DATA SECURITY IN INDONESIA: BIBLIOMETRIC ANALYSIS OF THE DEVELOPMENT OF PERSONAL DATA REGULATORY

Hanna Nurhaqiqi [1,*], Ratih Pandu Mustikasari [2], Kusnarto [3]

[123] Universitas Pembangunan Nasional Veteran Jawa Timur
[1] hanna.nurhaqiqi.ilkom@upnjatim.ac.id*; [2] ratih.pandu.ilkom@upnjatim.ac.id ;
[3]Kusnarto.ilkom@upnjatim.ac.id

* corresponding author

ABSTRACT

The Regulation on Personal Data Protection was passed at the end of 2022. Personal data is very significant in this modern era. The world of communication today is closely related to personal data. The internet network makes data an expensive and invaluable asset. Many journals research personal data and current problems. This study identifies and analyzes research developments regarding the Regulation of Personal Data. The research method is Quantitative Descriptive. Data were obtained from 200 journals from Google Scholar using Publish or Perish software in the last five years. This data is then visualized through the Vos Viewer software. The results found are seven clusters, 35 items, and 284 links. The keywords that appear are Privacy, Data Privacy, Data, Security, etc. If there had been more research on Data Security on Blockchain in the previous five years, the research trend would have been more on Privacy Concerns, Users, and Personal Data.

1.Introduction

The development of information technology creates an independent, critical and dynamic information society. This has made digital activity increasingly intense and strengthened in the last two decades. The rate of information flow on the internet continues to grow, and not only for communication. Changes and shifts in communication today make it easier for someone to communicate without knowing the boundaries of space and time (Shawila Nolanda Destiano Lestari, 2021). However, the existence of the information itself becomes so valuable and priceless. This user information, namely Personal Data, then needs to be protected by the state. In order to maintain stability and protect the public so that personal information does not fall into irresponsible hands, information spread on the internet can also trigger verbal abuse or cyberbullying. Misuse of data and inadequate skills in using technology will be a loophole for misusing the person's identity and the user's character (Holle, 2019).

The development of information technology and the speed of spreading digital information is related to the freedom of human behaviour that tends to make people do things

they like, so they become addicted (Rahardaya, 2022). Personal data is also related to personal needs; in this case, it can be categorized as public Personal Data that moves from the real world to the digital world. This personal data is then managed and protected by every Government in every country. Protecting personal data is essential for the public because it includes complex things, such as diverse cultural, economic and social differences. Furthermore, this information is to assist both the Government and the community administratively in obtaining more effective and efficient public facilities.

Information is about personal identity in ID cards and digital information in Applications. The crowd information on social media gives Social media a chance to construct individual needs (Ratna Puspita Sari, 2022). Discussing this, regulating personal data is considered a representation of a country in providing legal and constitutive attention to protecting its citizens. This paper also covers more than just government actions in protecting personal data. Nevertheless, citizens and social media management companies are also involved in developing personal data protection. Social media interactions also include digital information on what to talk about around society digitally (Dwi Prasetyo, 2023). It also deserves the attention of researchers to find out how far the law has met the dynamic needs of society and its power in regulating the public.

In this regard, this study tries to see how the development of research studies related to "Data Privacy" in the last ten years. Research not only on aspects of Government but also on any aspect that discusses Personal Data. This study presents a bibliometric presentation by providing data sets that are appropriate and relevant to the development of study trends because they are displayed based on the citation level of each article. An explanation of the sequence of data processing and analysis will then be explained in the research methods section.

Personal Data and Digital Security

The development of digital technology provides extraordinary opportunities and challenges to today's information society. Freedom of access to information and communication, making interactions and activities unlimited and open. This enables every level of society to pursue any activity as long as they get adequate knowledge. Activities undertaken to enter the virtual world are like opening doors to the digital world. By using an account as initial access in cyberspace, the public or users must create an account such as a digital version of KTP (Resident Identity Card).

According to Law No. 27 of 2022 concerning the Protection of Personal Data, Personal Data is data about individuals who are identified or can be identified separately or combined with other information directly or indirectly through electronic or non-electronic systems. Creating an account to be able to enter cyberspace indeed uses personal data. The more accounts you create, the more it is necessary to manage and protect personal data. As stated in the same law, Personal Data Protection is the overall effort to protect Personal Data in the Personal Data processing chain to guarantee the constitutional rights of Personal Data subjects. It should be underlined that Personal Data is included in the Constitutional Rights protected by the Government.

