AI Training Data and Transfer Pricing

By Anthony Pastore, Jason Osborn, and Lucas Giardelli

ANTHONY PASTORE is a Partner at Mayer Brown's Tax Controversy & Transfer Pricing practice. **JASON OSBORN** is a Partner and Co-Leader, International Tax and Transfer Pricing Team at Mayer Brown. **LUCAS GIARDELLI** is a Partner and Co-leader of International Tax & Transfer Pricing at Mayer Brown LLP.

"Data is the new oil." When a mathematician made this observation two decades ago, he was addressing an audience of marketers who were sitting on mountains of data but struggling to "refine" it into something valuable. In the artificial intelligence ("AI") era, hyperscalers, AI labs, and other companies often face the opposite challenge: they know how data can train their models, yet they struggle to find enough of it. One option might be to acquire data from third parties. But a better option could be to obtain it from a foreign affiliate, since that data might be more directly relevant to the company's business.

This article examines the transfer pricing questions that can arise in this space. Suppose a U.S. technology company wants to create a new AI model. It writes the model code and owns the relevant IP in the United States, but it needs specific user or customer data to train the model. This data includes not merely contact information like names and addresses, but also the richer digital footprint that users generate through their online activity: browsing and search history, purchase behavior, location data, and the countless other behavioral signals that platforms collect. In our hypothetical, the U.S. company's subsidiary in Europe owns some or all of this data, so that subsidiary provides it for purposes of training the model. How is this transaction classified and priced for purposes of transfer pricing under Code Sec. 482 of the U.S. Internal Revenue Code?

We will argue that this scenario presents several challenges. First, the user data must be analyzed to see if it fits into any of the categories in the recently finalized Code Sec. 861 digital-content regulations. Second, the user data transaction must be priced under the Code Sec. 482 regulations. And under either the Code Sec. 861 regulations or general tax principles, the form in which the data is *delivered* could be a relevant factor for determining the Code Sec. 482 framework for analysis. The subsidiary could make the data available for download (which could be a tangible transaction or an intangible transaction) or make the data available to the parent on the subsidiary's own systems (which could be a service transaction). Non-tax issues—such as data-privacy laws—might influence which delivery mechanism the company chooses. And uncertainty could arise because cutting-edge technology is must be analyzed under regulations that do not expressly address AI.

The Stakes

User data can be a highly valuable resource to companies creating AI models. Consider this recent headline: "Meta's $14.3 billion investment in Scale AI represents the social media giant's most significant move to secure high-quality training data for artificial intelligence models .... This acquisition addresses Meta's most pressing challenge in the AI race: access to the specialized datasets required to train competitive large language models."[1] This is just one example: other AI labs have been shelling out big money to secure user data. And while these public examples have garnered media attention, related parties may also often transact in user data on an intercompany basis.

Datasets for AI training present some unique characteristics that make them especially hard to classify using available analytical tools:

- *Data is arguably non-rivalrous*. Data can be used by many users for AI training purposes without diminishing its value. Unlike oil, data is never depleted, although its value might naturally diminish over time as it becomes stale. This could perhaps suggest that the provision of data has a lower market value. That said, it is possible to look at data as an "anti-rivalrous" good, which *increases* in value as more users interpret and process it, thereby implying a higher value.[2]
- *Data changes the AI*. The data that trains an AI model can change the model, even if the data is never actually downloaded. A model is trained not by storing memories from the data, but by calculating a mathematical error signal from the data that alters the model's internal parameters and ultimately increases the accuracy of the model's output.[3] Said differently, the model "learns" a lesson from the data, so that it becomes more accurate over time. This could perhaps suggest that the use of the data (no matter the delivery mechanism) is a transfer of intangibles. It might also imply a higher value to the data.
- *Data may be private or public*. Although some AI models are trained on proprietary data (as in our scenario), other models are trained on publicly available, open-source data. Others still are trained on "synthetic" data, which is artificially generated to mimic real-world data.[4] The provision of public or semi-public data might seem unlike a transaction in high-value property and more like a service. Even for that kind of data, however, the process of curating or processing the data could make what was initially public seem more like a trade secret or know-how and therefore a transaction in intangible property.

