Towards Ethical Guidelines for Fair Data Economy – Thematic Analysis of Values of Europeans

Minna M. Rantanen[0000-0001-8832-5616]

Department of Management and Entrepreneurship, University of Turku, Turku, Finland-
minna.m.rantanen@utu.fi

Abstract. Data has become a modern commodity and a resource that is produced, shared and consumed in increasing amounts. This development has led to creation of data economy ecosystems. These economies are complex socio-technical systems, where the role of individuals as consumers and producers of data is vital. Although recent events in existing data economies have raised awareness of value of the personal data, there is very little research done about individuals and their needs in these ecosystems.

In this paper individuals and their needs are researched in context of fair data economy – an on-going project that aims to develop an ethical European data economy that is fair to all of its actors. To create ethical governance guidelines, values of European individuals are researched by means of thematic analysis from 4,792 answers to two open question about fair data economy conducted in Finland, France, Germany and the Netherlands. As a result, a value basis for the governance model is formed.

Keywords: data economy, data ecosystem, human values, thematic analysis

1 Introduction

Production, sharing and consuming data has increased drastically over the recent years and data has become a commodity. This has created new environments that are called data ecosystems – complex networks of organisations and individuals that exchange and use data as their main resource. (Oliveira and Loscio, 2018). These ecosystems are platforms of data economy, which is often used as a synonym to data ecosystem.

Data economies can have many forms: big data, open data, governmental data, small data and personal data can all have their own economies that can also be connected to each other (see e.g. Thinyane, 2018). In this paper I focus on individuals as data producers and data subjects generating data. It is noteworthy that the data in these kind of data economies are not necessarily personal or identifiable but can also be anonymously generated.

Alas, in research individuals are often seen as merely data subjects, although their role has become more active in the data economies due to pervasiveness of technology

and possibilities of responsibilisation (Lammi and Pantzar, 2019; Birchall, 2015). Simultaneously, awareness about the value of personal data has increases in past decades. Currently personal data are seen as data subject's property rather than collectors' – although legally this is often not the case (Igo, 2018). Also awareness of the value of personal data and rights to control them have risen due to for example Facebook and Cambridge Analytica scandal, and General Data Protection Regulation (GDPR) (van Ooijen and Vrabec, 2019).

The aim of this research is to set ground for fair data economy. Fair data economy is an ongoing project of Suomen itsenäisyyden juhlarahasto (henceforth Sitra) that aims to develop an ethical European data economy that creates value to all actors in it (Sitra, 2019). Governance of a data economy is not simple since it requires maintaining a cooperation between multiple actors with different incentives. Viable data economy is not only about data collection, but also producing, using and sharing data, since data not used has no value (Zuiderwijk et al., 2016). Thus, governing a viable fair data ecosystem requires insight to motivations of individuals besides technical solutions and business models.

It is necessary to explore the needs of the individuals of data economy to achieve "the fair" of the fair data economies. There are participatory development methodologies that stem from this idea, but they are often meant to development of an information system in an organisation – thus insufficient for development of complex and vast ecosystems such as data economy. Values do however offer a viable alternative for needs, since they guide our actions and judgements, are reasonably permanent and basic (Schwartz, 2012).

This paper focuses on empirical research of values in relation to data economy. The research question is:

RQ1: *What values people relate to fair data economy?*

To answer this question, I analyse responses of market survey regarding fair data economy conducted TNS Kantar for Sitra. This survey was conducted in Finland, France, Germany and the Netherlands. In total the survey received 8,004 responses. In this paper I focus on the results of thematic analysis of open answers in order to find out how individuals perceive fair data economy. The questionnaire focused on what people would see as fair. In this paper results are analysed as values – what people find important - in the context of data economy. Open questions were answered in total by 2,385 (29.8%) and 2,407 (30.1%) respondents.

This study contributes to the emerging field of data ecosystems by providing insight to the needs and values of individuals and serves as basis of ethical guidelines for fair data economy governance. In order to adapt results of this study to these ethical guidelines more analyses of values, their relations and hierarchy are needed. These analyses are left for future research.

