



WIPO/IP/CONV/GE/21/INF/4
ORIGINAL: ENGLISH
DATE: DECEMBER 20, 2021

WIPO CONVERSATION ON INTELLECTUAL PROPERTY (IP) AND FRONTIER TECHNOLOGIES

Fourth Session
Geneva, 22 to 23 September 2021

SUMMARY OF FOURTH SESSION

Document prepared by the WIPO Secretariat

INTRODUCTION

1. The WIPO Conversation provides a global, leading forum to engage with and facilitate discussion and knowledge on the impact on IP of frontier technologies, including artificial intelligence (AI).
2. WIPO began this series of conversations in 2019 as a way of convening a diverse group of stakeholders to exchange information, build knowledge and support well-informed policy choices in an era of digitalization and technological transformation that is rapidly reshaping our economies and societies.
3. Frontier technologies represent a vast set of new, emerging technologies that operate at the intersection of scientific breakthroughs and practical implementation. It is estimated that frontier technologies currently represent a \$350 billion market, predicted to grow to a \$3.2 trillion market by as soon as 2025.¹
4. While artificial intelligence (AI), which the first three sessions of the WIPO Conversation focused on, remains central to these shifting dynamics, AI is only one of many frontier technologies such as the Internet of Things (IoT), robotics, blockchain and others that are transforming innovation, creation and business models. As such, the scope of the WIPO Conversation was expanded in 2021 to cover all frontier technologies.
5. Moreover, the digital transformation has been accelerated by the COVID-19 pandemic. Intangible assets and intellectual property (IP) are fast becoming key economic drivers, and the global intangible asset value is at an all-time high of US\$ 65.7 trillion.² This highlights the relevance and importance of the WIPO Conversation of building awareness, sharing information and making the discussions around IP and frontier technologies understandable and accessible to a wide set of stakeholders.
6. All frontier technologies are fueled by accelerating digitalization, rely on data and often generate data. For example, the current advances in AI are being driven by three main factors: advances in computing power, better algorithms, and the availability of data. In 2017, it was reported that 2.5 quintillion (2.5×10^{18}) bytes of data were produced every day, while 90 percent of the world's data had been generated in just the previous two years, illustrating the exponential increase in data.³
7. As data is the driving force behind all frontier technologies, including AI, and due to its importance in the digital economy, the topic of the fourth session of the WIPO Conversation on IP and Frontier Technologies was titled "Data: Beyond AI in a fully interconnected world". The session focused on why this intangible asset increasingly matters and is changing how we do business, innovate and create.
8. The fourth session of the WIPO Conversation on IP and Frontier Technologies took place over the course of two days, on September 22 and 23, 2021, in a virtual format. There were over 1,300 attendees from 110 countries, which underscores the extensive interest in and importance of the topic.
9. The present paper summarizes the discussions undertaken during the fourth session.

¹ UNCTAD/TIR/2020, Technology and Innovation Report 2021, <https://unctad.org/webflyer/technology-and-innovation-report-2021>

² Brand Finance Global Intangible Finance Tracker 2020, <https://brandirectory.com/reports/gif-2020>

³ Domo (2017), Data Never Sleeps 5.0, <https://www.domo.com/learn/infographic/data-never-sleeps-5>

OPENING

10. The fourth session of the WIPO Conversation was opened by WIPO Director General, Mr. Daren Tang. The Director General emphasized that it is crucial to understand the nature and value of data in order to design appropriate regulatory frameworks for the emerging data ecosystems.

11. As an intangible asset, data highlights the increasingly important role IP plays in the global economy and raises a number of complex questions for the international IP system. The Director General stated that these considerations must continue to be front and center of the emerging debate at WIPO, as well as broader policy debates at the international and regional level.

12. In holding this Conversation, said the Director General, WIPO seeks to bring together the global innovation community to raise awareness for how data relates to the global IP ecosystem and how to strike a balance between protecting data rights and encouraging data sharing.

STRUCTURE OF THE FOURTH SESSION

13. The session comprised four expert panels (each including two to five speakers led by a moderator and followed by a Q&A portion) and an open floor discussion.⁴

14. Day 1 was structured around three panels setting the wider scene of what data is, why it is key to fueling future economic and social development and how it is regulated by current frameworks. Day 1 concluded with a number of innovators and creators explaining the role that data plays in their business models and the challenges they face.

15. Day 2 then took a deep dive into data and IP in the form of a panel followed by an open floor discussion. The day centered on how data fits into the existing global IP system and whether the current IP provisions are sufficient.