The stakes here are high. These unique characteristics suggest that the value of user data is uncertain, which gives the Internal Revenue Service ("IRS") and other taxing authorities space to challenge the valuation. Further, under U.S. tax laws, if the provision of user data is or is treated by the IRS as an intangible transaction, then that could unlock the possibility of periodic adjustments under the Commensurate with Income ("CWI") standard.

The transfer pricing analysis depends on how the transaction is classified, and there are at least three possibilities. The transaction could be a transfer of tangible property governed by Reg. §1.482-3, on the theory that the transferee receives only a copy of the data rather than any underlying intellectual property rights. It could be a transfer of intangible property governed by Reg. §1.482-4, on the theory that the data itself constitutes a trade secret or similar proprietary asset. Or it could be a provision of services governed by Reg. §1.482-9, on the theory that the affiliate is merely providing access to the data rather than transferring ownership of anything. Each classification entails different pricing methods and, potentially, different outcomes. Addressing our scenario therefore requires a two-step analysis: classification and then pricing.

Step 1—Classifying the Transaction

Classifying the transaction requires its own two-step analysis. We must determine whether the Code Sec. 861 regulations classify it and then, if not, we must determine how it should be classified under general tax and transfer pricing principles.

Step 1a—Code Sec. 861 Regulations

We start with the Code Sec. 861 regulations. Those regulations (which are set forth in Reg. §§1.861-18 and 1.861-19) govern digital-content and cloud transactions, respectively. Their primary purpose is to provide sourcing rules for these transactions. But they also apply "for purposes of Internal Revenue Code section [] ... 482," so they are potentially relevant to the transfer pricing for data transactions.[5] By controlling the classification of transactions for Code Sec. 482 purposes, the regulations ensure that there will be no inconsistency in characterization for sourcing and transfer pricing purposes.

What is a digital-content transaction? The regulations define "digital content" as a computer program or other

content in digital format that is either protected by copyright law or not protected by copyright law solely due to the passage of time or because the content was dedicated to the public domain.[6] The regulations provide four categories[7]:

(i) A transfer of a copyright right in the digital content;
(ii) A transfer of a copy of the digital content (a copyrighted article);
(iii) The provision of services for the development or modification of the digital content; or
(iv) The provision of know-how relating to development of digital content.

What is a cloud transaction? In contrast to a digital-content transaction, "[a] cloud transaction is a transaction through which a person obtains on-demand network access to computer hardware, digital content ..., or other similar resources."[8] As the regulatory preamble explains, "when a customer streams digital content, there is no transfer of digital content. Instead, the customer receives access to the digital content through the provider's servers. Especially for large file-size content, performing the hosting function in order to allow the customer continuous access places a higher burden on the provider."[9]

How do these regulations categorize transactions? If a transaction fits within the digital-content regulations, it could be classified as either an intangible transaction (if it is a transfer of a copyright right or the provision of know-how) or a service transaction (if it is a provision of services). It might also be a tangible transaction if it is a transfer of a copyrighted article in either physical or electronic form,[10] although the preamble to the regulations explicitly reserved on this point.[11] If the transaction is a cloud transaction, then it will be automatically classified as a service.[12]

What does all this mean for the provision of user data? As the reader might have noticed, these categories exclude a category that is highly relevant to our analysis. Specifically, the digital-content regulations exclude non-copyrightable data from the definition of digital content, despite commentators requesting that the definition include it.[13] Because user data is typically not subject to copyright, this exclusion could mean that user data is not digital content. But one might argue that even if user data cannot be digital content, it could nevertheless fall within the cloud transaction framework as an "other similar resource," as further explained below. That leaves two likely possibilities: the provision of user data might be a cloud transaction under the Code Sec. 861 regulations (and thus a service transaction for Code Sec. 482 purposes), or the regulations might not address the provision of user data at all.

