

Quantifying the Intangible

A Practitioner Framework for Enterprise Data Valuation

Executive Summary

The valuation problem

Corporate value creation has shifted decisively toward intangible assets, but financial reporting still captures only a fraction of the resources that drive enterprise value. Ocean Tomo [47] reports that intangible assets represented approximately 92% of S&P 500 market value at the end of 2025, compared with 17% in 1975. Data is one of the most strategically important components of this intangible economy, yet also one of the most difficult to evaluate. It is non-rival, reusable, context-dependent, and often produced as a by-product of operational transactions rather than acquired in a clean market exchange [34] [61]. These characteristics make data economically powerful, but they also make it hard to recognize on balance sheets, price in transactions, or govern as a financial asset [8] [12].

This tension raises a central question: how should organizations account for and value data as an asset? Two extreme responses are equally unsatisfactory: ignoring data simply because it fails traditional recognition tests, or capitalizing all internally generated data at fair value despite the measurement difficulties described above. A more defensible approach is a layered valuation architecture that matches the method to the purpose: conservative accounting recognition where existing criteria are met, richer disclosure for material data resources, transaction-based evidence drawn from licensing and M&A, and internal management valuation to guide investment, monetization, and governance decisions [8] [30] [32].

What this whitepaper contributes

This whitepaper provides a practitioner-oriented framework for understanding, measuring, and realizing the economic value of enterprise data assets. It integrates evidence from accounting standards, investor surveys, academic economics, information systems research, purchase price allocations, AI licensing deals, and emerging data-sharing regulation [8] [9] [34] [38].

The central argument is deliberately bounded: data can be a material enterprise asset, but its value is not intrinsic. Data becomes valuable when it is legally usable, technically accessible, of high-quality, analytically exploitable, embedded in workflows, and combined with complementary organizational capabilities [2] [6] [35] [46].

Key Findings

Dimension	Findings
Valuation gap	The macro evidence on intangibles is strong, but it must not be equated mechanically with data value. The most robust claim is that financial reporting underrepresents intangible value drivers, including but not limited to data [8] [47].
Accounting treatment	IAS 38 and U.S. GAAP recognize only narrow categories of internally generated intangibles. The current reform trajectory is disclosure-first rather than broad immediate recognition of internally generated data assets [12] [30] [60].
Economic value	Data is non-rival and reusable, but marginal returns often decline once learning thresholds are reached. Data network effects exist, but they are conditional on freshness, exclusivity, model learning, and complementary assets [25] [34] [61].

Dimension	Findings
Valuation methods	Cost, income, market, real options, and Shapley-based approaches are all useful, but each answers a different valuation question. No single method establishes a universal fair value for data [23] [32] [38].
Transactions	M&A and licensing markets reveal data value indirectly. Data is often absorbed into customer relationships, databases, developed technology, or goodwill rather than reported as a standalone data asset [14] [40] [44].
GenAI	Generative AI has created visible price signals for rights-cleared content and expert-labeled training data, but these prices are contract-specific and should not be generalized into universal per-record valuations [52] [53] [64].
Future trajectory	By 2035, data valuation is likely to operate across three layers: accounting and disclosure, internal management valuation, and transaction-based market pricing [8] [30].

Table 1. Key Findings

Definitions used throughout this whitepaper

Data refers to observations, signals, records, or structured information generated or collected in business processes [34] [61]. Data alone is not necessarily an asset in an accounting or valuation sense [8].

Data asset is an identifiable dataset, database, or data stream that an enterprise controls or has the legal right to use and that can generate economic benefits for the enterprise or another party [29] [32].

Data resource refers to a governed, documented, and purpose-bound data source that may be eligible for internal management valuation, disclosure, or, in limited circumstances, accounting recognition [30] [45].

Data product refers to a governed bundle of data, metadata, access interfaces, documentation, quality controls, and service commitments designed for internal or external users [11] [28].

Data rights refer to legal, contractual, or technical permissions to access, process, share, sell, or use data for specific purposes [19] [20].

