Private Data for Social Good

2019-07-17

Academia Sinica
Taipei, Taiwan

organized by
The Board of Science and Technology (BOST), The Executive Yuan
and
The Center for GIS, Research Center for Humanities and Social Sciences, Academia Sinica

Speakers:

Nicholas de Cordes (Orange Group; OPAL Project)

Data for Development: Lessons learned from using telecom data in Africa and Latin America

Dr. Chiou, Wen-Tsong (Academia Sinica)

Is an Algorithmic Leap of Faith Rational?

Prof. Peng, Wen-Chih (National Chiao Tung University)

Mining crowd flows and behavior from Geo-Data

Dr. Chan, Ta-Chien (Academia Sinica)

Challenges for GeoAI: Geoprivacy and geomasking

Moderator:

Dr. Chuang, Tyng-Ruey (Academia Sinica)

Good Data: Data for policy, policy for data

Good Data:
Data for policy, policy for data

2019-07-17

Tyng-Ruey Chuang
Academia Sinica

Public Data? Private Data?

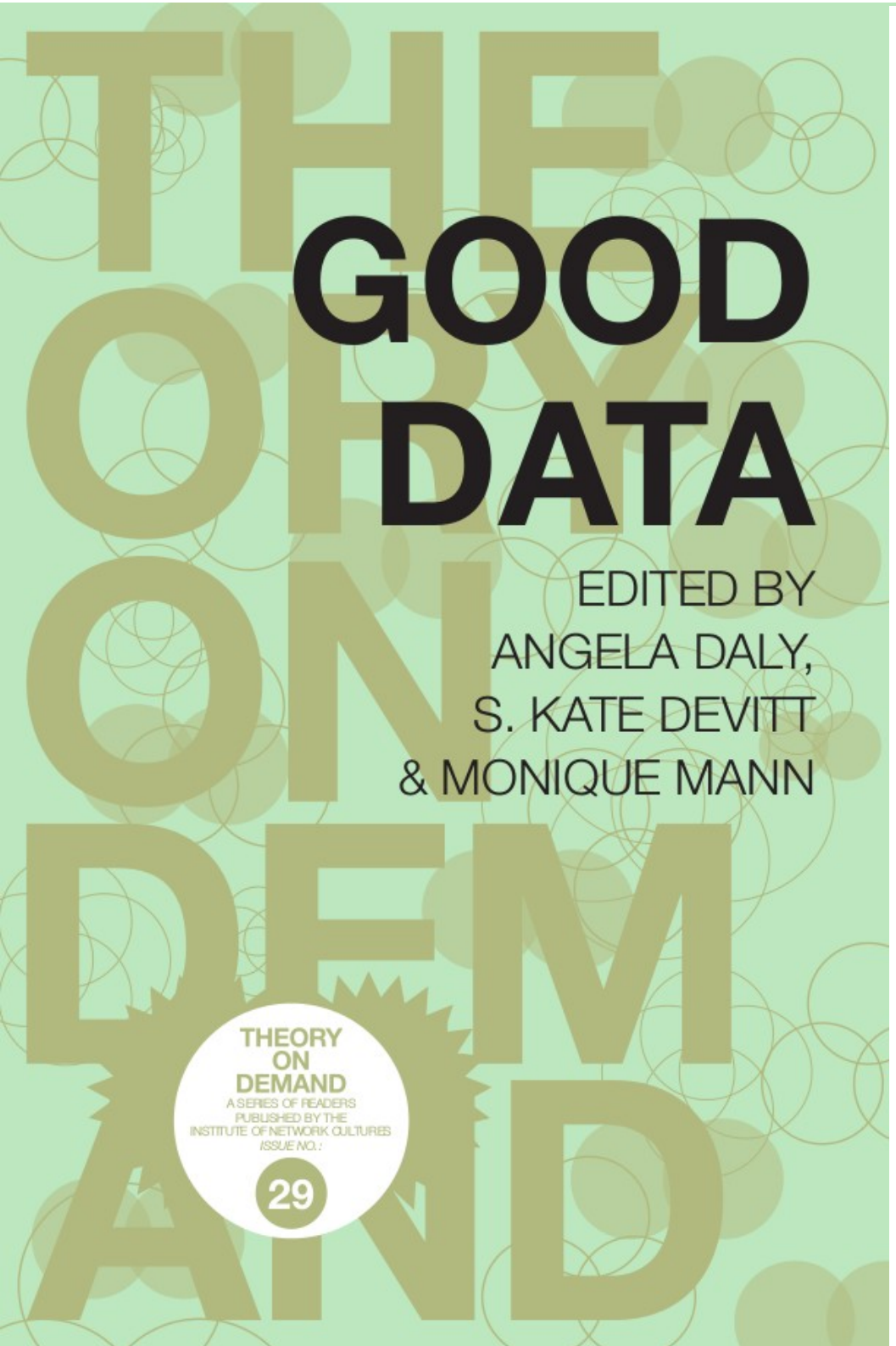
- Public Data: reusable by all without restrictions
  - in the public domain
  - public section information
  - publicly licensed (or CC0) datasets/databases/collections
  - **public datasets need not be accessible or open**
- Private Data: not accessible to others
  - not revealed to others; confidential; restricted
  - may be used by others with conditions (e.g. by contracts)
  - private data is not necessarily personal data and vice versa
  - **datasets produced by corporates are private by default**