Step 1b—General Tax Principles

If the Code Sec. 861 regulations do not provide a characterization, then the taxpayer must use general characterization and Code Sec. 482 tax principles to determine the character of the transaction.[14]

For data used in AI models, this would begin by determining whether the data is property, rather than a service. For the provision of non-copyrighted, non-patented data to be treated as property, it generally must qualify as know-how or a trade secret. The leading authorities on this are *E.I. DuPont de Nemours & Co.*[15] and Rev. Rul. 64-56. If the data is secret, then it would likely be considered property. Otherwise, it likely would not, although it would depend on the facts.

Until further clarity emerges from IRS guidance or perhaps even caselaw, taxpayers transferring user data across borders must navigate a regulatory framework that was not expressly written with AI in mind.

If the data is property, then we need to determine whether that property is being sold, licensed, or leased. "Under the substantial rights test, a transfer of the exclusive rights to use the trade secret or know-how *and* to prevent any unauthorized disclosure thereof will be deemed a sale of the asset."[16] "[A] transfer of the exclusive right to use a trade secret or know-how for a period less than the estimated useful life of the asset results simply in a license of the asset."[17] Sales or licenses of intangible property are governed by the intangible-property rules in Reg. §1.482-4. The transaction could also be treated as a leasing transaction subject to the tangible-property regulations in Reg. §1.482-3.[18]

Application

Perhaps strangely, the categorization question could very well hinge not only on the nature of the data, but also on the method of delivery.

The cloud regulations arguably resolve the categorization question when the U.S. company accesses the data in the subsidiary's digital environment (*e.g.*, by accessing the

data on the subsidiary's servers, but never downloading a copy). As noted, cloud transactions are always treated as service transactions. The caveat is that the regulations provide that "[a] cloud transaction does not include network access to download digital content for storage and use on a person's computer or other electronic device."[19] Reg. §1.861-19(b). As noted above, data used to train AI models actually changes those models, so that unique feature would need to be studied fully—even merely accessing the data to train the model could be considered a "download" for "use" on the parent's systems.[20] With that caveat noted, if this step resolves in favor of a cloud transaction, then the transaction is a service.

If the subsidiary instead delivers an electronic *copy* of a non copyrightable dataset for local U.S. use, the digital-content regulations likely do not govern the classification of the transaction because those regulations explicitly exclude user data from their remit. Delivering a copy, as noted, would also not be considered a cloud transaction. Therefore, the Code Sec. 861 regulations would not characterize the transaction, and we would categorize the transaction under general tax and transfer pricing principles. Proprietary user datasets qualify as intangible property for Code Sec. 482. The provision of that intangible therefore could be a licensing transaction or a sale transaction, depending on how the transaction is structured. Again, the provision of such datasets could also arguably be a tangible transaction.

Step 2—Pricing the Transaction

Once the taxpayer classifies the transaction, the taxpayer must price it under the Code Sec. 482 regulations. While the Code Sec. 482 regulations provide one arm's length standard for all types of transactions, the regulations provide different analytical frameworks for analysis depending on the transaction type. Assuming the taxpayer is not operating under a cost-sharing arrangement, two likely candidate frameworks are Reg. §1.482-9 (for a service transaction) or Reg. §1.482-4 (for an intangible transaction). A transaction characterized as a tangible property transfer could be priced under Reg. §1.482-3.

At a theoretical level, the economic value of the user data could be priced identically under any of these regulatory frameworks. Like every intercompany transaction, a user-data transaction should be analyzed under general transfer pricing principles with due regard to the parties' functions, assets, risks, and contractual terms, among other relevant factors. Under such general transfer pricing principles, the facts and circumstances of the particular case, rather than the analytical framework for analysis, should determine the appropriate comparables, the "best method," and ultimately the arm's length price (or arm's length range of prices).