Data value refers to the economic contribution of data under a specified purpose, user, method, and rights regime. It is therefore always context-dependent rather than universal [32] [38].

Contents

Executive Summary	2
1. The Valuation Gap — Why Data Is Hard to See	7
1.1. From intangible dominance to data-specific valuation	7
1.2. Accounting standards as structural barriers	7
1.3. Investor demand for better disclosure.....	8
1.4. Standard-setting trajectory.....	8
1.5. Strategic implications	8
2. How Data Creates Value — Lifecycle Dynamics, Returns, and Decay.....	10
2.1. Why “data is the new oil” is analytically misleading	10
2.2. Startup phase: threshold effects and option value.....	10
2.3. Scale-up phase: learning effects and diminishing returns.....	11
2.4. Maturity phase: complementarity, reusability, and architecture.....	11
2.5. Data decay and useful life	11
2.6. Data gravity and architecture	12
3. Valuation Methodologies. What Each Method Can and Cannot Answer	14
3.1. The central valuation problem	14
3.2. Cost approach: replacement or reproduction cost	14
3.3. Income approach and MEEM	14
3.4. Relief-from-royalty	14
3.5. Market approach.....	15
3.6. Real options valuation	15
3.7. Data Shapley and contribution-based valuation.....	16
3.8. Method selection framework	16
4. Data in Transactions.	
How Value Becomes Visible.....	18
4.1. The PPA visibility paradox	18
4.2. Data-driven M&A premiums are conditional.....	18
4.3. Cautionary transactions	18
4.4. Subscription data utilities	19
4.5. GenAI licensing as a new price signal	19
4.6. Expert data and post-training markets.....	19
4.7. Data as collateral and structured finance	20
5. Data Sharing, Regulation, and Federated Value Creation	22
5.1. Data spaces and value without central data ownership	22
5.2. EU Data Act	22
5.3. European Health Data Space	23
5.4. Contribution-based compensation in data ecosystems	23

6. A Decision Framework for Enterprise Data Valuation.....	25
6.1. Step 1: Define the valuation object.....	25
6.2. Step 2: Define the valuation purpose.....	25
6.3. Step 3: Assess rights and constraints.....	25
6.4. Step 4: Assess quality, freshness, and provenance	25
6.5. Step 5: Identify the value mechanism	25
6.6. Step 6: Select and triangulate methods	26
6.7. Step 7: Convert valuation into governance	26
6.8. Enterprise Data Valuation Canvas	26
7. The 2035 Horizon	29
7.1. Three-layer architecture	29
7.2. What data is likely to command a premium.....	29
7.3. Managerial imperatives	29
CONCLUSION: From Data Hype to Data Discipline.....	31
References.....	33

Chapter 1



1. The Valuation Gap — Why Data Is Hard to See

1.1. From intangible dominance to data-specific valuation

The rise of intangible assets is one of the defining features of modern capital markets. Ocean Tomo's 2025 intangible asset market value study reports that approximately 92% of S&P 500 market value was intangible at the end of 2025, continuing a decades-long shift from tangible to intangible value drivers [47]. This shift creates a legitimate concern for investors and boards because many strategic resources are not fully captured by traditional balance-sheet accounting [8] [39].

The analytical bridge from "intangibles" to "data" must, however, be made carefully. Intangibles include brands, patents, software, customer relationships, human capital, organizational routines, network effects, and goodwill, and only some of this value is attributable to data [32] [39]. Data valuation therefore requires a narrower question: which data resources can be identified, controlled or legally used, economically linked to benefits, and valued with a method appropriate to the decision at hand [12] [32]?

Macroeconomic estimates also require caution. Recent work suggests that the value of data is economically material, but the estimated magnitude depends heavily on whether the analyst measures data investment, data rents, unpaid data flows, or productivity effects [9] [48]. Data rents are estimated at approximately 2.2% of U.S. GDP and, when combined with investment in data assets of approximately 0.7% of GDP, yield a total contribution approaching 3% of U.S. GDP [7] [9].