The rest of this study is structured as follows. The next section shortly introduces related research about the main topics of the data economy and the values. In the third section research process of the thematic analysis is presented and it is followed by the results section. In the section five the values found from the survey are discussed. Finally, conclusions are presented.

2 Background

2.1 Data economy

Data economy is still emerging research field. Due to the infancy there is still lack of consensus in terminology and common understanding about what should be considered as part of data economy. Thus, there is a need to define, what meant when talking about data economy. The idea of data economy is launched by data giants and various international bodies (Lammi and Pantzar, 2019.), but it is used in official reports and research.

The European Union describes data economy as follows: *"… is characterised by an ecosystem of different types of market players – such as manufacturers, researchers and infrastructure providers – collaborating to ensure that data is accessible and usable. This enables the market players to extract value from this data, by creating a variety of applications with a great potential to improve daily life (e.g. traffic management, optimisation of harvests or remote health care)."* (European commission, 2017.)

Data economy as a concept is misleading, since it is often used to describe a system of humans and technologies rather than abstract economy, thus ecosystem metaphor is often used to describe its ever-changing form and balance. Oliveira et al. (2019) describe data ecosystems as *"socio-technical complex networks in which actors interact and collaborate with each other to find, archive, publish, consume, or reuse data as well as to foster innovation, create value, and support new businesses."* (Oliveira et al., 2019, p. 1.) This definition reveals the multitude of the actions in a data economy.

Koskinen et al. (2019) present their own definition in an attempt to avoid the illusiveness of data economy through ecosystem metaphor, but add also normative aspect to their definition*: "Data economy ecosystem is a network, that is formed by different actors of ecosystem, that are using data as a main source or instance for business. Different actors and stakeholders are connected directly or indirectly within network and its value chains. Data economy ecosystem also incorporates the rules (official or unofficial), that direct action allowed in network. "* (Koskinen et al, 2019) For the sake of clarity the simplest term – data economy – is used in this paper although agreeing with all definitions presented.

Actors of data economies can have multitude of roles which can also overlap (Oliveira et al. 2019). In this paper I focus on individuals that do not have any formal obligations towards organisations in the data economy. These individuals are often described also as consumers, producers, or data subjects etc. depending on the context. Aguilera et al. (2015) refer to citizens that produce for and consume data from smart city apps "prosumers". This term describes well the overlapping roles of individuals in a data economies.

2.2 Value based development of data ecosystems

To truly consider data economies as socio-technical systems, we should consider needs of individuals using these systems while creating them in a participatory manner

(see e.g. Mumford, 2006). However, in the case of large data economies – such as European fair data economy – that is the desired user group is too large to consider traditional participatory methods. This does not however mean that the users should not to be heard since there are also other methods that can be used to find out what people find important.

What people find important are affected by our values and thus, values can offer an option for studying needs. From the perspective of personal values, values can be defined as something that a person or a group find important in their life (Rokeach, 1973; Schwartz, 2012). Each person holds numerous values with different level of importance. Values and value systems also vary between persons – something that is important to one person, might be unimportant to another. (Schwartz, 2012.) Despite this some basic values have been identified. For example, Schwartz (1992; 2012) has developed a theory of basic values and identified ten basic values: self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, benevolence and universalism.

As these values are basic, it could be assumed that values of the individuals in context of data economy reflect some of these values, but also more specific values might be seen important. Thus, there is a need to study values of individuals specifically in the context of data economy from empirical perspective.

3 Research process

The data of this study was collected in online questionnaire that was conducted by Kantar TNS in Finland, France, Germany and the Netherlands in 2018. The questionnaire was developed and ordered by Sitra. The survey included questions that aimed to clarify the use of digital services and attitudes towards collection and use of personal data. The aim was to clarify how individuals experience the potential use of their data from perspective of data protection and privacy. The survey included four sections: 1) background information, 2) rights in relation to data, and attitudes towards terms of use and privacy settings, 3) trust towards service providers and increasing trust, and 4) disclosure of information and its management and the concept of fair data service.