Practically, though, there are several reasons why characterization as an intangible transaction could, all else equal, signal possible higher value or invite greater scrutiny:

- *Commensurate-with-income standard and periodic adjustments.* If the transaction is classified as an intangible transaction, the IRS can assert periodic adjustments under the commensurate-with-income standard. This means that, if the user data proves more profitable than anticipated, there is a risk that the IRS will assert that such unexpected value should be taken into account with hindsight. Indeed, in a memo from early 2025, the IRS's Office of Chief Counsel appeared to take the position that the commensurate-with-income standard gives the IRS the right to treat after-the-fact profit information as determinative, even if that after-the-fact profit information was not foreseeable at the time of the transaction.[21] Moreover, non-U.S. tax administrations that follow the 2022 Organisation for Economic Co-operation and Development ("OECD") Transfer Pricing Guidelines (the "OECD Guidelines") may take a similar hindsight approach under the OECD Guidelines guidance on Hard-to-Value-Intangibles ("HTVI").[22] In contrast, if the transaction is classified and respected as a service transaction, it should not be subject to periodic adjustments under the Code Sec. 482 regulations or the OECD Guidelines' HTVI regime.
- *Potential differences in comparables sets*. The comparables sets for pricing services and intangibles typically differ, and these differences could yield different pricing outcomes. For example, when applying the Comparable Profits Method to a service provider, the method typically hinges on comparable *service* companies that do not own significant intangibles, whereas to analyze an intangibles transaction, it would be necessary to locate comparable transactions or companies that *do* employ similar intangibles. These are different universes of comparables that can produce different routine return benchmarks and therefore different residual calculations.
- *DEMPE*. Tax administrations that follow the OECD Guidelines may scrutinize intangible transactions to determine whether the allocation of profit from the intangibles' exploitation is consistent with the parties' contributions to the development, enhancement, maintenance, protection and exploitation ("DEMPE") of the intangibles.[23] This is an important

consideration because some tax administrations may interpret the OECD Guidelines' DEMPE guidance in a manner that diverges from taxpayer or IRS positions grounded in the Code Sec. 482 regulations, giving rise to potential double taxation. In contrast, the OECD Guidelines' guidance on pricing service transactions does not contain a concept analogous to DEMPE and is generally assumed to have less potential for divergence in interpretation with the Code Sec. 482 regulations.

- *Default expectations*. It is sometimes assumed that services are priced at lower values than intangibles. Indeed, as we know, one of the pricing mechanisms for services is the Services Cost Method, which allows taxpayers to charge services out at cost with no markup. To qualify for that method, a service must either be a specified covered service listed in Rev. Proc. 2007-13 or a low-margin covered service with a median comparable markup of 7% or less.[24] Although it is debatable whether most transactions in user data for purposes of training AI could qualify for this method, it does reflect Treasury's understanding that many (but certainly not all) types of services are routine or relatively lower value in nature.

All that being said, characterization as a service transaction will not necessarily be enough to avoid intensive transfer pricing scrutiny. As noted, value does not turn on the analytical framework alone. Moreover, the IRS can scrutinize whether a transaction characterized as services contains a purported embedded intangible element as described in Reg. §1.482-9(m).[25]

One final note about the pricing bears mentioning: If the taxpayer is operating under a cost-sharing arrangement, the classification of the user-data transaction is arguably less important from a transfer pricing perspective. If the group operates a cost-sharing arrangement for the development of AI intangibles, the subsidiary's provision of user data could constitute either a "platform contribution transaction" or an "operating contribution," depending on the proximity of the data to the intangible development activity. In either case, the subsidiary must receive arm's length compensation from the cost-sharing participants reflecting the anticipated value of that contribution.