Personal Data Protection activities have become part of the company's business marketing strategy in convincing buyers and potential buyers to remain as consumers (Saura, Ribeiro-Soriano, & Palacios-Marques, 2021) and referred to as the Data-Driven innovation strategy used by a number of companies today. It has become a trend to gain a wider audience by understanding the importance of user privacy in the digital market. This is also the ideal foundation for the Government to gain strong trust from the public, namely by protecting the security of the digital data of its residents so that it does not cause anxiety and can provide public comfort and peace.

Regulation and Role of the State

The data theft scandal of tens of millions of Facebook social media user accounts was carried out by Cambridge Analytica in March 2018. This was allegedly one of Trump's winning strategies as President of the United States at that time. It was recorded that 70.6 million Facebook users from the United States leaked (Aria, 2020). Even Indonesia was not spared from the data leak, namely the number of 1.1 million users; this is not a small number. Facebook's stock price immediately slumped in the same month. However, there are other highlights. Instead, the Government's role has become so central and primary in protecting data for the public.

Not long ago, the United States Congress and the European Parliament summoned Zuckerberg and explained the case. The following year, the United States Federal Trade Commission (FTC) imposed sanctions with a fine of US$ 5 billion or the equivalent of Rp. 70 trillion, to Facebook in relation to the Cambridge Analytica case. This case also warns government authorities in other countries, such as the European Union, by creating the General Data Protection Regulation (GDPR).

The Indonesian Government, through the passage of the Personal Data Protection Law last year is expected to be a fundamental constitutional step to continue to monitor and regulate the pace of security of the Indonesian people's personal data. This requires public transparency, in which the public must have the ability and understanding of the importance of data security and its designation in living as a civil society (Isaak & Hanna, 2018). Referring to Article 58 paragraph (1), namely, the Government has a role in realizing the implementation of Personal Data Protection in accordance with the provisions of this Law.

One example is privacy in big health data. Administrative ease and bureaucratic efficiency in obtaining easier health services. It turns out that it is accompanied by consequences of the vulnerability of digital health data (II & Cohen, 2019). With the Health system built by the Government since 2011, namely the Social Security Administering Body, the protection of personal data is crucial and needs to be monitored by the public. Has it been running well in accordance with the law mandated? Alternatively, are there obstacles in the field that the authorities must address? This research study can serve as a reference in seeing how far the development of research related to Personal Data has been going on in the last decade.

2. Method

The implementation of this research was carried out by studying bibliometric analysis. Bibliometric analysis is included in the quantitative method by analyzing bibliographic data that has been collected previously. Bibliographic data is usually mixed using software to make it easier for researchers to obtain data according to the required research criteria. It can also be collected manually, which requires more time. A bibliography, commonly called a Bibliography, is very much needed in conducting research, deepening ideas, and tidying up the systematics of the research document itself. According to Limestone University Library (Library, 2023), a bibliography is a data set from sources used for a series of research works. This bibliography generally includes the author's name, title, name and location of the issuing organization/institution, publication date, and page on the required source.

Bibliographic data collection techniques in this study used software. There are various kinds of software for obtaining bibliographic data; one is Publish or Perish (PoP). Anne-Wil Harzing, a Middlesex University, UK professor, designed this PoP device. The birth of this device was originally intended for completing studies and was used internally. Until finally, it was inaugurated as open-source software so that academics worldwide could use it for free.

Figure 1. Publish or Perish
Source: Harzing.com

Publish or Perish, PoP can collect bibliographic data through eight sources, namely Crossref, Google Scholar, Google Scholar Profile, OpenAlex, PubMed, Scopus, Semantic Scholar and Web of Science. Then using Citation Metrics, this PoP can be obtained further. Consists of: Publication years, Citation years, Papers, Citations, Cites/year, Cites/paper, Author/paper, and various intensity indexes for each paper obtained. Furthermore, data collection has been completed, and visualization of data that has been stored using the RISmanager file is performed. Data visualization through software, one of which is Vos Viewer. This Vos Viewer will provide an overview of the distribution of bibliographic data based on files that have been stored according to the keywords the researcher wants.

