Available Online: 23 July 2026 DOI: 10.54254/2753-7102/2026.35616

# Can the concept of derived data subsume data intellectual property rights in China?

*Guangyu Liang*

Jishou University, Jishou, China

lianggy_1@163.com

**Abstract.** The 2026 Guidelines for the Registration of Data Property Rights place derived data within a unified framework for data property registration. Although derived data and data intellectual property rights overlap in subject matter, the former concerns the recognition and allocation of rights in processed data, while the latter concerns whether and how an intellectual contribution should receive legal protection. The overlap arises because both regimes often concern lawfully processed and value-enhanced datasets. Nevertheless, they answer different legal questions. Data property registration identifies interests in holding, using, and operating data and supports valuation and transactions. Data intellectual property rules concern whether the selection, arrangement, or processing of data embodies a sufficient intellectual contribution. The two systems should therefore be institutionally coordinated rather than merged. Treating the derived data framework as a substitute for the data intellectual property regime would therefore confuse value enhancement with intellectual creation, weaken the connection with copyright and international rules on data compilations, and risk treating routine processing as sufficient for data intellectual property protection.

**Keywords:** Data Property Rights, data intellectual property, derived data

## 1. Introduction

Several localities have developed data intellectual property registration schemes to facilitate data transactions and provide support for rights claims. These local experiments responded to a practical gap. Data had become an important factor of production, but there was no unified national framework for identifying data-related interests, stabilizing transaction expectations, or issuing standardized evidence that could be used in licensing and valuation. Local intellectual property authorities therefore used existing administrative institutions and registration platforms to record processed data collections and support market transactions.

Against this background, the 2026 Guidelines for the Registration of Data Property Rights (Trial) raise a new question: does the national data property registration system replace this path by relying on the concept of "derived data"? The question is important because the Guidelines do more than create a filing procedure. They place derived data within the broader policy framework that separates the rights to hold, use, and operate data. Once derived data becomes an expressly recognized object of registration, it may appear that data collections previously registered as data intellectual property can simply be reclassified under the national data property system.

Because derived data is formed through professional processing, modeling, analysis, and information extraction, it resembles the processed and valuable data collections covered by local data intellectual property registration frameworks. Yet overlap in the data objects concerned does not mean that the two systems are identical in legal nature or function. One system principally asks who may hold, use, process, circulate, or commercially exploit a body of data. The other asks whether the intellectual contribution embodied in a data collection deserves protection against unauthorized appropriation. These questions may concern the same dataset, but they call for different standards and lead to different legal consequences. This article first explains their conceptual overlap and then argues that differences in doctrinal foundation, protection threshold, institutional function, and international coordination require them to remain distinct.

## 2. Conceptual overlap and functional differentiation

The Guidelines separate the rights to hold, use, and operate data. The right to use data includes using it "to form derived data through processing, aggregation, analysis, and other methods." According to the Guidelines, derived data means data formed when a data processor, with respect to data over which it enjoys a right of use and while protecting the lawful rights and interests of all relevant parties, uses professional knowledge to process the data, conduct modeling and analysis, extract key information, or employ other methods to bring about substantive changes in the content, form, or structure of the data, thereby significantly enhancing its value.

This definition contains three elements. First, the processor must have a lawful basis for using the source data and must respect the legitimate interests of relevant parties. Second, the processing must produce a substantive change rather than a mere copy. Third, the change must significantly enhance the value of the data. The concept thus recognizes that lawful processing may produce a data object with a legal and economic identity distinguishable from its source for registration and transaction purposes, even though the resulting data remains technically dependent on that source. The principal function of the derived data framework is to determine the status of processed data within a system of circulation. This is especially useful where raw data is supplied by one party, processed by another, and commercialized by a third. The separation of holding, use, and operation allows interests to be allocated among multiple contributors without forcing every data-related claim into the traditional categories of ownership or intellectual property.

However, local data intellectual property rules address a related but different legal question. Hangzhou is particularly instructive because it has established both local data property and data intellectual property registration systems. Its definitions therefore reflect an institutional effort to distinguish between the two categories. Hangzhou defines data intellectual property rights as follows: "The rights and interests enjoyed by a data processor in a data collection that has been lawfully acquired, formed through processing according to certain rules, and possesses practical value and the characteristics of an intellectual achievement." This definition also requires lawful acquisition, processing, and practical value, but it adds an intellectual contribution requirement. Economic value and technical transformation are therefore insufficient by themselves. The inquiry must consider whether the data collection embodies a sufficient intellectual contribution.