1.2. Accounting standards as structural barriers

Under IAS 38, an intangible asset must be identifiable, controlled by the entity, and expected to generate probable future economic benefits [29]. However, internally developed brands, mastheads, publishing titles, customer lists, and comparable items do not qualify for recognition because the related expenditure cannot be reliably separated from the broader costs of developing the business as a whole [12] [29].

This creates a structural barrier for data. Internally generated data is often intertwined with software, customer relationships, operational processes, and analytic models, making it difficult to isolate as a separable asset [12] [32]. Privacy and data-protection regimes also complicate the meaning of "control" because enterprises may have rights to process or use data for specific operations without owning it in a property-law sense [18] [19].

IFRS and U.S. GAAP differ in their treatment of development expenditure, but both remain conservative relative to the economic role of data. Under IFRS, development expenditure can be capitalized only when six cumulative criteria are met. These include demonstrating technical feasibility, an intention and ability to complete and use or sell the asset, probable future economic benefits, sufficient resources to complete the project, and the ability to measure the related costs reliably [12] [29]. Under U.S. GAAP, internally generated intangible development costs are generally expensed unless specific software guidance applies [12]. This results in many companies reporting non-GAAP earnings that are often times higher compared to GAAP earnings and the question if a company is making profit or a loss is becoming increasingly difficult to answer [24].

The result is not that accounting standards are "wrong" in a simple sense. The result is that financial statements are designed for reliability, comparability, and auditability, while data value is often contextual, probabilistic, and dependent on future uses that cannot be reliably measured at the time of collection [8] [28] [32].

1.3. Investor demand for better disclosure

Investor demand for better information on intangibles is well documented [24] [56]. A CFA Institute survey of investment professionals found that 86% of respondents needed better disclosures regarding intangibles, while only 39% found current intangible disclosures useful for valuation purposes. Furthermore, the survey revealed that 91% of respondents considered better disclosure of internally generated intangibles beneficial, and 87% supported a sequenced approach in which improved disclosure precedes broader recognition [8].

This evidence supports a disclosure-first agenda. Investors do not necessarily need all data resources to be recognized at management-estimated fair value on the balance sheet; they need decision-useful information about data-related investments, governance, scale, quality, risk, monetization, and assumptions [8] [60].

A practical disclosure package for material data resources could include: capitalized and expensed data-related costs, description of major data resources, data provenance, governance and access rights, monetization channels, useful life assumptions, cybersecurity and privacy risks, and key operating metrics such as active data products, API usage, refresh frequency, or data-quality indicators [2] [8] [32].

1.4. Standard-setting trajectory

The IASB has an active project on intangible assets, but it has not yet decided to introduce broad balance-sheet recognition for internally generated data assets [30]. The UK Endorsement Board has also conducted extensive research on intangibles and continues to monitor international developments, but the evidence suggests an improvement in reporting and disclosure rather than immediate wholesale capitalization [60].

China represents the most visible jurisdictional experiment with enterprise data resources. The Ministry of Finance of the People's Republic of China [45] issued interim provisions on the accounting treatment of enterprise data resources that took effect in 2024, allowing qualified data resources to be presented under existing categories such as intangible assets or inventories when recognition criteria are met. This approach is significant because it operationalizes data-resource accounting, but it remains conservative because recognition still depends on existing asset criteria and reliable measurement.

The global direction is therefore best characterized as cautious experimentation. The most likely near-term path is richer disclosure, better disaggregation, more explicit useful-life and governance assumptions, and selective recognition where existing criteria can be met [8] [30] [32].

1.5. Strategic implications

For CFOs, the valuation gap means that data-intensive investments may create economic value without producing a corresponding balance-sheet asset [8] [39]. For boards, it means that data governance is not merely a compliance issue but a capital-allocation and fiduciary issue [8]. For investors, it means that reported book value is often a weak proxy for the value of data-rich enterprises [39] [47]. For acquirers, it means that data due diligence must examine rights, provenance, quality, integration feasibility, and monetization rather than relying only on accounting labels [36] [50].

