
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Nikander, Pekka; Elo, Tommi

Will the data markets necessarily fail?

Published: 01/01/2019

Document Version

Publisher's PDF, also known as Version of record

Please cite the original version:

Nikander, P., & Elo, T. (2019). Will the data markets necessarily fail?

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Will the data markets necessarily fail?

A position paper

Pekka Nikander, Tommi Elo

Department of Communications and Networking Aalto University School of Electrical Engineering Espoo,
Finland

pekka.nikander@aalto.fi, tommi.elo@aalto.fi

Abstract—With the billions of Internet of Things devices connected via the 5G and other networks, loads of useful data are produced. However, the majority of these data are disappearing into the silos of cloud and IoT companies. This problem is exacerbated by the current economic system creating perverse incentives that push companies to keep their data private and not to sell or share them. From the society point of view, this leads to severe inefficiencies. More structurally, Adam Smith's invisible hand does not work: in the data markets, *the public and private interests are not aligned* by the current market forces.

Based on these observations, we present a conjecture wherein we state that any attempts to fix the market failure in the data markets within the current economic structures are bound to be inefficient. Only by *redefining fundamental economic concepts*, such as ownership and money, we can efficiently align the interests, clear the markets, and gain welfare potential. Furthermore, we briefly suggest an urban community currency experiment wherein this conjecture could be empirically tested.

Keywords — 5G, data markets, market failure, non-rival, anti- rival, creative destruction, perverse incentives

I. INTRODUCTION

The 5th generation cellular networks are being launched now. It can be safely assumed that in the 5G the role of machine-to-machine communication will be even larger than in the 4G networks. This will result in *very large amounts of data*, including “raw” data from sensors, cameras, etc; processed and annotated data resulting from such data; and metadata describing the datasets. Furthermore, the datasets will be combined with much richer metadata and meta-meta-data structures than what is available today.

At the same time, it looks extremely likely that the large majority of this data will be considered to be “owned” by someone¹, and *kept private* [6,7,12]. There are several reasons for this. One reason is that it is extremely hard to evaluate the monetary value of any given data set, since the value depends on the context in which it is used [7]. Hence, companies are reluctant to sell their data,

¹ Please note that we write “own” in quotes. The reason for this is that there are many reasons why data should not be owned, and in certain cases, even cannot be owned. See e.g. Carballa 2016 [4] or Davidson et al 2016 [5].

since they cannot easily evaluate the monetary value of the data to the purchaser and consequently are afraid of setting a price. Another, perhaps more obvious reason is that many companies are afraid of their data being used against them [6]. That is, companies may think that if they give or sell their data, that data may help their competitors or other companies so much that they gain a competitive advantage against the company where the data has originated from, eventually reducing its profits or even throwing it out from the markets. There are also other reasons related to, for example, the difficulties in controlling where the data may end up, issues with legislation that make it practically impossible to sue third parties gaining access to the data, difficulties in determining which parts of the data contain sensitive information, etc. As a result of such factors, the majority of the industrial data is very likely to remain private to the companies, just as today.

From the efficiency point of view, this privatisation of industrial and other data is very problematic. In general, the more relevant data is available about a production process, the better the process can be optimised. For example, according to Müller et al, typical (private) big data and analytics assets are associated with 3–7 percent improvement in production [9]. The compound effect of sharing the data sets within industries may be estimated to be of at least the same magnitude, with further expected efficiency gains if datasets can be shared also between industries. While there appears to be no reliable works on this, we surmise that it may be estimated that on the global level shared big data and analytics may lead up to at least 15 – 20% efficiency gains across many industries.

From a more general point of view, the lost potential gains are even larger. For example, combining mobility and health data could allow us to estimate and perhaps even prevent the spread of many infectious diseases much more efficiently than today, leading to both people feeling better and allowing them to be more productive [13].

For the 5G network, we surmise that the current general unwillingness to share data will lead to severe inefficiencies in both running the network itself and especially to failures in attempts to create efficient data markets to the industrial and other data available through the network.