The number of respondents was in total 8,004. In survey there were total 23 questions of which three were open questions. Amount of answers open questions was more modest (n=4,792) than answers to quantitative questions. Rates of responses are presented in Table 1. This study focuses on analysis of two questions since the third one was answered only 35 times in total data set and often referred to previous open answer.

Analysed data was formed from the answers to the questions:

- Q21: Service providers collect a lot of data of you. In your opinion, how should this data be managed for you to feel that it is fair for you?
- Q23: If services that use personal data would have a "fair data" label, what would be the minimum requirement for it?

Table 1. Answer rates to the open questions

Country	Number of respondents	Number of open question respondents
Finland	n=2,000	Q21 n=705 (35.3 %) Q23 n=762 (38.1 %)
Germany	n=2,004	Q21 n=615 (30.7 %) Q23 n=581 (29.0 %)
The Netherlands	n=2,000	Q21 n=568 (28.4 %) Q23 n=537 (26.9 %)
France	n=2,000	Q21 n=497 (24.9 %) Q23 n=527 (26.4 %)
Total	n=8,004	Q21 n=2,385 (29.8 %) Q23 n=2,407 (30.1 %)

Answers were analysed by using thematic analysis. In analysis we followed guidelines of Braun and Clarke (2006). First, author and a supervisor familiarised themselves with the data. Next, answers were coded by each person. Codes were not predefined, but they were created based on the repetitive themes in answers. After initial coding, codings were compared and similar themes were combined. In this phase differing interpretations were discussed to avoid bias. Unclear answers were left uncoded to avoid overreaching interpretations. Nvivo Pro 12 was used as tool in the coding of the themes.

These themes included all of the coded answers. Majority of answers included 1-3 distinctive value statements. Some statements could have been. For example, notions such "no selling to third parties" could have been handled could refer to negative attitude instead of willingness to control data sharing. These selections were discussed by researchers, but still it is noteworthy that boundaries of these themes are not clear and presented themes are based on interpretations about similarities.

Additionally, it must be noted that not all answers were coded because some answers were too brief to draw any conclusions. For example, in both questions some of the respondents replied with a name of an organisation without context or explanation. Since it is impossible to know what exactly was argued with these answers they were not considered as a theme. Amount of uncoded answers was fairly low (< 3 %).

4 Results of the thematic analysis

The thematic analysis resulted in eight themes that are presented in Table 2. In total 6,017 values were coded from the data. In the following subsection Themes 1-7 are discussed in further detail. Theme 8 is not discussed further, since 61 respondents who stated that they are not able to answer did not provide any explanations. Additionally,

these answers were relatively rare and there is no apparent link between question or nation of the respondent.

Table 2. Major themes and amount of the references

	Major themes	References
1	User's control over data and data sharing	1,829
2	Transparency and being informed	1,396
3	Security	1,346
4	Trust and fairness	796
5	Compensation or benefits for users	251
6	Supervision and rules	226
7	Negative attitudes towards data collection and data economy	173
8	Unable to answer/does not know	61
	Codings in total	6,017

4.1 User's control over data and data sharing

Responses that included some notions about need to control data and data sharing formed the largest theme of values occurring 1,829 times. Most commonly people wanted to decide about the collection, use and secondary use of their data and be aware what data are collected, how and by who it is used and to who it is shared. They wanted an option to prevent secondary use and a possibility to withdraw their consent (dynamic consent). It was common that respondents wished for possibility to give informed consent. Clarity of the terms and conditions were seen as important part of the consent in fair data economy, but also as enabler of transparency (Theme 2).

Access to data and ability to correct data collected and possibility to delete some of it were seen as important factors in a fair data economy. Besides explicitly mentioning consent, many respondents stated that they should be the ones deciding whether data is shared and to who it is shared to. These kinds of answers highlight privacy as value, since privacy refers to ability to determine when, how and to what extend information is communicated to others.