The key implication is not that every company should put data on the balance sheet. The key implication is that every data-intensive company should know which data resources matter, how they create value, which rights attach to them, how quickly they decay, and which valuation method is appropriate for the decision being made [32] [38].

Chapter 2



2. How Data Creates Value — Lifecycle Dynamics, Returns, and Decay

2.1. Why “data is the new oil” is analytically misleading

The oil metaphor is misleading because oil is rival and consumed in use, while data is non-rival and can be reused across applications without being depleted. Data can also be recombined with other datasets, models, and organizational routines, which gives it potential for increasing returns under the right conditions [34] [61].

Non-rivalry does not mean that more data always creates more value. Varian [61] argues that data often exhibits declining marginal returns, because the first observations may substantially improve learning while later observations of the same type add little incremental predictive power. This is why data valuation must distinguish between volume, variety, freshness, exclusivity, quality, legal usability, and complementarity [25] [28] [61].

2.2. Startup phase: threshold effects and option value

In early-stage firms, data value often arises from threshold effects. A company may need enough observations, users, transactions, or labeled examples before a recommendation system, risk model, routing tool, or automation system becomes useful [25] [61]. At this stage, a dataset may generate limited current revenue while still having substantial option value due to its potential use in future model development, product-market fit, or partnership-based monetization [10].

Waze illustrates the strategic logic of a data network effect, although the acquisition value should not be interpreted as a pure data price. Users generated real-time traffic and route information, which improved the service, attracted more users, and increased the supply of fresh mobility data. Google’s acquisition of Waze was therefore consistent with a data-network-effect thesis, but the transaction also included brand, technology, team, user community, and strategic integration value [66].

The valuation implication is that real options logic is often more suitable than a conventional DCF for early-stage data resources. A DCF may undervalue a dataset that has uncertain but potentially high future use, while a real options approach recognizes the flexibility to deploy, combine, license, or abandon the data later [10] [59].



At pre-seed and seed stage, a DCF is simply not applicable — there are no reliable cash flows to project, and the value of a data asset lies entirely in what it enables, not what it currently generates. What we look for instead is option value: does the dataset give the company a head start in training proprietary models that competitors cannot easily replicate, and are the ownership rights clean and clearly documented? Exclusivity and legal clarity around data provenance are therefore hard diligence criteria for us, not soft signals.”

2.3. Scale-up phase: learning effects and diminishing returns

In the scale-up phase, data can become a growth engine when product usage generates data, data improves model performance, improved performance attracts additional usage, and the loop reinforces itself. This mechanism is the foundation of data network effects and is visible in digital platforms, routing applications, recommendation engines, and advertising systems [25] [61].

The empirical and theoretical literature cautions against treating this mechanism as automatic. Data network effects are strongest when additional data improves a shared model or service for many users, not merely when it personalizes a service for one user [25] [61]. The effect is also stronger when data is fresh, difficult to replicate, legally usable, and paired with superior analytics capability [6] [35].

The valuation implication is that income-based methods become more plausible once data contributes to identifiable cash flows. However, forecasts should model saturation and declining marginal returns rather than assuming that data volume creates linear or exponential value indefinitely [32] [61].

2.4. Maturity phase: complementarity, reusability, and architecture

In mature enterprises, the marginal value of additional homogeneous records often declines. What matters more is whether data is fresh, high-quality, reusable, rights-cleared, semantically interoperable, and embedded in operational or customer workflows [46] [61].

Mature data assets also create value through complementarity. Brynjolfsson and Hitt show that IT value depends heavily on complementary organizational investments such as process redesign, skills, and organizational change [6]. Kohli, Grover [35] emphasize that IT value materializes through actual usage, not through technical availability alone. Mithas, Ramasubbu, Sambamurthy [46] show that IT capability improves firm performance through customer management, process management, and performance management capabilities.