The rest of this paper are organised as follows. First, in Section II we describe the underlying theoretical underpinnings on the nature of data. In Section III we introduce our structural market failure hypothesis. In Section IV we discuss the implications and some about potential future work.

II. RELATED WORK

A reason for the general unwillingness to share data seems to be embedded into the very structure of our economy.

In economics, Arrow discussed widely the nature of immaterial goods in production and pointed out to the difference between *rival* and *non-rival* good [1]. Most of the production factors are rival, basically meaning that if I use such a good, such as drink a cup of coffee, you cannot use the very same good, i.e. the same cup of coffee. However, non-rival goods are structurally different. For example, if I visit a museum, my visiting the museum does not hamper you ability to visit the same museum.² Hence, *experience goods* — or “club goods” in the economic parlance — are non-rival.

² Of course, such a distinction is only an abstract approximation. If enough of people visit a museum, you will end up queuing.

While discussing value for open source software, Weber introduced the term *anti-rival* to economics [11]. Essentially, an anti-rival good is one that *gains value the more it is used*. Today a classic example of an anti-rival good is a massive multiplayer online game, which becomes more enjoyable to its users the more other people play the same game.³ Bauwens extended this analysis of peer production [2]. Coming closer to our argumentation, Kemkes et al stated that “Payments are an inefficient policy approach when an ecosystem service is anti- rival,” basically indicating that using money to compensate for anti-rival production will lead to economic inefficiencies, at least in the context of ecosystem services [8]. Bauwens and Kostakis went even further and claimed that “the primary driver [for economic activity in the future] is the sphere of abundance of knowledge available for all, which is not a market driven by supply and demand dynamics.” [3]

Finally, in his working paper Carballa argues eloquently about the structure of (personal) data and why, exactly, the current market structures lead to the near-monopolistic market structures in the platform economy, where the winner-takes-all is more a rule than an exception [4].

The resulting extension to the economics standard types of goods matrix has been nicely illustrated at least by “Review Yobo” in 2015; see Table 1.



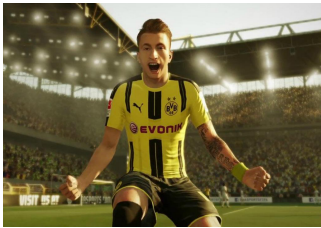



	Rival	Non-Rival	Anti-rival
Excludable	 Private good (coffee)	 Club good (museum)	 Network good (game)
Non-excludable	 Common good (ocean fish)	 Public good (beach)	 Symbiotic good (knowledge)

Table 1. Types of economic goods

Adapted from: Review Yobo: Economics of video games - the knowns and unknowns! [24]

³ One should make a distinction between the game itself, which is anti-rival, and the underlying technological products, which are rival goods.

III. INTERLOCKING STRUCTURAL INEFFICIENCIES HYPOTHESIS ?

Given this, we surmise that the reasons for the winner-take-all phenomenon and the general industrial unwillingness for sharing data are *structural* in nature. We take the argument of Kemkes and surmising that payments, i.e. *money*, are an *inefficient compensation method* for data. Furthermore, reflecting Bauwens, Carballa, and e.g. Davidson et al [5], we further surmise that *ownership is an inefficient governance model* for data.

Our first hypothesis is, it is *impossible to efficiently* use money to compensate for sharing data. That is, money simply does not work as an appropriate “payment” method for parties that would be willing to share their data as money is structurally rival, and cannot be copied. Indeed, if it can be copied, it ceases to be a good form of money. Anti-rival goods, such as data, on the other hand, are by default infinitely shareable. In practical terms, data can be copied at an essentially zero cost. Thus other methods of compensation are needed. We further suggest that such other methods should themselves be anti-rival in nature. In other words, these new compensation methods should be ones that “gain value the more they are used.” In practical terms, that means that they should be more like *reputation* or even *honour* than money. Reputation is earned, a little bit like money, but it cannot be disbursed, unlike money or tokens, which may be given out so that you no longer hold them. Not so with reputation. You cannot give your reputation to someone else, but you can use your reputation to affect theirs, which often leads to reciprocity and both of you gaining more reputation, in a quite anti-rivalry manner. However, it should be noted that at this writing this hypothesis is still quite controversial and remains theoretically and empirically unproven.