PANEL 1: DATA, BEYOND AI IN A FULLY INTERCONNECTED WORLD

16. The first panel, moderated by Mr. Dean Joliffe, set the scene by introducing data in a broader economic context, how it drives many elements of Industry 4.0, and why it is important to discuss it in a regulatory and IP context.

17. Mr. Dean Joliffe opened the discussion with an example from Odisha, India, where the use of data played a key role in minimizing the devastation caused by cyclones. Although data had been available previously, it was not until this was compiled into interoperable and accessible systems and analyzed to produce insights that action could be taken and thousands of lives saved. Such success demonstrates the creation of value from data, but uncertainty remains around how to make value creation replicable, measurable, and based on reliable inputs.

18. Ms. Diane Coyle proposed that value is created when information obtained through data is used to do something different from (and more beneficial than) the status quo. Data has monetary/economic (analytical) value, but also social (contextual/informational) value that is hard to measure directly but becomes evident in decision-making ability. She provided an example of Marriott vs. Airbnb, where the value of data became evident in retrospect when studying the difference in decision-making patterns between the two companies.

⁴ See the agenda for further details https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=550571

19. Ms. Aruba Khalid highlighted that the value of data is linked to the scale at which data-driven insights can be derived and applied. Insight generation is linked to new advances in storage technology, improvements in analytical software, and the ubiquity of data collecting devices. Insights can create high value out of previously low-value data by analyzing it for the purpose of, for example, cost efficiencies, improved processes, a better understanding of behavior, and highly personalized products. The full value of data is not known until it is put to a specific use, including new business models and innovations that are yet to be invented. For some companies, such abilities can build strategic advantage and keep out new entrants, creating the need for regulation to maintain a balance.

20. When asked if different types of data need different regulations, Ms. Coyle said that, in industry, private players might come up with their own governance arrangements and that national technical standards will play a key role in providing oversight. For public data, such as urban data on transport or air quality, the institutions that could put this data to use may lack funding to do so. Policy interventions may therefore be needed to promote access to data and value creation from that data.

21. Ms. Khalid confirmed that the scale and rate of growth of Industry 4.0 sets it apart from prior industrial advances, where it took longer for society to recognize (and benefit from) a certain innovation. Data plays a role in this as the interconnectedness of new data-driven technologies, such as AI, cloud computing and 3D printing, drives scale and leads to much faster adoption.

22. The panel concluded by discussing key barriers to increased data use and the interventions that would unlock greater value. Agreement was found around the importance of building skills to use data and promoting data science education, democratizing access to data currently held by a small number of companies or countries, and balancing private interest and social benefit. It was also noted that building new types of institutions to achieve these balances and adapt social contracts around data use would be needed.

PANEL 2: THE REGULATORY MATRIX OF DATA

23. The second panel, moderated by Mr. Aaron Shull, considered the multiple regulatory frameworks that can be applied to data, depending on the interest or value that is sought to be regulated. Regulatory approaches can also differ across cultures. This panel introduced policy elements such as ethics, data for the public good, security, privacy and cultural approaches to data.

24. Mr. Shull framed the panel discussion in terms of the public policy questions that arise around data ownership and the ownership of algorithms built upon it, which drive wealth creation, accelerate productivity, and can potentially disrupt business models. Companies derive value from their ability to collect, organize, control and commercialize stores of data, and their methods for doing so can blur the lines of responsibility between institutions and individual actors. The resulting risks, including security breaches, discrimination, and loss of autonomy, pose a significant challenge due to the need to regulate complex and interdependent policy areas.

25. Mr. Kung-Chung Liu proposed three different classes for data: data generated for AI training, big data, and copyright-protected data. Standardization is needed to increase the generation and collection of data, and to enable interoperable processing. However, such standardization can be problematic for data considered private. Issues of data trade, data localization, data sovereignty and cross-border data flows must also be considered, as lack of regulation or overregulation can pose an obstacle to innovation. The importance of making big data available for social and economic benefit is so high that regulation should generally favor open access to data. Mr. Liu suggested that current anti-trust laws are not suitable to address

the issue of access to big data and too time consuming to implement, therefore sector specific regulation is needed.

26. Ms. Caroline Wanjiru Muchiri focused on contextualizing data, the different circumstances in which personal data is collected and appropriate frameworks for data governance. Developing countries have often adopted GDPR-like policies despite a context that is very different from Europe, and it may be worth considering a more locally applicable approach. Some local policies in Africa, such as the introduction of digital IDs in Kenya, have actually increased personal data collection. The Global South has unique data owners, such as religious institutions and humanitarian/development agencies, which hold data stores that are not fully governed by local laws or shared with local authorities. Therefore, unique frameworks that reflect the social, political, and international context are needed.