The valuation implication is that a mature data estate should not be valued simply by storage volume or number of records. It should be valued by its contribution to revenue, cost reduction, risk reduction, decision quality, compliance, customer retention, or market access [32] [38].

2.5. Data decay and useful life

The value of data decays at different rates depending on the use case. Mobility, advertising, fraud-detection, and short-term behavioral data typically lose value quickly because current information is more useful than older observations [61]. Credit, clinical, industrial, and longitudinal datasets may retain value for years, but their predictive utility still depends on refresh cycles, regime changes, and model retraining [49] [67].

The practical point is that accounting useful life, technical data retention, and economic value may decay according to different timelines. A credit database may have a 10–15-year useful life for purchase price allocation purposes because it is continuously updated and embedded in workflows, while individual records may lose their predictive value much faster [14] [58]. Valuation models must therefore distinguish between the value of a continuously maintained database and the value of a static historical data extract [32].

Data Type	Typical Valuation Logic	Valuation Implications
Real-time mobility and routing data	Freshness dominates because current conditions change quickly [61].	Useful life can be minutes or hours for operational routing; historical archives may still support planning models.
Advertising and purchase-intent data	Recent behavior often matters more than large historical stocks [61].	Short amortization, flow-based pricing, or operating-expense treatment may be more defensible than long-lived capitalization.
Credit and financial behavior data	Historical patterns matter, but fresh transactions improve predictive power [49] [67].	Continuously refreshed databases can justify longer useful lives than static extracts.
Clinical and EHR data	Longitudinal history can be valuable, but clinical relevance depends on context, protocols, and population changes [51].	Useful life should be tied to use case, regulatory basis, and model updating.
Industrial IoT and equipment data	Value depends on asset context, sensor quality, failure modes, and integration with maintenance workflows [42].	Real options and process-value approaches are often more appropriate than per-record pricing.

Table 2. Valuation Logic and Implications by Data Type

2.6. Data gravity and architecture

Data gravity describes the tendency for applications, analytics, and services to move toward large data concentrations because moving large volumes of data can be costly, slow, risky, or constrained by regulation [13] [41]. The concept is especially relevant for cloud architecture, M&A integration, and federated data-sharing ecosystems [13].

For valuation, data gravity matters because it can create both value and lock-in. A deeply embedded data estate can increase switching costs and strengthen competitive advantage, but it can also increase migration costs, cloud egress costs, integration complexity, and regulatory risk [13] [68].

Chapter 3



3. Valuation Methodologies. What Each Method Can and Cannot Answer

3.1. The central valuation problem

Data rarely produces value alone. It produces value with software, models, domain expertise, processes, brand, distribution, customer relationships, and organizational adoption [6] [35]. This creates the central valuation problem: isolating the economic contribution of data from the contribution of the surrounding system.

Because of this isolation problem, data valuation should always begin with the valuation purpose. A data asset may need to be valued for accounting recognition, purchase price allocation, internal investment decisions, data-sharing compensation, litigation, insurance, collateralization, or licensing [32] [38]. Each purpose implies different standards of evidence, different assumptions, and different acceptable methods [32].

3.2. Cost approach: replacement or reproduction cost

The cost approach estimates what it would cost to reproduce or replace a data asset, including collection, acquisition, cleaning, labeling, structuring, storage, documentation, and quality assurance [32] [38]. It is most useful when future cash flows are difficult to isolate, when the data is early-stage, or when the objective is to establish a conservative floor value [32].

The limitation is that cost does not equal value. A dataset that cost little to collect may create large economic value if it is unique and legally usable, while a costly dataset may be worthless if it is obsolete, duplicative, non-compliant, or unused [32] [38].

3.3. Income approach and MEEM

The income approach values an asset by estimating the present value of expected future economic benefits attributable to that asset [32]. The multi-period excess earnings method (MEEM) is commonly used when an intangible asset is the primary driver of an identifiable revenue stream. The method isolates the earnings attributable to that asset by deducting contributory asset charges for the other assets required to generate the cash flows [32] [36].