Figure 2. Vos Viewer
Source: Vosviewer.com

The Vos Viewer tool, which originates from the Netherlands, can also be accessed free of charge in constructing and visualizing a data set. Displays a bibliometric network that can make it easier for researchers to discover what research is developing in a certain period. Keywords in research are very important and significant because they are a source of readability of studies from time to time. The linkages between networks have different narratives, and researchers then analyze this ability due to bibliometric analysis. Therefore, this research will sequentially retrieve the keyword "Data Privacy" on the Scopus page search with a span of the last ten years, namely, 2013-2023, under the maximum limit allowed by PoP devices, namely 200 Scopus-indexed articles. All of these data Bibliographic data is stored in the form of a RIS Manager file—furthermore, data visualization using Vos Viewer. The results obtained are analyzed and discussed in the Results and Discussion section.

Figure 3. Data Collection Technique
Source: Researcher's Data, 2023

3. Results and Discussion

The results and discussion of the bibliometric analysis will be explained in four sections. First are the data results obtained from the Publish or Perish device. Second, analysis of the tables compiled by the authors based on the distribution of the number of publications over the last ten years. Third, the ten journals with the highest number of citations. Fourth, Cluster and Network Analysis.

(1) Citation Metrics Publish or Perish: "Data Privacy"

Based on the explanation in the Research Methods section. This Bibliometric analysis begins with Publish or Perish (PoP) tools with the Keyword "Data Privacy" in the last ten years. Bibliometric data collection is taken from Scopus-indexed article pages with a maximum number of 200 articles.

The following are the results of the data collection obtained.

Figure 4. Citation Metrics PoP
Source: Researcher's Data, 2023

Seen Publication years display 2013-2022 because in 2023 there are still no articles found that match the keywords. The citation metrics data shows that the number of journal citations is 27,926, with an average citation of 139.63 times for each journal. Specifically for searching Scopus-indexed articles, an Application Programming Interface (API) is required by clicking "Request Key API" from the Scopus page. The code obtained is entered into the PoP device. This device also displays the number of journal citations as a bibliographic data set feature. Namely, h-index: 103; g-index: 158; hI, norm: 103; hI, annually: 10.30; hA-index: 35. Each article has been cited at least 103 times. This shows that the need for articles on Data Privacy is relatively high among academics.

(2) Distribution of Annual Publications
The table below shows that the distribution of article publications from 2013-2019 was quite volatile. Even so, the minimum production level is 23 articles. This shows that the productivity of researchers related to Data Privacy remains high even though it fluctuates. It will be different in the following year, namely from 2020 to 2023. The data shows that there has been a sharp decline. The decline in the level of scientific publications. This is suspected because, at that time, there was a pandemic. Based on the data, this year, namely 2023, there still needs to be Scopus-indexed research/article that discusses Data Privacy.

In 2013, there were 23 articles discussing Data Privacy. The article with the highest citation at the time was "AppIntent: Analysing sensitive data transmission in Android for privacy leak detection" (Yang, 2013). Together with other articles, this year discusses the ability of technology to maintain data confidentiality from leaks. Then it continues from Health data security with the title "Big Healthcare data: preserving security and privacy" (Abouelmehdi, 2018) to political preferences with the title "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection" (Isaak J. , 2018).

Until 2022 related to covid and data security with the title "Towards Secure and Privacy-Preserving Data Sharing for COVID-19 Medical Records: A Blockchain-Empowered Approach" (Tan, 2022).

Figure 5. Publication Distribution 2013-2023
Source: Reseacrher's Data, 2023

Figure 6. Publication Distribution 2013-2023
Source: Researcher's Data, 2023

(3) Journal with the highest citation based on Publish or Perish

   The table below shows the publications with the highest citations from the Publish or Perish tool database. There are 1553 citations to the article "Decentralizing privacy: Using blockchain to protect personal data". This paper discusses three things: Data Ownership, Data Transparency and Auditability, and Fine-grained Access Control (Zyskind, 2015). The discussion regarding blockchain and data security which its users can only access is implemented in all big data-based applications. This knowledge is general and must be understood by all big data companies and their derivatives. Then the second article with the most citations is "Privacy-preserving multi-keyword ranked search over encrypted cloud data", with 812 citations (Cao, 2014). This paper describes the strict set of rules for maintaining data confidentiality and the integration system for each user. In the next article, "Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control". This article describes health-related applications that follow blockchain protocols and rules that must be followed (Yue, 2016).