These functional differences do not eliminate overlap at the level of the data objects concerned. Local institutional practice likewise indicates that data property rights and data intellectual property rights may overlap in their protected subject matter [1]. The same data collection may fall within both frameworks. However, derived data emphasizes substantive change and value enhancement, whereas data intellectual property rights ask whether intellectual input satisfies the relevant threshold for legal protection. Their legal

standards and effects therefore differ. In other words, the derived data framework primarily addresses the legal status of processed data, while data intellectual property rights concern the nature and scope of the protection that may be granted to intellectual contributions embodied in such data.

## 3. Institutional coordination and the limits of registration

Conceptual overlap makes coordination necessary, but it does not justify absorption. A single data product may simultaneously involve transactional interests, intellectual contributions, rights in source data, personal information, trade secrets, and copyright. The legal system must therefore preserve distinct routes of analysis. Otherwise, a simplified registration label would conceal rather than resolve the substantive conflicts.

The distinction between the two systems first affects their relationship with existing intellectual property protection. Treating data products that may embody intellectual contributions solely as derived data could obscure, or in practice displace, the application of copyright and related forms of intellectual property protection. In (2007) Hai Min Chu No. 5904, merchant profiles prepared for restaurants were recognized as original works; in (2017) Jing 73 Min Chu No. 1914, a digital navigation map was also protected. These cases show that the data-based nature of a product does not exclude copyright protection. Data products must still be examined for originality in their selection, arrangement, expression, or structure. These cases do not imply that every database is copyrightable. They show instead that characterization must focus on the protectable elements of the particular product. A merchant profile may contain original expression, while a navigation map may reflect creative choices in selection, organization, symbolization, or presentation. If such products were examined only as derived data, the legal test could shift from originality to value enhancement. An economically oriented concept would then displace a doctrinal standard designed to separate protected intellectual creation from unprotected facts and routine labor.

The distinction is equally important for China's coordination with the international intellectual property framework. Article 10.2 of the Agreement on Trade-Related Aspects of Intellectual Property Rights provides: "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such." This provision thus places qualifying data compilations within an intellectual creation framework, particularly where their selection or arrangement constitutes an intellectual creation. If China were to transfer such subject matter wholesale from the intellectual property framework into an experimental data property framework, the domestic system could lose its doctrinal correspondence with established international rules. The concern is therefore not simply that the concepts would differ, but that an abrupt change in the institutional route could make cross-border coordination and recognition more difficult. Article 10.2 further separates protection for the compilation from protection for the underlying data or materials. This layered approach is essential because one database may contain unprotected facts, copyrighted works, personal information, confidential business information, and an original selection or arrangement. Intellectual property law protects only the qualifying intellectual structure and leaves the legal status of the underlying elements to other rules. The derived data concept, which evaluates transformation and value of the data product as a whole, is not designed to reproduce that layered analysis.

Domestic institutional development also demonstrates why the two systems should be coordinated rather than absorbed into one another. Local data intellectual property registration relied on China's existing intellectual property infrastructure to support transactions in the absence of unified substantive legislation. Existing agencies and application procedures reduced the cost of institutional experimentation. Registration helped applicants describe the data collection, record processing rules, identify the claimant, and establish a

time-stamped record for licensing or later disputes. Its experimental nature, however, means that its concepts remain incomplete [2].

The incompleteness of these local experiments does not mean that the intellectual property route lacks an independent function. Rather, it makes clarification of the possible substantive bases of data intellectual property protection even more important. Possible paths for data intellectual property protection include granting limited exclusivity in exchange for public disclosure [3], and adapting compilation copyright or database protection to data-related needs. Both paths exceed the basic allocation of data holding, use, and operation. Under a disclosure-based approach, a processor that reveals prescribed information about the source, rules, structure, or scope of a dataset might receive a limited right against certain forms of commercial appropriation. Such a right would require rules governing its term, scope, defenses, and the treatment of independent creation. A compilation-based approach would protect only those aspects of selection or arrangement that satisfy an originality or intellectual creation threshold. Although these approaches differ, both concern the basis, form, and scope of legal protection, not merely the identification of who may use or operate data.