27. Mr. Jiro Kokuryo discussed how the East and West differ in their cultural approach to data and how this, in turn, influences the balance between perceived individual and social value. The western approach sees the individual pursuit of happiness as the route to social good, therefore leaning toward individual ownership of data. However, traditional markets created for the exchange of non-shareable goods and property rights may not be efficient for data due to the unique, non-rival nature of data. With data, an efficient market may be one where both the contribution and reward come to and from society as a whole, a “potluck” economy. As a result, it is necessary to re-examine and re-prioritize fiduciary duties. Employing data stewards as central agents within a data regulation framework could ensure that loyalty to the data subject is balanced with benefits to the potential data user and society as a whole. Re-thinking the philosophical foundations and ethos of the digital economy is necessary to preserve the dignity of all parties.

28. Ms. Dafna Feinholz shared the UNESCO Recommendation on the Ethics of AI.⁵ These principles were created to address AI readiness from design to implementation. Ethics is about making explicit the values that lie behind policies and how and who decides “what is good”. As AI evolves, it must constantly be accompanied by ethical considerations to ensure that human rights and freedoms are protected. The UNESCO Recommendation includes elements of data ethics and regulation, including (a) developing data governance strategies that ensure continuous evaluation of training data, proper security, and feedback mechanisms; (b) application of privacy by design and regular privacy impact assessments; (c) designing effective accountability mechanisms; (d) promoting open data and (e) collaborative platforms. It offers two innovative tools to member states: (a) a methodology for assessing AI readiness and (b) a framework for evaluating impact, risk prevention, and redress mechanisms.

29. Ms. Muchiri commented on how data governance frameworks can benefit vulnerable communities and uphold human rights. She pointed out that it is important to recognize that the value of data lies in both the collective and the individual. Minorities need to be proactively included in the data market, with deliberate data collection incentives and institutional structures. Mechanisms need to be designed that assure benefits of data accrue to vulnerable populations, that data is used only for the intended purposes, and there are avenues of redress and enforcement where necessary.

30. On the topic of how data localization and data sovereignty fit into the frameworks discussed, Mr. Liu stated that this depends on the greater question of whether data belongs to the individual or to society and will need international negotiation. Many institutions charged with regulation are pre-internet and not ready to address these issues. Generally, data localization runs counter to the principles of the data economy and the drive to generate, collect and

⁵ <https://en.unesco.org/artificial-intelligence/ethics>

process as much data as possible. This is because data sovereignty and data localization slow everything down through property rights and national borders, respectively.

31. When asked whether there exists a holistic framework for data governance, Mr. Kokuryo responded that the best approach is through implementing fiduciary duties and mediating conflicts of interest, thereby assuring that data collectors retain loyalty to the data providers. Technology exists to verify contributors and fairly allocate acknowledgement and rewards. Alternative frameworks, such as self-governance or regulations by anti-trust laws, are not practical because they create information overload or only apply to large players, respectively.

32. Ms. Feinholz proposed that the unique context of public data requires public entities to safeguard such data and use it only for specific reasons. Therefore, public data must be treated with vigilance and care. However, it is also important to ensure the free flow of information to maximize social benefits, including access to open data and international cooperation. The UNESCO Recommendation sets out ground rules for a balanced approach.

33. To close the panel, each speaker summarized their view of the key elements in data governance frameworks, including equity, fairness, and inclusivity (Ms. Muchiri); fiduciary duty to one another (Mr. Liu); dignity, loyalty, altruism (Mr. Kokuryo); and international cooperation, solidarity, and transparency of purpose (Ms. Feinholz).

PANEL 3: DATA AND BUSINESS MODELS – A BUSINESS VIEW

34. The third panel, moderated by Ms. Clara Neppel, provided insights into how innovators and creators use data and what part IP (and other regulatory frameworks) plays in their enterprises.

35. Ms. Clara Neppel started the discussion by highlighting the business and technical challenges of data and AI. All AI solutions require data as an input and generate output ranging from, for example, language creation to moving a robotic arm. The input data can be sensitive or proprietary, but business models exist to secure such data. Usually, these models require some form of compensation, which reflects the intended purpose and use of the data. Licensing provisions commonly restrict the use of data or technologies for the intended purpose. Protecting the AI output is less straightforward as it is not directly created by humans and therefore may not be protected under traditional IP laws. Developing an AI model itself can represent a significant investment of resources, as does training the model. Some models are currently protected behind firewalls, others remain open source, and some are partially available for reuse and adaptation to specific tasks. There is an interesting interaction when it comes to ownership, access and control. Patent protection is a possible way to protect AI systems, but patents are generally only available for technical solutions and would generally not cover data.