Self-supervision of data and data use were seen especially important in situations where data are shared with third parties, though it seems that people have fairly negative attitudes towards secondary use, because answers such as "*no forwarding of data of any kind*", "*Not shared to everyone.*" and "*No sale to third parties*" were common. However, some did see secondary use of data possible in fair data economy, if they have a possibility to control the data use.

Only few respondents directly mentioned the ownership of data and suggested that they should have it. However, demands for possibility to control their personal data suggest that people perceive them as their property especially when data is used as merchandise. Seeing data as property is more distinct in answers emphasising compensations or benefits for people (Theme 5). Thus, control over the data and data sharing

also relates to the idea that personal data is something also individuals see as valuable asset.

Responses also give practical requirements for giving control to individuals: informed consent that can be withdrawn and ability to also control to whom the data sets are passed on. In order to control consents the individuals should also be able to supervise the status quo of the data collections and their use. This is also a requirement from those that wish to supervise the correctness of data about themselves. In return people could be willing to edit the data, which could enhance the quality of data sets.

4.2 Transparency and being informed

Transparency and being informed were mentioned in 1,396 answers that often overlap with the ability to control data and its usage. Transparency in this context refers to organisations and its activities being open to public scrutiny. It is a precondition to the ability to be aware of the actions an organisation. Transparency was often presented without further explanations, which implicates that people have strong intrinsic conception of it. It was often used as a synonym with openness or being open.

Small difference to Theme 1 is that people wanting to be informed did not implicate that they need some power over data, but just wanted to get information about the data usage. Thus, this theme could be described also as being aware of what data are collected and how they are used within and outside of an organisation. As being aware of the data and data usage are crucial in terms of control over data, informing can be seen as enabler of Theme 1, such as making informed consents. Also, transparency was often seen as a mean to gain trust (Theme 4).

4.3 Security

Values reflecting security were presented in the responses 1,346 times. Although data security should be sufficient in any technological solution people felt the need to highlight it. Respondents noted that fair data economy and fair data label should have high standards of security and confidentiality. They often linked to trust and fairness in parallel with security, which links this theme to theme 4. Some also connected security to informing, thus linking this theme also to previous theme.

Besides generally pointing out that security is a priority, people also saw that anonymisation of sensitive data must be guaranteed. However, in some cases people's perceptions about what data should be anonymized or what constitutes as sensitive information were varying. Some respondents saw data security as something that should be guaranteed in all situations and all times – which unfortunately is not possible. Data misuse was often mentioned, but perceptions of it varied from criminal activities to profiting from data.

Respondents also emphasized the nature of data collected, especially that only necessary data should be collected, and that data should be deleted after use. Some also wanted fair data economy to work only on restricted areas such as their home country or Europe. In these cases, security could have been seen as a safeguard of privacy. There were some contradictions between the ways that people felt that data should be handled

in a fair data economy. Some noted that preventions of human errors or curiosity are risks to security and wished that data would only be handled by technical solutions such as artificial intelligence whereas some disagreed strongly.

4.4 Trust and fairness

Trust and fairness related values occurred 796 times in the responses. Trust is generally understood as strong belief in the reliability, truth or ability of someone or somebody. Thus, trust is a subjective feeling that someone or something is trustworthy. As stated earlier, respondents saw transparency and security somewhat enablers and facilitators of trust. Thus, this theme is interconnected with previous themes. Mentions of distrust, were not uncommon and thus formed their own theme (Theme 7).

Trust and fairness were often mentioned when describing benevolence of fair data economy. Trustworthiness was generally seen as keeping one's promises, which in the context was often seen as following and not changing the terms and conditions agreed upon. Also reliability, fair and honest were seen as attributes of trustworthy actor in a data economy.

Ethicality or morals were used multiple times -mainly as adjectives - but rarely explained. Respondents who elaborated their often meant using data for good causes and in respect of their moral values. Also appropriate data management, respectful behaviour, integrity, responsibility, equality and doing no harm were mentioned multiple times, but rarely explained.