Communal Data!

- Data produced and managed by a group of people for their mutual benefits
  - Project Gutenberg; OpenStreetMap; Wikidata
  - Mozilla Location Service https://location.services.mozilla.com/
  - Common Voices by Mozilla https://voice.mozilla.org/
  - “Citizen Science”
  - Publicly licensed (CC BY; CC BY-SA; ODbL); CC0
- Policy for data ← formalized community norms

Good Data

- Data for policy
  - evidence-based policy making
    - Where is the data? Is it accessible? Is it authentic?
  - computational reproducible research
    - replicated research findings ←open data + free software
- Policy for data
  - public money, public code (& data)
    https://publiccode.eu/
  - governance of communal data sharing
    - public licenses alone not sufficient; sharing sensitive data
  - FAIR data
    - Findable, Accessible, Interoperable, and Reusable



12: GOVERNANCE OF COMMUNAL DATA SHARING

CHIH-HSING HO [1] AND TYNG-RUEY CHUANG [2]

Introduction

The rapid development of the data economy calls for innovative research into its social and ethical impacts. When enormous opportunities emerge along with making use of vast amounts of data, challenges are generated and concerns arise around monopoly and market enclosure. Current legal and regulatory frameworks for data protection fail to address these devastating problems. By focusing on consent and the anonymisation of data, these legal techniques echo the neoliberal methods of governance which promise individual autonomy and choice as an advanced liberal strategy. This article proposes theoretical and computational approaches to the analysis of an alternative data sharing model, which is based on community participation in decision making and self-governance. We consider several examples, such as user data cooperatives and collaborative data projects, to further explore how a community is formed and how the governance of communal data sharing is being established. We will then develop frameworks for the governance of communal data sharing by combining common pool resource management and a socio-legal perspective on the commons.

Today we see many states as well as *private* initiatives to promote a *data*-driven industrial revolution across the globe. Data, said to be like oil a century ago, has been cast as a new type of resource fuelling an emerging, lucrative digital-era industry.[3] However, the wealth derived from this digital revolution is not being evenly distributed. According to a study by the Economist, all five of the most valuable listed companies in the world - Apple, Alphabet (Google's parent company), Amazon, Facebook and Microsoft are tech titans.[4] Digital wealth is being monopolized and concentrated in very few hands. Such dominance has led to such side effects as unfair competition, manipulation, routine intrusion of privacy, and the undermining of democracy.[5] These tech giants provide the infrastructure undergirding much of the data economy, and stand to gain the most from it. Although most of their services appear to be free, what underlies the transactions of the digital economy is an exchange of services for control over data. The challenges posed by capitalist accumulation of data raise the question: is this monopoly inevitable?

1 Assistant Research Fellow, Institute of European and American Studies, Academia Sinica, Taipei Taiwan; LLM (Columbia), JSM (Stanford), PhD in Law (LSE), Email: chihho@sinica.edu.tw (corresponding author).
2 Associate Research Fellow, Institute of Information Science, Academia Sinica, Taipei, Taiwan, PhD (NYU), Email: trc@iis.sinica.edu.tw.
3 'The world's most valuable resource is no longer oil, but data', *The Economist*, 6 May 2017.
4 'Tech firms hoard huge cash piles', *The Economist*, 3 June 2017.
5 An example can be illustrated by the Facebook scandal, see: Tam Adams, 'Facebook's week of shame: the Cambridge Analytica fallout', *The Guardian*, 24 March 2018, https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence.