

Designing governance models
to meet the needs of
data capitalism

- Interview with Thomas Hardjono
by Shigeki Hayashi -

5 June 2023

**Special
Edition**

lakyara vol.371

Executive Summary

With data protection becoming an ever more crucial imperative for companies and governments, we spoke with Thomas Hardjono, CTO of MIT's Connection Science Group, about how to strike the right balance between data access and protection and why the co-op model of data governance holds promise.

.....



Thomas Hardjono

*CTO of Connection Science,
Massachusetts Institute of Technology*

Dr Thomas Hardjono is the CTO of Connection Science and Technical Director of the MIT Trust-Data Consortium at MIT in Cambridge, MA. USA. He is an early pioneer in the field of digital identities and trusted hardware, and instrumental in the development and broad adoption of the MIT Kerberos authentication protocol. His activities include leading standard development efforts, notably at the IETF (Internet Engineering Task Force), IEEE, Trusted Computing Group, Confidential Computing Alliance and others. He is currently involved in several startups around the MIT community. His current area of interest is Web3 Digital Assets, with focus on the interoperability of asset networks and survivability of these networks against cybersecurity attacks.

Shigeki Hayashi

*Adviser,
Nomura Research Institute*

Joined NRI in 1988. Initially assigned to PMS Development Department; later transferred to Insurance Systems Department. Promoted to General Manager of Project Development Department, Financial Solution Sector and then General Manager of Innovative Systems for Financial Industry Department. Seconded to Nomura Holdings in 2007. Returned to NRI in 2009. Appointed Executive Officer and Deputy Division Manager–Insurance Solution in 2012. Promoted to Division Manager–Insurance Solution Division in 2014. Appointed Senior Corporate Managing Director in 2016. Division Manager–Financial Technology Solution since 2017. Appointed Senior Executive Managing Director in 2021. Current position in April 2023.



Striking the right balance between data utilization and data protection

Shigeki Hayashi: I understand that MIT conducts practical research to solve societal problems in collaboration with governments and companies globally. What kind of research does your Connection Science Group do?



Thomas Hardjono: The group's goal is to utilize diverse data sources in what we call computational social science, a discipline that brings data and social behavior together. Current economic models from the 1950-60s see human beings only as logical units that transact with each other. Our research recognizes that people actually do more than that. They also exchange ideas, news, etc.

Hayashi: What kind of projects are you working on?

Hardjono: One example is a project sponsored by a credit card company that came to us to improve its ability to predict which of its cardholders will be experiencing financial stress within six months. We recommended combining data from different verticals. If you combine card payments data with mobility data from mobile network providers, you can better predict credit card churn and, in turn, financial stress [Reference: Money Walks: Implicit Mobility Behavior and Financial Well-Being, by V. K. Sing, B. Bozkaya and Alex Pentland, PLOS ONE Journal].

Hayashi: One question on my mind in recent years is how to strike the right balance between the public interest and protection of individuals' rights with respect to data. For example, public hygiene sometimes took precedence over privacy during the pandemic. Also, many countries are now working on enacting data regulations. What are your views on the legal frameworks being developed by countries globally?

Hardjono: Governments, corporations and communities all need to use data. Data has become an important third form of capital alongside money and human capital.

In terms of legal frameworks, most countries have the same requirements. Namely, they have to figure how to protect data while ensuring it can be utilized as needed by the government, companies and other parties. Data protection encompasses

protection of not only of individuals' privacy but also national sovereignty. There's now so much hacking going on, much of it perpetrated by state actors. A few years ago, hackers stole a huge dataset on US federal employees [Reference: Washington Post, "Hacks of OPM database compromised 22.1 million people, federal authorities say", Washinton Post, July 9, 2015].

Hayashi: What should we be focusing on in terms of technology?

Hardjono: The challenge is developing technology that allows data to remain usable but in a protected state. We have been looking at many different elemental technologies such as multiparty computation and homomorphic encryption, which enables computations to be performed on encrypted data without decrypting it. Our goal is to develop ways to safely share insights. For example, a company may want to do computations using data in the possession of another company. In such cases, instead of moving the data, it is better to allow access through question-and-answer mode: you send a query and receive a response. If such an approach had been used for COVID contact tracing, the process likely would have gone more smoothly. This Q-and-A approach is applicable not only within countries but also to international collaborations.