36. Three speakers presented their use of data in their business models and the IP challenges that they face.

37. Uncanny Valley describes itself as a music, sound and technology collective. It has produced a proprietary AI model that uses sound and music samples and allows users to generate new music content and find new revenue streams. Mr. Charlton Hill shared Uncanny Valley's example of producing the winning entry of the 2020 AI Song Contest using that software. However, he also highlighted that the company could only assign creative rights to the human songwriters and were unable to assign any rights to the AI model that had also been a significant creative contributor to the song. This exposed the legal uncertainty around augmented creativity. By way of example, the data used to train Uncanny Valley's AI model comprised over 200 songs. There are open questions regarding the use and ownership of this data and it is important for the future to make sure proper accreditation is possible to avoid piracy problems and an inability to keep up with technology.

38. Mr. Javier Fernandez of CropLife Latin America introduced Agriculture 4.0, a new paradigm of digital farming using mobile technologies, remote-sensing services and distributed computing to improve crop production and decrease environmental impact. New technologies can significantly improve farming practices by collecting and processing data, leading to better decision making by farmers, optimizing use of resources such as seed selection, and reducing negative environmental impact. Visualization of farmlands is key; however, much of the imagery and crop data is proprietary and subject to restricted access. Analytical algorithms can also be covered by patents and copyrights.

39. Mr. Saar Safra shared the work of BeeWise, a company that creates a physical structure for bees where they can be monitored by AI algorithms and taken care of in real-time if they show any signs of distress, without the human caretaker needing to be physically present. Taking care of bees is critical to maintaining biodiversity as bees pollinate 75% of all edible crops, and their colonies are in severe decline. Data and IP are very important to BeeWise's business model as the data collected by the bee housing structures powers AI algorithms that are protected by patents.

40. Two further speakers, Ms. Lucie Arntz and Mr. Shuo Wang, presented potential solutions for data-driven businesses to help deal with data privacy and ownership issues.

41. Ms. Lucie Arntz of Aspheris explained that the company's service uses federated learning, a solution for training AI algorithms on non-centralized data. This method allows an algorithm to learn securely using proprietary sets of data individually, without the need to move data into central storage and hence without compromising data security. The algorithm identifies correlations in each dataset and then combines these into one overarching model. This service presents a technical solution to the legal challenge of using sensitive data, for example, personal data or data that may contain trade secrets. The resulting algorithmic output cannot be traced back to its source nor reveal any of the individual inputs used for analysis. IP plays an important role for Aspheris as the company has IP rights over its proprietary algorithm, but the data sets used may be protected by the IP rights of other entities.

42. Mr. Shuo Wang introduced Bryea, a big data provider from China, and also discussed the challenge of training AI algorithms while maintaining the confidentiality of data sets. Bryea helps its customers digitalize their services using IoT devices to collect unstructured data and edge computing for data analysis. The lack of unified data can be an issue for analysis. Some types of data sources and storage mechanisms cannot be easily unified, creating "stovepipe" systems or data islands. Data ownership is another challenge, including potential legal uncertainties concerning the ability to share or cache data breaching access authorization. These challenges can be solved through edge computing and the middle platform. Edge computing means that data is processed close to the source and not moved to a centralized location. The middle platform provides analytics through an application layer, establishing a real-time connection between a data source and the data lake, thereby not changing the location of the original data and maintaining the security, integrity and quality of the data.

43. Ms. Neppel invited each speaker to share their response to the following question: Thinking about data and business IP, what challenges do you face, and what would you like to change or have more certainty on?

44. Mr. Hill responded that solving the problem around creative data and how data can be reused without losing attribution to the original – for example, in a way similar to the way songs can be "covered" – would be important. Another challenge is tracking creativity to its source since AI manipulates content away from its origin, while allowing remuneration to flow to the artists.

45. Mr. Fernandez focused on harmonization within the industry and clarity in the debate around AI and patentable subject matter. He also mentioned the need to clarify whether any of the existing data sets are proprietary or open source and the role IP should be playing in this space.

46. Ms. Arntz agreed that clarity around IP protection of databases is needed, as current regulations do not seem to account for how difficult it is to build up stores of data or recognize how valuable this is.

47. Mr. Safra stated that there is a lot of data already flowing, but distributed systems will soon be needed to compile the data, keep it safe, robust and consistent. Therefore, clarity around standardization protocols would be very valuable.

48. Mr. Wang responded that there generally is an issue of data overload rather than a lack of data. Therefore, it is important to balance IP protection for data with access to data to generate future innovation based on data.

49. Ms. Clara Neppel closed the session by recalling that it is also important to keep in mind that the right social contracts must be in place to balance the needs of different stakeholders, including businesses, where it is appropriate to have proprietary solutions and what should be in the public domain.