The different thresholds for protection provide a further reason why derived data cannot absorb data intellectual property rights. Not all derived data reflects a sufficient intellectual contribution. Derived data may be produced through standard industry techniques of data collection, selection, and storage. Even where processing changes the content, form, or structure of data and increases its value, the result does not necessarily constitute an intellectual achievement [4]. This difference should be reflected in examination. A data property registrar may focus on lawful acquisition, the right to use source data, the processing activity, the boundaries of the resulting dataset, and apparent conflicts with third-party rights. Intellectual property examination must additionally ask whether the contribution meets the applicable originality or intellectual creation threshold, which elements, if any, are protectable, whether the claim improperly extends to facts or source materials, and what limitations apply. The same platform may collect information for both examinations, but the legal tests cannot be identical. Absorption would also create under-inclusion. A carefully curated index or compilation may embody an original selection or arrangement even if the underlying data remains substantially unchanged and thus does not satisfy the Guidelines' definition of derived data. Relying on a single derived data category could consequently treat some routine processing as sufficient for protection while potentially overlooking creative compilations that do not involve substantial transformation.

Finally, although the data intellectual property regime remains experimental, its design can and should draw upon the basic doctrines and principles of intellectual property law [5]. Existing intellectual property principles provide established methods for defining protected subject matter, determining the scope and limits of legal protection, balancing private interests with the public domain, and coordinating new forms of protection with copyright and related regimes. The experimental character of data intellectual property therefore does not justify replacing it with the concept of derived data. Rather, it means that the regime should be developed by reference to the established principles of intellectual property law.

Regardless of the doctrinal path chosen, the regime must address questions familiar to intellectual property law, including the threshold for protection, the distinction between protected intellectual contributions and unprotected data, the scope and duration of exclusivity, limitations and exceptions, and the relationship with independently created or lawfully obtained data. These questions cannot be resolved merely by identifying a dataset as derived data or by registering the rights to hold, use, and operate it.

Accordingly, the continued development of data intellectual property within an intellectual property framework has independent systemic value. It enables new forms of data-related protection to draw upon established doctrines while remaining connected with copyright, compilation protection, and the broader

international intellectual property system. By contrast, incorporating data intellectual property entirely into the derived data category would sever this doctrinal connection and reduce questions of intellectual contribution and legal protection to questions of data processing and value enhancement. This does not prevent procedural coordination between the two registration systems. However, the substantive examination should remain differentiated. Data property registration should identify interests relating to the holding, use, and operation of data, whereas data intellectual property examination should determine whether the claimed intellectual contribution satisfies the relevant intellectual property standards and what protection should follow. Procedural coordination is therefore compatible with, and indeed depends upon, the preservation of their distinct doctrinal foundations.

## 4. Conclusion

Derived data and data intellectual property rights overlap in subject matter but perform different legal functions. The former concerns the status and allocation of rights in processed data, while the latter concerns whether and how intellectual contributions embodied in data should receive legal protection. The two systems should therefore allow for coordinated registration while requiring separate examination. Data property registration may provide public notice and transaction evidence, whereas the protection of intellectual contributions should remain governed by intellectual property principles. The derived data framework cannot absorb the distinct functions of the data intellectual property rights regime.

This conclusion follows from the distinctions identified above. Most importantly, value enhancement is not equivalent to intellectual creation. Derived data may be commercially valuable without embodying a sufficient intellectual contribution, while an original compilation may qualify for intellectual property protection without undergoing the substantive transformation required for derived data. Treating the two categories as interchangeable would therefore risk treating routine processing as sufficient for data intellectual property protection, weaken established copyright analysis, and reduce compatibility with international rules on data compilations. China's national data property registration system should be understood as a mechanism for identifying data-related interests and facilitating data circulation, not as a comprehensive replacement for intellectual property law. The appropriate model is coordinated coexistence: registration infrastructure may be integrated, but the substantive bases, thresholds, and legal effects of data property rights and data intellectual property rights must remain distinct.

## Funding

This research was funded by Scientific Research Foundation of Hunan Provincial Education Department (25C0330).

## References

[1] Cao, X. (2026). Data property rights and intellectual property rights: Complementarity, compatibility, typology, and international adaptation. *Journal of Southwest University of Political Science & Law, 28*(2), 162–181.

[2] Yao, J. (2025). Discriminating between data property rights and data intellectual property rights. *Journal of Shanghai University of Political Science and Law, 40*(1), 65–76.

[3] Cui, G. (2019). Basic theory of limited exclusive right in big data. *Chinese Journal of Law, 41*(5), 3–24.

[4] Hu, K. (2025). Questioning the data intellectual property rights empowerment theory. *Law Science*, (10), 114–131.
[5] Wang, G. (2026). On the prerequisites for data intellectual property registration. *SJTU Law Review*, (3), 78–92.