Fair treatment of employees was a surprisingly big subtheme. Aspects such fair wages, good working conditions and not using child labour were mentioned multiple times. This could result from respondents' confusion between fair data label and fair-trade label, but fair treatment of also employees of organisations in the fair data economy should be treated fairly.

4.5 Compensation or benefits for users

Respondents demanded both compensations and benefits from their data 251 times. From the answers it became apparent that current benefits for personalised advertisement were not enough and people were aware of the monetary value of their personal data. Some respondents demanded that they should have a monetary compensation for their data especially in case of secondary use. Other however demanded for more free premiums or other benefits such as discounts or better personalisation in general. On contrast to hedonistic benefits, some respondents stated that data economy is fair if it clearly supports common good and their data are used for example for non-profit research purposes.

4.6 Supervision and rules

Supervision and rules in the fair data economy were mentioned 226 times. Compliance to rules that are implicit and fair in general were seen as a safeguard of fair data economy alongside of supervision. Appropriate supervisor of actors in data ecosystem was

often described as some institute that does not profit from the ecosystem. Many mentioned that supervision could be a task for (their) government. Sanctions for violating the rules were sometimes emphasised as a measurement to maintain the integrity of fair data ecosystems and fair data label system.

4.7 Negative attitudes towards data collection and data economy

Negative attitudes towards any data collection and data economies formed a clear theme with 173 mentions in responses. Mainly respondents expressed cynical attitude towards fairness in data economy or wanted to seize any data collection. Fair data label was sometimes seen as an ineffective way to support fairness in data economy. This theme emphasises that the current ways the data economies work have provoked mistrust in the respondents.

5 Discussion

The answer the research question *"What values people relate to fair data economy?"* is that European people value express rather uniform values when considering fair data economy. The largest three: user's control over data and data sharing, transparency and being informed and security are clearly valued by majority of respondents. Trust and fairness ia also rather large theme, whereas, compensation and benefits and supervision and rules are less often mentioned. Negative attitudes towards data collection are good to acknowledge, since they express the unwillingness to of some individuals to take part in a data economy.

However, if we analyse these further, we can find more fundamental basic values behind them. Theme of control over one's data highlights basic values of power and self-direction, more specifically values of privacy and autonomy – an ability to make your own decisions without being controlled by anyone else.

Transparency and being informed naturally connect to the first theme through information needed to make an informed consent, but also link to trustworthiness. Thus, transparency and being informed can also be seen as an enabler of other values. From perspective of basic values transparency and being informed connect to self-direction, universalism, conformity, security and power. Thus, they have a wide value basis. Links between different themes and value categories emphasise the complex nature and effects of transparency and being informed. Links between different themes and values emphasise the importance of transparency and informing individuals as a corner stone of fair data economy.

Security in itself is a value and it seems that it is highly valued also in the context of fair data economy. Answers of the respondents do also have aspects that reflect value of power, since security is seen as a safeguard of privacy. Additionally, security enables trust and fairness that are discussed in the next subsection. Thus, although security is valued in itself, it does also serve as an enabler of other values.

As a theme trust and fairness compressed desirable attributes of a fair data ecosystem or appropriate behaviour within it. From perspective of values this theme and its sub-themes could be represent values of universalism, benevolence and conformity, which can be described as moral values, since they describe "oughtness" of fair data economy and how people should be treated within it. Also, justice is a cross-cutting value in this theme, since respondents do highlight just behaviour or treatment.

Theme of compensation and benefits reflected two different kind of values. Some answers highlighted basic values of hedonism, power and self-direction, but some highlighted moral values such benevolence and universalism. Compensation or benefit to oneself is clearly an issue of what is seen just – it is not okay to collect and use data without also users profiting from it in some manner. It is notable, that this theme also connects to transparency. Organisations should communicate their intentions with data clearly and point out what are the benefits of sharing data for individuals and for society.