Table 1. Most Cited Journals

| Rank	Title	Cites
1	Decentralizing privacy: Using blockchain to protect personal data	1553
2	Privacy-preserving multi-keyword ranked search over encrypted cloud data	812
3	Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control	745
4	ProvChain: A Blockchain-Based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability	490
5	Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?	473
6	Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT	379
7	Privacy in the Age of medical big data	332
8	A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT	330
9	Identity-Based Remote Data Integrity Checking with Perfect Data Privacy Preserving for Cloud Storage	315
10	Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud	312

Source: Researcher's Data, 2023

(4) Clustering and Network Analysis with Vos Viewer

The data set that has been obtained through the Publish or Perish device is then stored in the form of a RISmanager file. Then it is visualized through the Vos Viewer to carry out an analysis regarding the development of the Data Privacy study. The figure below displays a network between research through interrelated keywords and begins with "Data Privacy". In 200 journals, there are seven clusters, 35 items and 284 links. Next, the analysis will be continued for each cluster as follows. The five clusters can be seen from the network in each colour, namely green, blue, red, yellow and purple. The five colours represent different topics.

Cluster 1, coloured red, has nine items: big data, big data privacy, covid, internet, IoT, privacy preservation, thing, user and work. Big data becomes the hub with the most networks in this cluster, connecting all other cluster colours with as many as 27 links. Then in cluster 2, the colour green has eight items: blockchain, challenge, cloud computing, data security, data sharing, privacy, privacy protection and security. It can be seen that the point "Privacy" has the most links among other items in the green cluster, totalling 33. Then in cluster 3, the blue colour has eight items. Namely data protection, future, health data privacy, personal data, privacy concern, protection, research and study. Cluster 4, yellow, has seven items. Namely concern, data, data privacy, data privacy concern, researcher, review, and use. The last three clusters only have 1 item in each. Cluster 5, purple, article. Cluster 6, Tosca colour, cloud. Cluster 7, orange colour, data aggregation.

The research results above indicate that future trends may be on Data Privacy. Where the branches are privacy protection, health data privacy, and security, in each article, the trend discusses its relation to the regulations running at that time. Because the behaviour of Data Privacy is relevant to implementing personal data policy regulations and other related regulations, given this, knowledge of ongoing policies will support public openness and accountability.
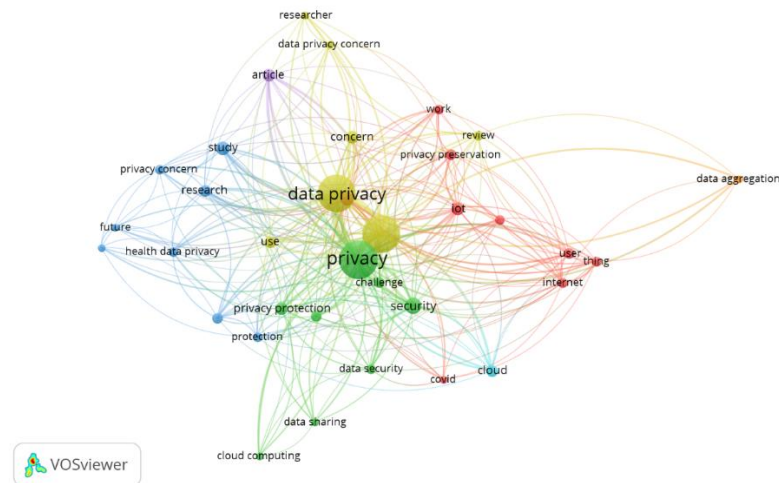
Figure 7. Data Vos Viewer Visualization
Source: Researcher's Data (2023)
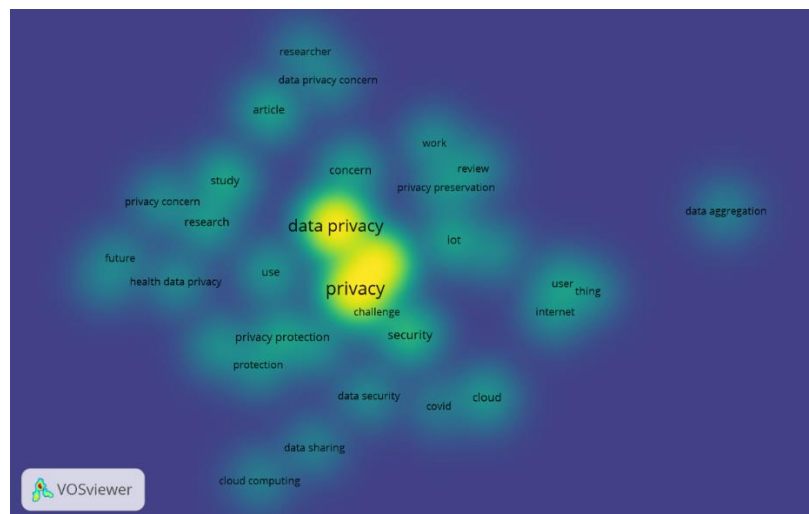
Figure 8. Density Visualization
Source: Researcher's Data (2023)