Practical Observations and Parting Thoughts

The different potential analyses described above might lead to disputes. Taxpayers might in many cases determine that a transaction is most appropriately analyzed as a service transaction for transfer pricing purposes, but the IRS or another tax authority might have an incentive to try to analyze the transaction as an intangible transaction. If the IRS were to do so, though, it might face an obstacle of its own creation: the cloud regulations, which—if applicable—force taxpayers and the IRS down the services path.

One non-tax limitation here is data privacy. For example, if the subsidiary were located in the European Union ("EU"), the General Data Protection Regulation ("GDPR") may prevent the user data from leaving the jurisdiction. In that situation, the U.S. model might legally need to pay a "visit" to the subsidiary's server or system, learn from the data locally, and send only the learning back to the United States. This looks more like a cloud transaction, which is more likely to be priced as a service. If the taxpayer were to characterize the transaction as a service, and the IRS were to successfully recharacterize the transaction as an intangible transfer, then data privacy regulations could create something akin to a "blocked income" problem (not because the income is actually blocked, but because a non-tax statute barring certain activity could impact the pricing analysis).

Given all these considerations, taxpayers should look for ways to reduce the possibility of disputes. The most obvious way is to structure the intercompany agreements in a thoughtful way that increases the likelihood that they will be respected on audit. There might be more proactive ways, such as seeking a private letter ruling on the Code Sec. 861 classification.

Finally, although not a silver bullet, this area might at some point also be ripe for use of a bilateral Advance Pricing Agreement ("APA") to reduce uncertainty. An APA would clearly delineate the nature and form of covered intercompany transactions (*e.g.*, as a license or intercompany service) and provide certainty on the transaction's pricing. And if the APA is bilateral or multilateral, it would provide for consistent characterization and pricing in the jurisdictions of each party to the relevant transactions, thereby preventing double taxation and avoiding the risk that the taxing authorities put the taxpayer in a whipsaw position by characterizing the transaction in different ways. This could be especially important because classification under the Code Sec. 861 regulations presents a U.S.-specific wrinkle.

Until further clarity emerges from IRS guidance or perhaps even caselaw, taxpayers transferring user data across borders must navigate a regulatory framework that was not expressly written with AI in mind. In a world where AI models can be trained in weeks and deployed globally in days, the "new" Code Sec. 861 regulations already are showing their age.

ENDNOTES

[1] Janakiram MSV, *Meta invests $14 Billion in Scale AI to strengthen model training*, Forbes, June 23, 2025, *www.forbes.com/sites/janakirammsv/2025/06/23/meta-invests-14-billion-in-scale-ai-to-strengthen-model-training/*.

[2] Aurel Stenzel, From data to information (and back), Medium, Apr. 8, 2021, *medium.com/frctls/from-data-to-information-and-back-7519b232d489*.

[3] What is model training?, IBM, *www.ibm.com/think/topics/model-training#1580786329*.

[4] Alexander Lindon, *Is synthetic data the future of AI?*, Gartner, June 22, 2022, *www.gartner.com/en/newsroom/press-releases/2022-06-22-is-synthetic-data-the-future-of-ai*.

[5] Reg. §§1.861-18(a)(1) and 1.861-19(a)(1).

[6] Reg. §1.861-18(a)(2)(i).

[7] Reg. §1.861-18(b).

[8] Reg. §1.861-19(b).

[9] *Classification of Digital Content Transactions and Cloud Transactions*, T.D. 10022, 90 FR 2977-01, 2979 (January 14, 2025) (preamble).

[10] Reg. §1.861-18(g)(2) ("The rules of this section shall be applied irrespective of the physical or electronic or other medium used to effectuate a digital content transaction.").

[11] *Classification of Digital Content and Cloud Transactions*, *supra* note 9, at 2986 ("The Treasury Department and IRS have determined that guidance on whether the categories of transactions in §1.861–18 are considered tangible or intangible property for purposes of such Code section s is outside the scope of these regulations.").