MEEM can be appropriate for credit databases, subscription data products, or data-rich platforms where revenue can be linked to a specific database or data service [14] [58]. Its limitation is that contributory asset charges are judgment-sensitive, and small changes in assumptions can materially change the residual earnings attributed to the data asset [32].

3.4. Relief-from-royalty

The relief-from-royalty method estimates the value of an asset by asking what royalty the company would have paid to license an equivalent asset from a third party [32]. The method is most appropriate when the data is identifiable and separable, can be licensed independently, and reliable evidence from comparable licensing transactions is available [32] [50].

This method is useful for content databases, market data, reference data, and data products with external licensing analogues (e.g., Bloomberg, n. d. [4]). Its limitation is that there is no deep, transparent, standardized royalty-rate market for most enterprise datasets, so the valuer must often use imperfect proxies [32] [50].

3.5. Market approach

The market approach values data by reference to comparable transactions or observable market prices [32]. It is attractive because market evidence can discipline management assumptions, but it is difficult because data transactions are often private, bundled, and contract-specific [32] [48].

AI licensing deals have improved price visibility for some rights-cleared content and expert-labeled data, but those deals should not be generalized into universal per-record values because they bundle access, rights, freshness, exclusivity, brand, legal certainty, and strategic partnership effects [52] [64].



Running a data marketplace taught us that the existence of a marketplace does not create a price. Even with standardized contracts and a functioning platform, most transactions were negotiated bundles — access, usage rights, freshness guarantees, quality assurance, and support — rather than clean per-record sales. Genuine comparables were rare because exclusivity, purpose limitation, and provider trust varied with every deal. The market approach worked best as a plausibility check on other methods, not as a standalone price source. The lesson for valuation is that observable marketplace prices are contextual evidence, not universal benchmarks.”

Prof. Chris Langdon,
Professor at Drucker Customer Lab

3.6. Real options valuation

Real options valuation is appropriate when current cash flows are limited but future uses are uncertain and potentially valuable [10] [59]. Data can create options to build new models, enter markets, support automation, partner in data spaces, or license insights later [10] [28] [27].

The limitation is that real options valuation requires assumptions about volatility, time horizon, exercise costs, and future opportunity values that are often difficult to observe for non-traded data assets [10] [32]. For this reason, real options analysis is often more useful as a strategic decision tool than as a precise external reporting number [10].

3.7. Data Shapley and contribution-based valuation

Data Shapley applies the Shapley value from cooperative game theory to estimate the marginal contribution of individual data points or data providers to model performance [23]. It is particularly relevant in collaborative machine-learning settings, federated data spaces, and data consortiums where multiple parties contribute data to a shared outcome [1] [23] [27].

Exact Shapley computation is computationally expensive because it requires evaluating many possible subsets of data, so practical approaches use approximation methods such as Monte Carlo sampling, KNN-Shapley, Beta Shapley, or distributional variants [23] [33] [37].

Data Shapley is powerful for fairness, contribution allocation, and model-quality diagnostics, but it does not establish a universal market price for data. The value changes with the model, performance metric, data distribution, and use case [23] [37].

3.8. Method selection framework

Valuation question	Preferred method	Rationale
What did it cost to recreate the dataset?	Cost approach	Establishes a conservative floor where future benefits are uncertain.
What cash flows are attributable to the data asset?	Income approach / MEEM	Useful where data drives an identifiable revenue stream.
What would licensing equivalent data cost?	Relief-from-royalty	Useful where licensing analogues exist.
What do comparable transactions imply?	Market approach	Useful where transparent and comparable transactions exist.
What future strategic flexibility does data create?	Real-options valuation	Useful for early-stage and uncertain future use cases.
How should value be allocated among contributors?	Data Shapley	Useful in collaborative ML and data-space settings.

Table 3. Key Questions for Selecting an Appropriate Valuation Method [10] [23] [32] [50].