Data as Capital

Hayashi: Circling back to your earlier comment that data has now become a key form of capital, I believe that just as countries have become affluent by virtue of capitalists amassing monetary capital to fund infrastructure projects and other enterprises, Big Data likewise has the potential to be a potent driver of wealth creation.



At the same time, I fear that Big Data accumulation could have adverse consequences à la the pursuit of monetary wealth under capitalism. For example, we have seen that information manipulation can influence how people vote. Other concerns include lapses in data ethics and cybercrime risks. What must be done to address such concerns?



Hardjono: The first thing is to ensure data quality. Data's source or provenance is particularly important. For any kind of dataset, you want to know who generated it. Ideally, data should be kept by whoever created it because when data gets copied over and over again, you lose track of where it came from. Doubts about data quality crop up in such cases.

Secondly, you really need to provide a governance framework around access to data. That means controlling who has access to the data for what purpose and for how long. You need transparency to gain public confidence.

A third point is algorithm quality. Our research has found that bias not only exists in data but can also be inadvertently embedded in algorithms. The goal is for both data and algorithms to be fair and unbiased.

Hayashi: Governments have played a major role in capitalism. What is their role in data capitalism?

Hardjono: I'll give you an example. We are involved in a data governance project in South Australia (SA). The SA state government is building a secure health data exchange. Its hospitals and other healthcare facilities have a huge amount of citizen health data. Many pharmaceutical and biotech companies are interested in this data. Some want to create a DNA biobank; others want to recruit people for clinical trials. It's not an easy project because it entails a lot of buy-in and support from the public. I think the government's role is to act as a mediator to bring everyone together and to ensure that citizen data privacy policies are followed.

Hayashi: Going back to what you said about algorithms, how is their fairness determined? In the US, where there is currently a big gap in worldviews between the left and right wings of the political spectrum, algorithmic manipulation of people's thinking or behavior is an issue. In such an environment, it seems to me that even an algorithm everyone agrees is fair may not necessarily be fair.

Hardjono: What it means for an algorithm to be fair is a very difficult topic. How to define a fair algorithm on a technical level is the subject of ongoing research.

Our current research has found that there is often intentional skew in social media data such as Twitter tweets. For example, we currently have a PhD student studying whether a given twitter feed contains any coordinated propaganda. We are finding that some countries are, for political reasons, deploying bots on Twitter to change the information people see in their Twitter feeds [Reference: “Detection of Coordination Between State-Linked Actors”, Keeley Erhardt & Alex Pentland, SBP-BRiMS September 2022].

We are trying to figure out how to detect and circumvent such bots. To do so, we believe people who use social media need to be authenticated. Right now, many Twitter users are seeing information generated by bots.

On the other hand, we also have research that shows that the left-right political schism in the US has been exacerbated by social media algorithms. We are now looking at how to create a more decentralized social media platform controlled by users, not by a single company like Twitter or Facebook.

Prospects of new data governance

Hayashi: Listening to you, I feel governance issues pose a bigger challenge to data capitalism than technological issues do. Who should govern data capitalism and what type of governance model would work best?



Hardjono: I think the individual citizen needs to participate in governance as a stakeholder. A top-down approach where the government imposes policies or regulations without buy-in from citizens will not work.

One governance model we are investigating is data cooperatives, similar to agricultural cooperatives, where rural farmers join together to collectively ship their crops to urban markets. A group of citizens could likewise get together and pool their personal data under a data-cooperative governance model. They would elect board members and establish a governing body to run the co-op. The co-op would set rules on who can access the data, what kind of algorithms can be run and so on.

I think the data-cooperative model has a lot of potential because citizens would be directly involved in governance, unlike in today's data ecosystems. Twitter users may feel they are involved in Twitter but they actually are not. Twitter is a company whose first priority is its shareholders. A data co-op would have a legal fiduciary obligation to its members. Its first priority is its members' interests.

Let me give you hypothetical example of a data cooperative organized by Uber drivers in Boston. Let's say Uber drivers in north Boston and south Boston currently cannot compare their incomes per mile with each other. What if they each were to get a copy of their trip logs, income records and so on from Uber and then form a data cooperative for Boston Uber drivers? If there are tens of thousands of drivers, the co-op would have enough data to do analytics for the benefit of its members.