What comes to our second interlocking hypothesis about ownership and governance, there appears to be a wider consensus that the situation is very unclear. Some authors argue that data should not or even cannot be owned (e.g. Carballa[4]), while others recommend full enclosure or privatisation of data (e.g. Jones and Tonetti [6]). It appears that in many cases these differences in opinions are politically motivated and without much empirical or even theoretical evidence. As an example of the related real world problems, consider the case of American leased tractors, where the farmers in many cases cannot maintain or repair their machines themselves due to insufficient access to data, leading to inefficiencies [10].

IV. SUGGESTED NEW MARKET STRUCTURES

While it may be too early to seriously envision what the data economics replacements for money and ownership would be, in this section we outline our current, very preliminary understanding. We first discuss a couple of alternative governance forms, data commons and data unions, already suggested in the literature, and then proceed to vision a potential anti-rival compensation mechanism. Finally, we attempt to bring these two together.

The term “data commons,” mostly in the context of scientific data, entered the discussion by several authors mostly in the early 2000s, partially stemming from David’s [14] Stanford discussion paper on public knowledge “commons.” Miller et al [15] , in parallel with Murray-Rust [16], discussed how the Creative Commons thinking had been applied to open data, and Yakowitz [17] compared the growing privacy concerns w.r.t. research data to Hardin’s 1968 tragedy of commons.

Extending the discussion of data commons beyond open and scientific data has merely started. Bloom suggested community data commons [18]. Eschenfelder and Johnson use the commons theory to study the repository rules for 24 open science web repositories, thereby shifting the attention from scientific data to the commons mechanism [19]. Contreras and Reichman went further and suggested a number of structural models for scientific data pools, with differing governance models [20]. To our knowledge, Carballa was the first to propose a “data commons” as a generic concept for sharing economy [4]. Ruppert et al defined the term “data politics” and mentioned passingly data commons as an alternative to ownership [21] .

In our opinion, data commons, largely as defined by Carballa, looks like a very promising governance model for many industrial cases. However, extensive care must be taken to design the rules on how all members in the commons will jointly benefit from any use of the data in the commons, so that instead of the tragedy we will have the comedy of commons [22]. We explore this in a little bit more practical terms in the next section.

Another alternative governance form the data is a “data union,” as experimented by The Data Union [23]. The basic idea is for the producers of private data, i.e. all citizens, to form a union similar to the trade unions, and collectively bargain with anyone who would like to use their personal data. Unfortunately this concept is still very underdeveloped both in public discussion and the research community.

Turning our eye towards compensation, we return to the introduced term “anti-rival compensation”, denoting compensation methods that gain more value as the underlying asset is being used or shared and thereby gaining more value. In today’s terms, such methods would be contractual arrangements, typically entered by the members of a data commons or a data union, basically agreeing to infinitely defer any direct monetisation of the value accumulating at the commons or unions. Of course, such deferment does not preclude indirect monetisation. In that respect such an arrangement may be compared to a fund or trust which by charter is not allowed to touch the principal but only to deal with its proceeds.

It should be noted that such anti-rival compensation methods are just starting to appear, thanks to the blockchain phenomenon. They also exist, more-or-less by accident, in various open source and open data projects, where the value is largely created as source code data or scientific data, captured through the non-monetised use of this data, and indirectly monetised through contract jobs. In such systems the reputation the members of the commons have accumulated, affects their ability to charge for contracted work.

Compensation methods where non-rival properties are being designed in to digital business platforms with cryptotokens, was studied in the context of ecosystem level value propositions and their implementation by Hakanen et al [26]. Using alternative monies to align ecosystem level and individual level incentives in the form of cryptotokens seems a promising new technology to structure the markets.

V. MODELLING AND EMPIRICAL TESTING

We aim to test our hypothesis by creating models and simulations of alternative currencies and their credit scoring, decentralised digital business platforms with unorthodox ownership structures, and

the necessarily resulting currency value interactions with fiat. Simulations can be used to compare the efficiency of different ownership structures, which we surmise to be interlocking with the nature of the compensation methods (currencies or monies) used.