PANEL 4: DATA IN THE CURRENT IP SYSTEM

50. The fourth panel, moderated by Mr. Igor Drozdov, set out to provide an overview of how IP applies to data and how IP fits into the wider framework. With this broader focus in mind, the panel also sought to address whether the current IP framework would need adaptation or whether there are situations in which IP could present a barrier for innovation related to data.

51. Mr. Igor Drozdov started the discussion by recalling that copyright law was created to protect the originality of expression rather than a source of inspiration, such as data. Patent law comes closest to protecting raw data, as it covers the collection of information needed to reach a certain result. With the assistance of machines, seemingly useless data has become valuable, and with this change, there has been a movement to commercialize it. At first, the regulatory reaction was focused on protecting databases as original structures through copyright. Later, the information itself was also included under *sui generis* database protection in some jurisdictions to create an investment incentive. The previous panels highlighted two trends: advocacy for the free flow of data as input and inspiration and IP protection as an incentive to collect it. This panel then discussed the options to protect data in the current IP system.

52. Mr. Bret Hrivnak began by making the distinction between raw and structured data. Under certain circumstances, intangible assets may be protected by IP rights, including patents, copyrights and trademarks. The social policy underlying the IP framework is to encourage creativity and to compensate for the investment required to generate inventions. Patents protect useful innovations with industrial application; trademarks are the backbone of branding; and copyright protects the original expression of an idea, but not the idea itself. Patents do not protect data itself but can protect how data is used, applied and created. Single data points are facts or statistics and are therefore not eligible for copyright protection as single units. However, there is a limited opportunity to protect data under copyright, and that is as a database, due to the creative selection and arrangement of data in a database. In the European Union (EU), *sui generis* rights can also grant protection for a database if a substantial investment was made to create it. The alternative is protecting data as a trade secret and relying on contract law to protect the agreements necessary for secrecy. There is a need to consider greater harmonization as part of the conversation on protecting data.

53. Mr. Taiwo Oriola reiterated that data can only be protected if it fulfils strict legal conditions. Data is the physical representation of information in a manner suitable for interpretation by humans or computers. Legally, there are two types of data, proprietary and non-proprietary data, which can both be protected under different laws. These laws have very strict definitions. For example, a trade secret must have commercial value, and its owner must prove that reasonable steps have been taken to keep it secret. Personal data, such as biometric data, does not fall under copyright; however, it may be protected under common law against unauthorized use under the law of confidence. Copyright does not protect facts *per se*, but only the expression of ideas. However, as mentioned previously, *sui generis* rights can protect databases. Technical information, if it represents an invention, can be protected by a patent. With the increased importance of data and regulation of access to its commercial use, it will be critical to identify when certain data falls under the current definitions for IP protection.

54. Ms. Tatiana Eleni Synodinou focused on the EU's *sui generis* right for the protection of databases. The EU database directive was issued in 1996,⁶ making it one of the oldest and most innovative in the area of computer and copyright law. It was the first step toward protecting the non-original contents of a database (and the substantial investment made by database creators) from extraction or replication of that database by an unauthorized third party. Since then, *sui generis* rights have been the center of fierce academic debate on the dangers of information appropriation and the necessity of data ownership. With the rise of AI, the question becomes: could or should *sui generis* laws extend to protect data produced by an AI system. Currently, the *sui generis* right is a distinct, autonomous legal framework from copyright law; therefore, lack of creativity or originality is not an obstacle, neither is the question of authorship in the case of AI-generated data. However, a 2004 ruling of the EU Court upheld a restrictive interpretation of the *sui generis* right: that it applies only to the investment made in obtaining data and not to resources used for the creation of new data.⁷ Therefore, the eligibility of AI-generated databases for *sui generis* protection depends on whether the output is data creation or the arrangement of existing data. On the one hand, the function of AI is to discover existing patterns in existing data. On the other hand, the algorithm itself can be seen as creating new information, and it is still unclear which will become the prevailing view in legal protections. Nevertheless, the question of whether the use of AI to arrange data in a database qualifies as a substantive investment would still apply.

