

Regulatory Issues and Approaches to Relevant Data

Professor Dr. Kung-Chung Liu, Lead, Intellectual Property Initiative, CAIDG, School of Law, Singapore Management University/Renmin University of China/Visiting professor, National Law University Delhi, India/Adjunct Chair Professor, Shandong University, China

Outline

Introduction

The typology of data

The protection of the relevant data

The access to the relevant data

Findings and conclusions

Introduction

1. Data have become the raw materials of production, a new source of economic and social values.
2. Also the development of AI has evolved from technology-driven to data-driven
3. The purpose of this talk is to identify regulatory issues of and approaches to data

The typology of data

1. Not All data are alike.
2. Definition of data is not feasible and non-existent
3. Typological approach: to identify types of data that are relevant, not exact, less “systematic” or “scientific,” with some overlap and flexibility
4. Whenever we talk about regulation of data, we have to be very mindful about what type of data we are talking about

Data in the Context of AI

We can identify three types of data:

data specifically generated for AI, data tailor-made as the training materials for machine learning;

big data with huge size (volume), rapid change (veracity) and update (velocity), and various data sources (variety);
and

copyright-protected data

Regulatory Issues of Data

1. Standardization?

A unified standard for data format, granularity, structure, internal organization, quality, data portability and interoperability.

With a unified standard of data format market players can use data from other sources without data transformation.

2. The more data, the better?

The generation, collection, agglomeration, processing and dissemination of data as much as possible

3. The trading of data

4. The cross-border flow of data and data sovereignty

Personal data and Its Regulatory Issues

Divergent national laws on personal data protection

As of April 2016 there are 108 countries which have implemented personal data protection laws governing collection and cross border transfer of personal data

Big Data and Its Regulatory Issues

1. **Proprietary entitlement over big data?** Position Statement of the Max Planck Institute for Innovation and Competition did not see any need and justification of creating exclusive rights in big data.
2. **Protection against tortious conduct via “regulation against unfair competition”**
3. Big data as the single most valuable capital of our society for economic and technological innovation are of crucial importance to our human being. **Therefore big data must be allowed to form, accumulate, process and free flow to the maximum extent possible, in order to maximize its values**

Big Data and Its Regulatory Issues

4. data localization

According to the study by Information Technology and Innovation Foundation (ITIF) data localization, which requires certain types of data must be stored within the country, has been adopted by some 34 countries.

Big Data and Its Regulatory Issues

Data trade: Free and fair

According to one study, cross-border data flows grew by 45 times between 2004 and 2014 and generated \$2.8 trillion in global economic revenue in 2014 alone.

If there are other market-related obstacles existing and preventing parties from conducting data trade in addition to standard (technical) obstacles

Access to big data

1. Indeed, scale economy and network effect would exhibit in data market and favor market concentration and dominance. This would be more obvious in two-sided markets where most consumers with data are concentrated in limited big platforms. Competition scrutiny is necessary to avoid data monopoly and to lower the entry barriers for small companies. Access to data under competition law can be obtained as a remedy in cases of abuse of market dominance.

2. However, as the Position Statement rightly concludes, **antitrust law in most cases is not suitable to address the issue of access to big data.** Rather, if at all, **sector-specific regulations might be required.**

Public-sector data and Its Regulatory Issues

A large part of big data comes from public sector, including public-funded education and R&D institutions, public utilities, government agencies etc.

The various **open government (public) data initiatives** are of high importance. The **OECD Council 2008 Recommendation on Enhanced Access and More Effective Use of Public Sector Information (PSI)** suggests openness, access and transparent conditions for re-use, quality control (to ensure methodical data collection and curation practices to enhance quality), and safeguarding the integrity of PSI

EU has recently promulgated **Directive (EU) 2019/1024 on open data and the re-use of public sector information** (20 June 2019)

To access proprietary data

1. With current legal tools, including flexible contractual arrangements and trade secret protection law, and technical measures, data holders could keep control of access to their data specifically generated for AI by granting license.
2. However, it is a completely different issue **whether and to what extent do we need to enhance the access to data specifically generated for AI.**
3. If data specifically generated for AI are PSI, the rationale discussed in “the protection of public-sector components of big data” should apply. However, if data specifically generated for AI are from private sector, the rationale discussed in “the protection of private-sector components of big data” could be applied by analogy.
4. We also need to explore if there are scenarios in which the use of competition law and/or ex ante regulation will be needed to ensure access to data specifically generated for AI that from the private sector and are of critical importance (essential facility?).

To access copyright-protected data

In 2019, the EU adopted the **Directive on Copyright** and related rights in the Digital Single Market and introduced specific clauses, which allow reproductions and extractions for TDM. However, the Directive foresees substantial limitations:

1. If TDM is done by research organisations and cultural heritage institutions, it must be for the purposes of scientific research, which excludes commercial uses.
2. If TDM is done generally by all those who enjoy lawful access to underlying mined materials, copyright holders have the opt-out option

In 2018, Japan amended the Copyright Act (effective since 2019) and introduced detailed clauses exempting related activities in data mining and analysis.

To construct a specific fair use clause that is more conducive to AI: preferring the UK and Japanese approach over the EU approach, no opt-out etc.