Based on our preliminary work, community currencies seem like an excellent platform for empirically testing at least some parts of our hypothesis. While most of the community currencies have focused on the needs of the community, such as providing liquidity to where there is too little, to decrease inequality, or to promote local businesses, some have experimented with novel economic structures, such as peer-to-peer monies.

Without going into details, we surmise that an urban community currency that has a strong public or common good element might be an ideal platform for testing our ideas. One possibility might be the “Digital Citizens”, a winning concept[27] developed at the ICT 2018 FIWARE hackathon for the City of Vienna. The concept was partially inspired by Pazaitis et al [25].

In this conceptual “Digital Citizens” community currency, Viennese could perform small tasks that would be good for the city or e.g. for their own health (thereby indirectly benefitting the city) and gain simultaneously city tokens and reputation. The tokens could be exchanged for services, such as paying for public transport or getting discounts on local businesses. The reputation, on the other hand, would grant more non-material benefits such as free use of city bikes or more voting power in deciding how funds are distributed in collaborative budgeting projects or how the rules of the community currency may be changed.

In this context, the citizen currencies and their associated credit scoring form interacting platforms, kind-of commons, with both a traditional monetary-like medium of exchange, the tokens, and a builtin anti-rival compensation mechanisms (e.g. the reputation).

VI. DISCUSSION AND FUTURE WORK

In this position paper, we have presented a hypothesis where we claimed that the current compensation and ownership structures are inefficient to clear the data markets. We have further surmised that the underlying structural mismatch of rival vs. anti-rival goods contributes directly to the winner-takes-all phenomenon, and to the failure to establish markets, in the private and industrial data markets, respectively.

Given that we have to doubt two interlocking cornerstones of our current economic system, it will necessarily take some time just to establish sufficient discussion to understand the scope of the problem. It may also turn out that it suffices to innovate in only either compensation or governance, in order to create efficient solutions.

Basing on these conjectures, in this paper we have proposed that we should consider data commons and data unions as the means of governance for data, instead of privatised data ownership.

In addition, we briefly elaborated the novel concept of anti-rival compensation, noting how digital reputation appears to be the best current approximation for this still largely non-existing conception.

Finally, we proposed that urban digital currencies, such as the “Digital Citizens” community currency suggested to the city of Vienna, might be an ideal platform for performing empirical

studies on these novel concepts. It is important that such currencies and their associated platforms take into account non-rival and anti-rival properties when the incentive structures of the currency are designed.

We believe that unless the structural inefficiencies of governance and compensation of data markets are seriously assessed and eventually solved, the industries building upon Internet of Things and 5G will work in a seriously suboptimal manner. And they will fail to utilise the full potential of the huge amounts of data that will be collected and stored. Consequently, we are calling for open collaboration and cooperation in this area.

ACKNOWLEDGEMENTS

We are thankful to our colleagues Timo Seppälä, Juri Mattila, Aija Leiponen, and Ville Eloranta for the fruitful discussions on this topic.

REFERENCES

- [1] K. J. Arrow, “The Economic Implications of Learning by Doing,” in *Readings in the Theory of Growth: a selection of papers from the Review of Economic Studies*, F. H. Hahn, Ed. London: Palgrave Macmillan UK, 1971, pp. 131–149.
- [2] M. Bauwens, “Class and capital in peer production,” *Capital & Class*, vol. 33, no. 1, pp. 121–141, Mar. 2009.
- [3] M. Bauwens and V. Kostakis, “Towards a new reconfiguration among the state, civil society and the market,” *Journal of Peer Production*, vol. 7, pp. 1–6, 2015.
- [4] B. Carballa Smichowski, “Data as a common in the sharing economy: a general policy proposal,” HAL, CEPN Working Paper, Oct. 2016.
- [5] S. Davidson, P. De Filippi, and J. Potts, “Economics of Blockchain,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2744751, Mar. 2016.