55. While Ms. Synodinou focused on data as an AI output, Mr. Carlo Scollo Lavizzari discussed input data obtained by text and data mining (TDM) and its importance in the development of AI. Quality AI outputs require quality input data, which may be protected by copyright. Some copyright protected data, such as dictionaries, is already being used as an input for AI. Data on its own does not mean very much without its unique context. The great advantage of text mining is that it can combine content from various contexts by serving as a type of translation service. To fully advance AI, collaboration is necessary between rights holders, researchers, corporate R&D, and commercial toolmakers who develop TDM products to create an ecosystem with incentives to provide high-quality input data. However, it is claimed that greater advances in AI are impeded due to the uncertainties in accessing copyright data for TDM. This is because of the lack of harmonization of fair use provisions under copyright law or narrow TDM exceptions. Article 3 of the EU Digital Single Market (DSM) Directive includes exceptions for TDM.⁸ It distinguishes non-commercial and commercial uses of data and offers

⁶ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009>

⁷ CJEU (9 November 2004). British Horseracing Board v William Hill Organization, C-203/02. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A62002CJ0203>

⁸ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC <https://eur-lex.europa.eu/eli/dir/2019/790/oj>

safeguards to ensure viable markets of text mining are created while researchers can continue TDM. Article 4 of the DSM directive also allows mining for commercial purposes as long as reserved rights are respected. Both the US and EU systems agree that copying for TDM is an infringing activity that requires licensing or an exception.

56. Ms. Elisabeth Kasznar Fekete addressed how data fits into international trade secrets regimes, such as TRIPS.⁹ Importantly for the consideration of data, trade secrets can protect the data itself or the contents, configuration and assembly of a database, not just the exact expression of data (as is the case with copyright), against unfair competition. However, trade secrets do not grant exclusive rights like a patent. Trade secret protection can be invoked in an action against breach of legal provisions, such as negligence, or breach of contract. In addition, they can be enforced through civil, criminal or labor law penalties depending on the jurisdiction. Trade secret protection can only be enforced when information is disclosed, acquired or used without the owner's consent and in a manner contrary to honest commercial practices. Data can also be safeguarded by non-disclosure agreements (NDAs) covering, for example, R&D, franchising, licensing, and M&A activities. However, NDAs are only binding on the contracting parties and cannot restrict the activities of third parties or the free flow of data. Technology impacts trade secret protection because data is now much easier to collect, use, reproduce and delete. Therefore, today's contracts are more precise and must distinguish trade secrets from consumer or personal data. It is important to protect efforts made in compiling data because it incentivizes business, and unused data is generally released to the public anyway. To avoid harming the IP system, predefined types of royalty-free licenses for data and systems that fast track communication between data owners and potential users should be considered.

57. Mr. Drozdov asked each speaker to reflect on whether the current IP system strikes the right balance between protection and the free flow of data.

58. Mr. Hrivnak said it was important to improve harmonization and to provide flexibility of protection for databases not only based on creative aspects but also on the considerable investment made by developers.

59. Mr. Scollo Lavizzari pointed out that in the EU, the *sui generis* database right was an attempt to harmonize between different countries and make it clear that sufficient investment was enough to afford protection. However, the creation of new IP rights, just as the wholesale exemption of certain technologies from IP rights, must be undertaken with extreme care. He also agreed with Mr. Hrivnak's point on harmonization.

60. Ms. Kasznar Fekete responded that she believed the current system was flexible enough to adapt and that there was no need for a complete change in the laws. Rather, a focus on how to use the current system, such as fast-tracking licensing, would be preferred.

61. Ms. Synodinou closed the session by saying that the IP system is in constant evolution, and we see some jurisdictions permitting better access to data while others focus more on protection and security. More empirical research is needed to understand if *sui generis* rights should be extended or if new methods are needed to strike the right balance.

DISCUSSION: IS THE CURRENT IP SYSTEM FOR DATA SUFFICIENT?

62. The Chair, H.E. Mr. François Rivasseau, welcomed participants to give further individual statements from the floor. The focus of these interventions was on whether current IP laws are sufficient to protect data, and if not, what adaptations may be necessary.

⁹ Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Section 7: protection of undisclosed information, https://www.wto.org/english/docs_e/legal_e/27-trips_04d_e.htm

General Considerations

63. The ICC shared its recent publication of Innovation Principles for Policy frameworks.¹⁰ These focused on three key elements: (1) effective, enforceable, and predictable IP rights enabling innovation and entrepreneurial risk-taking; (2) importance of trustworthy business environments; (3) promoting multi-stakeholder perspective in all data and IP discussions.

64. The UK IPO informed participants that it is holding an open consultation for innovative companies to share their experience in working with IP laws and investing in training AI systems.¹¹

Data Regulation in Japan and the Republic of Korea

65. Several speakers provided examples from Japan and the Republic of Korea and the recent new legal provisions concerning data regulation. Views were expressed both in favor of such efforts as well as in underscoring certain challenges that persist.