In practice, robust valuation requires triangulation. The analysis should justify the selected method, consider relevant alternatives, identify the key assumptions, and assess how changes in those assumptions affect the result [32].

Chapter 4



4. Data in Transactions. How Value Becomes Visible

4.1. The PPA visibility paradox

M&A is one of the few contexts in which data value can become visible, but public accounting rarely labels it simply as “data.” In purchase price allocations, data-related value is often embedded in databases, credit files, customer relationships, developed technology, user base, or goodwill [14] [44] [58].

Microsoft’s acquisition of LinkedIn illustrates the point. Microsoft framed the acquisition strategically as the combination of the Microsoft Graph and the LinkedIn Graph, but the purchase price allocation reported customer-related intangible assets and goodwill rather than a standalone “data asset” line item. The data logic was strategically central, but the accounting label did not isolate data as a separate asset [44].

Mastercard’s acquisition of Fincity similarly shows that data access can be a core deal rationale without producing a separate data line item in public PPA disclosure. Mastercard described Fincity as strengthening its open-banking platform, but the accounting disclosures grouped acquired intangible assets into developed technology and customer relationships [40].

The implication is that practitioners should not equate absence of a data line item with absence of data value. They should instead examine how data value is embedded in customer relationships, software, technology, network effects, and goodwill [32] [36].

4.2. Data-driven M&A premiums are conditional

The evidence does not support a universal rule that more data equals a higher acquisition multiple. Data creates deal value when it is monetizable, exclusive or difficult to replicate, legally usable, strategically complementary to the acquirer, and operationally integrable [36].

Academic research on data-driven M&A indicates that a target’s data orientation may influence acquisition activity. However, positive announcement returns and acquisition premiums are not automatic and depend on the circumstances of the transaction. Buyers with established AI capabilities may realize greater value from data-rich targets where the acquired data complements their existing resources and operating capabilities. This does not imply that the market consistently reflects the value of data through a distinct upfront acquisition premium [54].

4.3. Cautionary transactions

Data-related deal logic can fail when strategic complementarity is weak. eBay’s acquisition of Skype illustrates how user-network and communication assets may fail to create expected synergies when they do not fit the acquirer’s core transaction workflows [16]. Microsoft’s aQuantive impairment illustrates the risk that audience data, ad technology, and online marketing capabilities may underperform when market position and integration capability are insufficient [43].

Verizon’s Oath impairment after AOL and Yahoo acquisitions illustrates how data-driven advertising theses can be undermined by integration difficulty, platform dependency, privacy constraints, and competitive pressure [63].

These cases do not prove that data lacks value. They prove that data value is conditional on strategic fit, governance, rights, integration, and execution.

4.4. Subscription data utilities

Some firms make data value visible through recurring subscription or transaction pricing. Bloomberg, Experian, Verisk, Dun & Bradstreet, S&P Global, LSEG, and LiveRamp monetize proprietary data through access, analytics, workflow integration, and decision support [4] [15] [21] [62].

These companies demonstrate that durable data value is rarely just a file sale. It is typically a bundled proposition combining data coverage, freshness, identifiers, quality assurance, analytics, workflow integration, service support, and trust. The valuation lesson is that the most defensible data businesses convert data into recurring decision infrastructure. High retention, recurring revenue, embedded workflow usage, and hard-to-replicate data sourcing are stronger indicators of data value than raw data volume [15] [62].

4.5. GenAI licensing as a new price signal

Generative AI has created visible demand for rights-cleared, high-quality, fresh, and domain-specific data. Reuters reported that OpenAI signed a content agreement with News Corp, while The Wall Street Journal reported that the agreement was valued at more than \$250 million over five years [52] [64]. Press reporting also indicated that Reddit's data licensing agreement with Google was approximately \$60 million per year, while the financial terms of OpenAI's later Reddit partnership were not publicly disclosed [57]. An involuntary price signal emerged from Anthropic's \$1.5 billion copyright settlement with authors over the use of uncensored training data [5].