[12] This differs from the proposed rules: "Proposed §1.861–19(c)(1) would [have] provide[d] that a cloud transaction is classified ... as either a lease of property or the provision of services, based on all relevant factors. ... The *final* regulations treat all cloud transactions solely as the provision of services." *Id.* at 2982–83 (emphasis added).

[13] *Classification of Digital Content and Cloud Transactions*, *supra* note 9, at 2979 ("Several comments recommended broadening the definition of digital content to encompass content not protected by copyright law that is transferred electronically and is similar to copyrightable content, such as consumer or user data, text files of recipes, government-produced documents, and sets of font and typefaces. ... The final regulations do not broaden the definition of digital content beyond content protectable by copyright law.").

[14] The OECD Transfer Pricing Guidelines provide a comparable framework. Under Chapter I, the "accurate delineation" of a controlled transaction requires identifying the economically relevant characteristics of the arrangement, including the contractual terms, functions performed, assets used, risks assumed, and the economic circumstances of the parties. This functional inquiry closely resembles the analysis required under the Code Sec. 482 regulations, which similarly look to the functions performed, risks assumed, and resources employed by each party. Both frameworks permit—indeed require—looking beyond contractual labels to the economic substance of the transaction.

[15] CtCls, 61-1 USTC ¶9359, 288 F2d 904, 153 CtCls 274.

[16] Postlewaite, Cameron & Kittle-Kamp, Fed. Inc. Tax. Intell. Prop. & Intangible Ass. §4:4 (Sales Versus Licenses of Trade Secrets and Know-How).

[17] *Id.*

[18] *See* Reg. §1.482-2(f); *cf.* Reg. §1.861-18(f)(2) (classifying non-sale transfers of copyrighted articles as lease transactions); Reg. §1.861-18(g)(2) (means of transfer, *e.g.*, physical or electronic, not taken into account in determining characterization). But note that the rules of Reg. §1.861-18 do not apply to non-copyrighted data, so the relevance of these rules is by analogy.

[19] By its terms, this exclusion applies to "digital content," so it arguably would not apply to non-copyrightable user data at all, which would mean the cloud-transaction classification stands without further complication. But if any portion of the dataset is copyrightable, or if the exclusion is read to extend to the "other similar resources" referenced in the cloud-transaction definition, the analysis becomes less clear-cut.

[20] To illustrate by way of an analogy: When a child learns by reading school textbooks, the child does not memorize the books verbatim, but this "training" nevertheless changes and improves the way the child thinks. AI training works similarly: the model ingests data not to store it, but to adjust its internal parameters based on patterns in that data. Whether this training process constitutes "download ... for storage and use" within the meaning of the regulation is debatable.

[21] IRS CCM AM2025-001 (January 15, 2025), *www.irs.gov/pub/lanoa/am-2025-001.pdf*.

[22] *See* OECD Guidelines, ¶6.186 *et. seq.*

[23] *See* OECD Guidelines, ¶6.32 *et. seq.*

[24] Reg. §1.482-9(b)(3).

[25] Furthermore, similar considerations exist for transactions characterized as tangible in nature. While tangible property is also sometimes assumed to be less valuable (or at least have a more certain value) than intangible property, under the special rule in Reg. §1.482-3(f), any "embedded intangible property" in tangible property transfers may need to be taken into account (and if appropriate, analyzed under Reg. §1.482-4). Moreover, while not directly bearing on value, we would note that transactions involving "digital goods" are expressly excluded from the scope of the OECD's "simplified and streamlined approach" (also known as Amount B) for pricing certain "baseline" marketing and distribution transactions, perhaps on the ground that comparables for digital goods may be different than for physical goods. *See* OECD, Pillar One—Amount B (2024), available at *www.oecd.org/content/dam/oecd/en/publications/reports/2024/02/pillar-one-amount-b_41a41e1e/21ea168b-en.pdf*.