66. Japan introduced legislation titled “Shared Data with Limited Access” under the “Unfair Competition Prevention Act” in 2018, which is separate from trade secret protection. Under this regulation, the unauthorized use of data that is not protected as a trade secret but satisfies certain requirements can lead to a claim for damages and/or an injunction. Japan considered a new data right as a possible alternative to the current IP system but found that exclusive rights would be too restrictive and curtail innovation. Therefore, Japan opted for a defensive strategy that allows recourse in the case of malicious acts while maximizing overall data utilization. It is the view in Japan that the current IP system provides sufficient protection and does not need reinforcement; however, more steps should be taken towards making data broadly available.

67. The Republic of Korea has closely observed the developments in Japan and has also proposed new legislation. The main concern is regarding copyright infringement when data is extracted by TDM and used for training AI algorithms. The lack of a clear legal definition for “data ownership” and inconsistency with civil law definitions may also cause misunderstanding. Although efforts to establish IP rights recognizing ownership of data should be very cautious, it is necessary to consider protection.

68. A counterargument was also presented, stating that the Republic of Korea’s current IP system already protects data too tightly and that the recent amendments will impede innovation since copyright protection covering data used in TDM runs counter to the overall ambition toward greater innovation.

Copyright, TDM, Database Rights and Trade Secrets

69. Copyright protection covering the raw data used in AI training datasets and TDM was a central theme through many of the interventions. Arguments were presented both for and against enforcing copyright protection on data used in AI training. Concerns were raised regarding overly aggressive restrictions hampering AI output, as well as challenges with tracking and penalizing infringement and fairly distributing benefits.

70. One speaker pointed out that, historically, the idea underpinning an invention or creation has not been subject to IP protection. It is only the elaboration, expression or implementation of ideas that is protected. Fundamental to IP law is the freedom to use information, or data, which forms the building blocks of the creative and inventive process. Given the numerous public

¹⁰ <https://iccwbo.org/global-issues-trends/innovation-ip/innovation/>

¹¹ Artificial Intelligence and IP: Consultation on copyright and patents legislation, <https://www.gov.uk/government/news/artificial-intelligenceand-ip-consultation-on-copyright-and-patents-legislation>

policy motives to increase access to data, it is necessary to be cautious in creating an additional IP right, although there is a need for increased international dialogue and harmonization in applying existing provisions. It was also noted that the suggestion of depositing data for patent applications would place an undue burden on patent applicants and would make the patent system less attractive, instead incentivizing trade secrets.

71. A counterpoint was made that AI systems trained using artistic copyright works clearly represents an infringement. Any subsequent work generated can be considered an adaptation and if the results are commercialized this constitutes exploitation without any benefits accruing to the rights holder. In the EU, there is an exemption to copyright infringement for TDM for scientific research. However, if algorithms trained under this exception are later commercialized, this should retroactively negate the exemption. The recording industry has been addressing this challenge in various ways, including a broad range of licensing solutions to facilitate both creative and distribution processes. However, record-keeping and transparency obligations on those using third-party works or protected subject matter are needed to facilitate detecting infringements committed in the course of technical processes.

72. Proponents of stronger copyright protections also voiced that recent amendments passed in Japan and the Republic of Korea favor AI at the cost of creators, and similar developments should be avoided.

73. An alternative interpretation of using TDM on copyright data was proposed – specifically, that there is no copyright infringement because the informational value derived through TDM was not explicitly present in the original input. Therefore, AI output is original content and can be copyrighted itself. The speaker also called for global harmonization of *sui generis* rights to avoid a scenario in which contractual obligations lead to fragmentation.

74. In a separate point, an appeal was raised for the crucial nature of transparency in auditing, de-biasing, and explaining AI algorithms, which has been a persistent challenge. Protecting AI training data as trade secrets prevents reasonable review and oversight, which, for example, in the case of using AI algorithms to inform criminal proceedings, can violate due process.

Technology gap

75. The audience was reminded to keep present the perspectives of developing nations and the Global South in the discussion on data. The benefits of data generally accrue in highly industrialized countries, but IP regulations should not reinforce the technology gap by creating insurmountable barriers to entry. For less industrialized countries, their role as data providers should be reciprocated via affordable and reasonable licensing and proactive technology transfer to avoid data colonialism. Equity, fairness, and redress, as well as efforts to remove biases in AI should also be a part of the discussions.

Data Ownership

76. Several speakers called for improved legal definitions for data ownership in order to facilitate greater understanding and accountability between parties working with data, as well as to standardize certain concepts with the aim of future harmonization in data laws.

77. It was highlighted that clarity regarding data ownership is critical to drawing up effective contracts for business collaborations around data. Where one party is responsible for processing data owned by another party, this lack of clarity can lead to uncertain liability. Furthermore, it needs to be defined when a data set is to be considered “new”, and who is the owner of a data set created from merging several datasets would be. There was a reiteration of the concern that ownership can be regulated by contracts, which may be convenient for

individual parties, but that contracts create a fragmented landscape of ownership. Outputs also remain confidential, which is contrary to the aim of open innovation.

