirations (e.g., human dignity, privacy, freedom, non-commodification).

As the interaction between technology and human-beings becomes increasingly deeper and more inter-twinned, it is necessary to take into explicit consideration the factors that are shaping the global socio-technical trajectory. ICT standards play a crucial role in shaping the technology landscape and in influencing technological developments. The development of these standards

embeds and infuses certain values and perspectives, which have tended to be techno-centric and business focused. As suggested and highlighted in this research, it is crucial that human-centric perspectives that explore data and technology valorisation from the perspective of the individuals, are incorporated not only in the development of the technology artefacts, but also in the associated ICT standardization processes. Within the technology development domain, notions such as human-centred design, value-sensitive design, participatory design, and user-driven development already go a long way toward emphasising the need for amplifying the individuals' agency and voice in the design and development of technologies. These perspectives need to be increasingly infused in other domains, including ICT standardization, which are fundamentally shaping the future data-driven societies.

6 Conclusion

The ensuing fourth industrial revolution as well as the ubiquity and pervasiveness of data are going to increasingly permeate further 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. This paper has explored the engagement of individuals within the sustainable data ecosystem, by investigating through a non-probabilistic survey study and presenting through descriptive statistical analysis: the attitudes and perceptions of individuals 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 ICT artefacts 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 – and how these can be infused with more human-centric considerations.

The role of individuals in data-driven societies is paramount and the necessity to support individuals' active involvement and substantive 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 both the data revolution and the fourth industrial revolution accrue to all, without leaving anyone behind.

References

[1] Greenfield, A. (2006). *Everyware: The dawning age of ubiquitous computing*. New Riders.

[2] Schwab, K. (2017). *The fourth industrial revolution*. Crown Business.

[3] Mai, J. E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192–199.

[4] Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

[5] Chen, H., Chiang, R. H., and Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 36(4), 1165–1188.

[6] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 61–67.

[7] Loebbecke, C., and Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, 24(3), 149–157.

[8] Höchtl, J., Parycek, P., and Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 147–169.

[9] Gang-Hoon, K., Silvana, T., and Ji-Hyong, C. (2014). "Big-data applications in the government sector," *Commun. ACM*, 57(3), 78–85, 2014.

[10] General Assembly, United Nations, (2015). "Transforming our world: The 2030 agenda for sustainable development," *https//sustainabledevelopment.un.org/content/documents/7891Transforming%20Our%20World.pdf*, 1, 1–5.

[11] Letouze, E. (2012). Thoughts on Big Data and the SDGs. *United Nations Sustainable Development*.

[12] *A World that Counts: Mobilising the Data Revolution for Sustainable Development*, (2014). Economic Commission for Latin America and the Caribbean (ECLAC).

[13] Letouzé, E. (2012). Big Data for Development: Opportunities and Challenges (White p). New York: United Nations Global Pulse.

[14] SDSN, (2015). *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development*. Sustainable Development Solutions Network, New York, USA.

[15] Li, I., Dey, A., and Forlizzi, J. (2010). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, (pp. 557–566). ACM.

[16] Rooksby, J., Rost, M., Morrison, A., and Chalmers, M. C. (2014). Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1163–1172). ACM.

[17] Gurrin, C., Smeaton, A. F., and Doherty, A. R. (2014). Lifelogging: Personal big data. *Foundations and Trends$^{®}$ in Information Retrieval*, 8(1), 1–125.

[18] Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. (2014). Human-data interaction: The human face of the data-driven society. *SSRN Electron. J.*

[19] Estrin, D. (2014). Small data, where n = me. *Communications of the ACM*, 57(4), 32–34.

[20] Lindstrom, M. (2016). *Small data: the tiny clues that uncover huge trends*. St. Martin's Press.

[21] Best, M. (2015). Small Data and Sustainable Development. In *International Conference on Communication/Culture and Sustainable Development Goals: Challenges for a new generation*.

[22] Thinyane, M. (2017). Investigating an Architectural Framework for Small Data Platforms. In *Data for societal challenges-17th European Conference on Digital Government (ECDG 2017)* (pp. 220–227).

[23] Sen, A. (1999). *Development as freedom*. Oxford University Press.

[24] Freire, P. (2014). *Pedagogy of hope: Reliving Pedagogy of the oppressed*. Bloomsbury Academic.

[25] Stals, S., Smyth, M., and Mival, O. (2017). "Sharing and exploring Quantified-Self Data on In-Place experiences and emotions," in *ACM CHI 2017 - Quantified Data and Social Relationships*.

[26] Teles, N. (2007). Motivations and Incentives: from the "crowding-out effect" to "peer-production".

[27] Kreps, D. M. "Intrinsic Motivation and Extrinsic Incentives," *Am. Econ. Rev.*, 87, 359–364.

[28] Cobanoglu, C., and Cobanoglu, N. (2003). "The effect of incentives in web surveys: application and ethical considerations," *Int. J. Mark. Res.*, 45(4), 475–488.

[29] Dewey, J. (1986). "Experience and education," in *The Educational Forum*, 50(3), 241–252.

[30] Hallnas, L., and Redstrom, J. (2001). "Slow Technology? Designing for Reflection," *Pers. Ubiquitous Comput.*, 5(3), 201–212.

[31] Fleck, R., and Harrison, D. (2015). "Shared PI: Sharing personal data to support reflection and behaviour change," in *Workshop on "Beyond Personal Informatics: Designing for Experiences of Data"- CHI2015*.

[32] Ploderer, B., Reitberger, W., Oinas-Kukkonen, H., and van Gemert-Pijnen, J. (2014). "Social interaction and reflection for behaviour change," *Pers. Ubiquitous Comput.*, 18(7), 1667–1676.

[33] Puussaar, A., Clear, A., and Wright, P. (2017). "Better together: Reciprocal sharing and social sensemaking of data," in *Workshop on "Quantified Data and Social Relationships" - CHI2017*.

[34] Ibrahim, S. S. (2006). "From Individual to Collective Capabilities: The Capability Approach as a Conceptual Framework for Self-help," *J. Hum. Dev.*, 7(3), 397–416.

[35] Solove, D. J. (2006). "A Taxonomy of Privacy," *Univ. PA. Law Rev.*, 154(3), 477.

[36] Solove, D. J. (2002). "Conceptualizing Privacy," *Calif. Law Rev.*, 90(4), 1087.

Biography

Mamello Thinyane is a Principal Research Fellow at the United Nations University institute on Computing and Society. He leads research within the Small Data Lab investigating the use of data and information and communication technologies to empower individuals and community-level actors towards

the achievement of the Sustainable Development Goals. Mamello studied at Rhodes University in South Africa where he obtained his BSc in Information Systems, MSc in Computer Science, and a Doctorate in Computer Science. Prior to joining UNU-CS he was an Associate Professor in the Department of Computer Science at the University of Fort Hare, and a Director of the Telkom Centre of Excellence in ICT for Development.