Hayashi: Do you think any governments or other parties need to put forth a legal framework or regulations for establishing a data cooperative?. And wouldn't data cooperatives need a business model?



Hardjono: In terms of rules, data cooperatives could be modeled after credit unions. In the US, there are two types of consumer-facing financial institutions: banks and credit unions. Different laws apply to each. In terms of business model, data co-ops would presumably generate revenue by internally performing computational algorithms on their data and providing the resultant insights to external parties.

Returning to my hypothetical Uber driver example, let's say the drivers drive 50,000 or 60,000 miles annually on average. So they would need a new car every 3-4 years or so. Their data co-op could very easily perform computations to identify cars that will soon need to be replaced. With the mediation of the data coop, the auto dealers could then offer the co-op members a discount commensurate with the number of cars to be purchased. Data co-ops' purpose is to generate benefits, but not necessarily monetary benefits, for their members.

The key points are that a data co-op must have (1) enough members to compile a sufficient dataset, (2) a governance model that is working for the benefit of its

members and (3) voluntary membership.

Use of data consortiums to solve societal problems

Hayashi: Under capitalism, financial institutions have fulfilled the role of winning trust through strict regulation. I suspect that under data capitalism, certain entities will be called upon to play a role similar to data cooperatives. Do you envision any existing financial institutions turning into something like a data cooperative?



Hardjono: When you're talking about companies, the proper term is "data consortium," not "data cooperative." Data consortiums are composed of member companies typically having some sort of common interests. The problem is that companies belonging to data consortiums tend to be competitors of each other and they want to keep their data confidential from competitors. So how do you get a group of competitors to collaborate for the benefit of the entire consortium? The solution is to use technology to gain insights from the member companies' respective data while maintaining the data's confidentiality. This process might involve providing data, agreeing on some rules of behavior and collaboratively developing algorithms.

Historically, the consortium model has been successfully used in the credit card space. Open credit card networks are consortiums of many banks. They have voluminous rules, mostly in the form of member agreements and technical documents. I believe this consortium model could apply to data consortiums also.

If financial institutions do form data consortiums, they should start with very specific types of data like maybe futures trading data. If widely adopted, the data-consortium model could also enable financial institutions to detect early warning signs of financial crises.

Hayashi: NRI provides infrastructure support to many financial institutions. We may be able to play a role in financial data consortiums.



Hardjono: We'd be interested in exploring data consortium prototypes with Japanese financial institutions. In our data analytics projects, we don't necessarily need a lot of data. Even small datasets sometimes yield valuable findings. And we use only anonymized data.

The model with which we've been most successful at MIT Connection Science is projects involving two or three corporate sponsors and sometimes a government body. The South Australian project I mentioned earlier involves the SA state government, a bank and telecom and healthcare companies. MIT plays a matchmaker role, which we can do because we're a neutral organization. We're interested in research. We want to help companies and institutions to benefit society because MIT's mission is to solve humanity's difficult problems. Given an interesting challenge, we'd love to work together with Japanese companies, financial institutions and/or government entities.

Hayashi: That's exciting to hear. Thank you.

about NRI

Founded in 1965, Nomura Research Institute (NRI) is a leading global provider of system solutions and consulting services with annual sales above \$5.1 billion. NRI offers clients holistic support of all aspects of operations from back- to front-office, with NRI's research expertise and innovative solutions as well as understanding of operational challenges faced by financial services firms. The clients include broker-dealers, asset managers, banks and insurance providers. NRI has its offices globally including New York, London, Tokyo, Hong Kong and Singapore, and over 16,500 employees.

For more information, visit <https://www.nri.com/en>

The entire content of this report is subject to copyright with all rights reserved.
The report is provided solely for informational purposes for our UK and USA readers and is not to be construed as providing advice, recommendations, endorsements, representations or warranties of any kind whatsoever.
Whilst every effort has been taken to ensure the accuracy of the information, NRI shall have no liability for any loss or damage arising directly or indirectly from the use of the information contained in this report.
Reproduction in whole or in part use for any public purpose is permitted only with the prior written approval of Nomura Research Institute, Ltd.

Inquiries to : Financial Market & Digital Business Research Department
Nomura Research Institute, Ltd.
Otemachi Financial City Grand Cube,
1-9-2 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan
E-mail : kyara@nri.co.jp

<https://www.nri.com/en/knowledge/publication/fis/lakyara/>