78. “Raw data” also needs a comprehensive legal definition in order to effectively conclude contracts. Data is a non-rivalrous good, and the current status of the legal discussion is that raw data cannot and should not be assigned to an owner, so no property rights should be introduced. A proposal was made for WIPO to take up the matter of mediating access to competition relevant data, not only where there is a question of dominant market position. For example, a recent decision under German anti-trust law has created a specific rule in such cases.¹²

79. The difference between database ownership by legal persons and the proprietary rights over personal data by private persons also need greater clarity in definition. Currently, only privacy rights are provided to natural persons under the GDPR, but this creates a widening gap between the legal concepts of privacy and property, which are intertwined. A proposal was made to define a broad legal term for this kind of data, one which straddles property and privacy in order to facilitate the process of creating a new purpose-built regulatory system, for example, “life data”, containing the aggregate data footprint of a natural person both online and offline. This would facilitate the integration of personal data into markets and the treatment of personal data as the property of natural persons.

80. Finally, further clarity was requested regarding the legal liabilities related to data storage and whether this poses an act of reproduction for copyrighted material, as well as to clarify legal definitions for who owns AI output when addressing situations of possible infringement.

81. Many speakers agreed that there is a need for further harmonization among local laws in order to facilitate cross border data flows and international collaboration. Particular attention was called to the matter of business data, which often passes through different jurisdictions because of multi-national operations and partnerships. If such data is unstructured, it can only be protected as a trade secret, which limits collaboration, so a proposal was made to further study how to facilitate sharing while maintaining rights over data.

Data Use by IP Offices (IPOs)

82. The potential of IPOs collecting and sharing new, IP related datasets was discussed, and proposals were made for the type of information and new tools that could be used to facilitate the better functioning of IPOs.

83. Some of this data collected by IPOs is never published, such as that related to patent applications that are refused. A proposal was made to establish a WIPO working group for preparing such supplementary datasets for collaborative use. This would be in line with the current WIPO initiative to establish in-house data tools and best practices.

84. Business intelligence at the IPOs has transitioned from being led by IT experts to responding to on-demand requests from operational teams, which requires new forms of governance. Users empowered to perform data operations need training and there is an ambition to raise data literacy levels and create an empowered analytics community. To provide the necessary tools, the EUIPO is creating an agile, secure data platform that will act as a single point of access for data – internal or external – and deliver harmonized and reusable data sets, including a data catalogue and relevant metadata.

¹² Bundesgerichtshof, Mitteilung der Pressestelle, Nr. 80/2020, KVR 69/19 - Beschluss vom 23. Juni 2020, <https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&Datum=2020-6-23&nr=107146&linked=pm&Blank=1>

85. Machines have great advantages when processing huge amounts of data, and trained AI will soon be able to assist in the review of patent applications. When this occurs, it is very important to consider whether such tools will have legal protection of their own because publication of such AI algorithms may bias the process of future patenting.

CLOSING

86. Assistant Director General, Mr. Ken-Ichiro Natsume, closed the session by thanking all participants. Mr. Natsume presented two main takeaways. First, data is critical to powering frontier technologies and the debates around access, protection, security, privacy, and ownership are critical, always with multi-stakeholder perspectives in mind. Second, regulation must strike a delicate balance between fostering the free flow of data on one side and protecting rights and investment on the other.

87. Mr. Natsume announced that webinars will be organized in which Member States will share their data strategies, laws and approaches relevant to IP and data. The first webinar will discuss the national data strategies, laws and approaches relevant to IP and data in Japan and the Republic of Korea.¹³ Furthermore, the WIPO Secretariat has begun updating the AI and IP Strategy Clearing House to include data and IP information, currently representing input from 73 member states.¹⁴

88. Finally, Mr. Natsume proposed possible topics for the next sessions of the WIPO Conversation in 2022, including how data and IP technologies can improve the functionality of the IP system itself. IP Offices are starting to take advantage of data and frontier technologies to provide services to stakeholders, including enhanced ability to search for and register IP rights. Furthermore, it may make sense to look at concrete and practical areas, such as approaches to handling AI inventions, including examination guidelines.

[End of Document]

¹³ Presentations and video on demand: https://www.wipo.int/meetings/en/details.jsp?meeting_id=67128

¹⁴ The AI and IP Strategy Clearing House is available at https://www.wipo.int/about-ip/en/artificial_intelligence/policy.html#clearing_house