Supervision and rules -theme or demands for rules, compliance and sanctions reflect basic values of conformity and security, although there is a link benevolence through responsibility and power trough authority. Value justice can be seen even stronger value in these themes than in the previous themes. It is notable, how this theme overlaps with transparency and trust but also the rules, their supervision by a trusted party and noted misbehaviours should be understandable and available to the public.

These negative attitudes make visible some of the challenges of the fair data ecosystem. These respondents do not trust organisations collecting data nor that they could be benefit from mechanisms such fair data label. These kinds of attitudes could reflect basic values of security, tradition, self-direction and power. However, due to lack of explanations it is impossible to deduct reasons behind these attitudes. However, it should be noted that strong negative attitudes exist and challenge the idea of fair data economy.

6 Conclusions

To conclude, it seems that people find important that practices in a fair data economy should support their autonomy, protect their privacy, be secure, transparent, trustworthy, benevolent and just. In addition, it should also be benevolent and benefit individuals as well as the society. In designing the ethical governance guidelines for fair data ecosystems these issues should be taken into consideration and implemented in them.

This research contributes to the emerging research of data economies and data economy ethics. It explores the values of Europeans and deepens the knowledge about individuals' conception and desires in the context fair data economy. These results offer a basis for future research of user-centric data economies and development of ethical governance guidelines.

Naturally this research has its limitations. First, it is limited to only four countries and rather small sample of all European countries. Thus, it cannot be considered as a presentative research of all European countries though it seems that there is little vari-

ance between values. Second, the survey that was used to collect the data was not specifically designed for value research, but to study citizen's attitudes that are affected by their values. Thus, there is much possibilities for future work in this field. Nevertheless, this research offers a good starting point for this type of research.

Acknowledgements

The author is partly working in the Sitra's IHAN Ethics -project and would like to thank Sitra for sharing the questionnaire data.

Additionally, thanks belong to PhD Jani Koskinen, who helped the author with thematic analysis part of the study.

References

1. Aguilera, U., Peña, O., Belmonte, O., López-de-Ipiña, D. (2015). Citizen-centric data services for smarter cities. Future Generation Computer Systems 76, 234–247.
2. Birchall, C. (2015). 'Data.gov-in-a-box': Delimiting transparency. European Journal of Social Theory, 18(2), 185–202.
3. Braun, V., Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.
4. European commission (2017) COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS "BUILDING A EUROPEAN DATA ECONOMY". https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2017%3A9%3AFIN
5. Igo, S.E. (2018). Me and My data. Historical Studies in the Natural Sciences, 48(5), 616–626.
6. Koskinen, J., Knaapi-Junnila, S. and Rantanen M. M. (2019) What if we had fair – people-centred – data economy ecosystems? 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation**.**
7. Lammi M., Pantzar M. (2019). The data economy: How technological change has altered the role of the citizen-consumer. Technology in Society, 59.
8. Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. Information Systems Journal, 16(4), 317–342.
9. Oliveira and Lóscio, B. (2018). What is a data ecosystem?. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (dg.o '18), Anneke Zuiderwijk and Charles C. Hinnant (Eds.). ACM, New York, NY, USA, Article 74,
10. Oliveira, M.I S., Barros Lima, G..F. & Farias Lóscio, B (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowl Inf Syst, 01 January 2019, 1–42.
11. Rokeach, M. (1973). The nature of human values. New York, NY, US: Free Press.
12. Schwartz, S. H. (1992). Basics in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), Advances in experimental social psychology, 25, pp. 1-65.
13. Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture, 2(1).

14. Sitra (2019) Fair data economy. https://www.sitra.fi/en/topics/fair-data-economy/, accessed 10.8.2019.
15. Thinyane, M. (2018). Small data and sustainable development — Individuals at the center of data-driven societies," 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K), Nanjing, 2017, pp. 1-8.
16. van Ooijen, I. and Vrabec, U. (2019). Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. Journal of Consumer Policy, 42(1), 91-107.
17. Zuiderwijk A., Janssen M., Van De Kaa G., Poulis K. (2016) The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. Information polity, 21(3), 223-236.