- [6] C. Jones, C. Tonetti, and others, "Nonrivalry and the Economics of Data," in *2018 Meeting Papers*, 2018, vol. 477.
- [7] P. Koutroumpis, A. Leiponen, and L. Thomas, "Data Strategy," *Global Proc*, vol. Surrey, no. 2018, p. 85, Apr. 2018.
- [8] R. J. Kemkes, J. Farley, and C. J. Koliba, "Determining when payments are an effective policy approach to ecosystem service provision," *Ecological Economics*, vol. 69, no. 11, pp. 2069–2074, Sep. 2010.
- [9] O. Müller, M. Fay, and J. vom Brocke, "The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics," *Journal of Management Information Systems*, vol. 35, no. 2, pp. 488–509, Apr. 2018.
- [10] A. Shah, "Can You Repair What You Own? | Mechanical Engineering Magazine Select Articles | ASME DC." [Online]. Available: <http://memagazineselect.asmedigitalcollection.asme.org/article.aspx?articleid=2705094>. [Accessed: 03-Jun-2019].
- [11] S. Weber, "The Political Economy of Open Source Software," p. 42, 2000.
- [12] P. P. Tallon, "Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost," *Computer*, vol. 46, no. 6, pp. 32–38, Jun. 2013.
- [13] J. Roski, G. W. Bo-Linn, and T. A. Andrews, "Creating Value In Health Care Through Big Data: Opportunities And Policy Implications," *Health Affairs*, vol. 33, no. 7, pp. 1115–1122, Jul. 2014.
- [14] P. A. David, "A tragedy of the public knowledge 'commons,'" Global science, intellectual property and the digital technology boomerang. 2001b. MERIT, Maastricht, p. 42, 2000.
- [15] P. Miller, R. Styles, and T. Heath, "Open Data Commons, A License for Open Data," LDOW, 369, 2008.
- [16] P. Murray-Rust, "Open Data in Science," *Serials Review*, vol. 34, no. 1, pp. 52–64, Mar. 2008.
- [17] J. Yakowitz, "Tragedy of the Data Commons," *Harv. J. L. & Tech.*, vol. 25, p. 1, 2012 2011.
- [18] B. Goldstein and L. Dyson, Eds., *Beyond transparency: open data and the future of civic innovation*. San Francisco: Code for America Press, 2013.
- [19] K. R. Eschenfelder and A. Johnson, "Managing the data commons: Controlled sharing of scholarly data," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1757–1774, 2014.
- [20] J. L. Contreras and J. H. Reichman, "Sharing by design: Data and decentralized commons," *Science*, vol. 350, no. 6266, pp. 1312–1314, Dec. 2015.
- [21] E. Ruppert, E. Isin, and D. Bigo, "Data politics," *Big Data & Society*, vol. 4, no. 2, p. 2053951717717749, Dec. 2017.

- [22] C. Rose, “The Comedy of the Commons: Custom, Commerce, and Inherently Public Property,” *U. Chi. L. Rev.*, vol. 53, p. 711, 1986.
- [23] “Data Union”, <https://thedataunion.eu> [Accessed March 19, 2019.]
- [24] Review Yobo (pseudonym), “Economics of video games - the knowns and unknowns!,” GameSkinny. Published Oct 7th 2015.
<https://www.gameskinny.com/oirze/economics-of-video-games-the-knowns-and-unknowns>
[Accessed 19 March 2019.]
- [25] A. Pazaitis, P. De Filippi, and V. Kostakis, “Blockchain and value systems in the sharing economy: The illustrative case of Backfeed,” *Technological Forecasting and Social Change*, vol. 125, pp. 105–115, Dec. 2017.
- [26] Hakanen, E., Töytäri, P., Turunen, T., & Eloranta, V., (forthcoming). Strategic Logics Behind Blockchain Projects: Capturing Value in Decentralized Ecosystems. *Academy of Management Global Proceedings*, (2018), 409, 2018.
- [27] P. Nikander and E. G. Servest, “A concept developed at the ICT 2018 FIWARE hackathon.” [Online]. Available: <https://joinup.ec.europa.eu/collection/open-source-observatory-osor/news/city-district-live-data>. [Accessed: 03-Jun-2019].