4. Conclusion

Research development is very dynamic and adapts to the needs of the environment. Bibliometric analysis is one way to discover research trends that occasionally emerge. Even if the narrative is precise, it can also predict future research trends.

Collecting the required research articles, if done manually, will take quite a long time. Then a data collection tool is needed to facilitate researchers. There are various kinds of software available. Researchers use Publish or Perish and Vos Viewer tools in this paper to collect, visualize, and analyze data.

This study retrieved 200 Scopus-indexed articles with the keyword "Data Privacy". The period is 2013-2023. The results show that the trend from 2013 to 2019 fluctuated but was still productive by being above 23 articles each year. Until 2020 there has been a sharp decline, and even in 2023, there has been no research related to Data Privacy. This can be identified by the presence of a pandemic that began in 2019.

This study is expected to help researchers and academics studying data privacy regulation. Even so, this article must continue to be updated, given the dynamic movement of

technology. Regulations continue to evolve and adapt to the times. Each country does its best to manage and protect public personal data.

References

Abouelmehdi, K. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*.

Aria, P. (2020). *Cambridge Analytica dan Peran Negara dalam Perlindungan Data Pribadi.* Katadata.co.id.

Cao, N. (2014). Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on Parallel and Distributed Systems*.

Dwi Prasetyo, T. S. (2023). Celebrating Virtual Existence on Social Class and Social Media. *Journal Spektrum Komunikasi Vol.11 No.1* , 75-84.

Holle, A. N. (2019). Anonymous Chat as a Cyberbully Impact. *Jurnal Spektrum Komunikasi Vol.7 No .1 Juni*, 51-61.

II, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 37-43.

Isaak, J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *The Policy Corner. IEEE Computer Society*.

Isaak, J., & Hanna, M. J. (2018). *User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection.* The IEEE Computer Society.

Library, L. U. (2023, February). *Citation Guide.* Retrieved from Libguides.limestone.edu: https://libguides.limestone.edu/c.php?g=545011&p=3736441

Rahardaya, A. K. (2022). CONTROLLING FREEDOM OF BEHAVIOR USING DIGITAL TECHNOLOGY WITH DIGITAL DETOX. *Journal Spektrum Komunikasi. Vol.10 No.3 September*, 276-283.

Ratna Puspita Sari, J. K. (2022). Tourism Journalism as A Regional Tourism Promotion Media. *Journal Spektrum Komunikasi. Vol.10 No.4 December*, 378-384.

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marques, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*.

Shawila Nolanda Destiano Lestari, N. A. (2021). BANDUNG STUDENT'S SELF-DISCLOSURE BEHAVIOR THROUGH ON INSTAGRAM. *Jurnal Spektrum Komunikasi Bol.9 No.1 Juni*, 66-77.

Tan, L. (2022). Towards Secure and Privacy-Preserving Data Sharing for COVID-19 Medical Records: A Blockchain-Empowered Approach. *IEEE Transactions on Network Science and Engineering*.

Yang, Z. (2013). AppIntent: Analyzing sensitive data transmission in Android for privacy leakage detection. *Proceedings of the ACM Conference on Computer and Communications Security*.

Yue, X. (2016). Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. *Journal of Medical Systems*.

Zyskind, G. (2015). Decentralizing privacy: Using blockchain to protect personal data. *Proceedings - 2015 IEEE Security and Privacy Workshops, SPW.*