These deals should be interpreted as contract-specific bundles rather than clean per-record prices. They combine rights to access current and archived content, legal certainty, attribution, retrieval or display rights, strategic partnership effects, and model-improvement value [52] [64].

Taken together, these transactions and settlements suggest that generative AI is accelerating price discovery for legally usable data. They do not, however, establish a generalizable market price for enterprise data, as the observed amounts reflect highly specific rights, risk allocations, and strategic contexts [5] [32] [52].

4.6. Expert data and post-training markets

The market for expert labeling and post-training data is becoming increasingly important. Reuters reported that Turing, an AI start-up, tripled revenue to approximately \$300 million as AI labs increased demand for expert human data and evaluation services [53]. This suggests that the scarcity premium in AI is shifting toward data that is specialized, verified, rights-cleared, expert-generated, or useful for evaluation and alignment [53] [65].

Synthetic data does not eliminate the value of real data. Research on model collapse shows that recursive training on generated data can degrade model quality by eroding rare but important features of the original distribution [65]. Synthetic data is valuable as augmentation, privacy-preserving simulation, and low-resource support, but it increases rather than eliminates the importance of trusted ground-truth data [55] [65].

4.7. Data as collateral and structured finance

Data can support financing when it is embedded in cash-generating assets such as loyalty programs, subscriptions, receivables, or recurring customer relationships. Airline loyalty-program financings during the COVID-19 crisis demonstrated how customer data, loyalty economics, and partner cash flows can support secured financing structures [\[3\]](#).

These examples should not be read as pure “data-backed lending”. The collateral package typically includes intellectual property, contracts, cash flows, customer relationships, and operating systems, with data as an important enabling component [\[3\]](#). Pure data collateralization remains a frontier area because lenders need reliable rights, liquidation pathways, valuation evidence, and enforceable security structures [\[32\]](#) [\[36\]](#).



Chapter 5



5. Data Sharing, Regulation, and Federated Value Creation

5.1. Data spaces and value without central data ownership

European data spaces are designed to enable data sharing without requiring central ownership of all data. They rely on federated governance, standardized contracts, identity, access control, usage policies, and interoperability mechanisms [17] [19] [27].

In federated contexts, value is often created through access, reciprocity, compliance, ecosystem coordination, and process optimization rather than direct sale of datasets [26] [27]. This makes traditional market pricing difficult but increases the relevance of contribution-based, contractual, and operational valuation methods [1] [23].

5.2. EU Data Act

The EU Data Act sets out rules on access to and use of the data generated by connected products and their related services. Its core framework has applied since 12 September 2025, though individual obligations take effect later: the requirement to design products so that data is “accessible by default” applies only to products placed on the market after 12 September 2026, and the phase-out of switching charges for cloud services takes effect from 12 January 2027. The Data Act does not create a blanket duty to share data: it gives users the right to access the data generated by their device free of charge and to have it shared with a third party of their choosing, while personal data and – to a certain extent – trade secrets remain protected. Where a company is legally obliged to share data with a party other than the user of connected products or related services, either under the Data Act or any other law, data holders can charge compensation for the shared data. However, that data must be made available on fair, reasonable, non-discriminatory and transparent terms.



WAGOI would not exist without the Data Act: before it applied, access to in-vehicle data depended entirely on manufacturer discretion, with no enforceable right to request data and no benchmark to challenge a price. FRAND and the cost-recovery ceiling for SMEs matter less for the price level than for its predictability; a knowable cost floor is what makes a cross-brand data business financeable. The harder constraint in practice is the trade-secret carve-out, which lets data holders condition rather than refuse access; where the Act clearly succeeds is in moving value away from gatekeeping raw signals and towards the layer that normalises data, manages user consent and makes it usable”

Quote by Mario Schraepen,
CEO WAGOI

The Data Act changes valuation because it limits how this mandated data access can be priced, thereby constraining the ability to charge “monopoly-style” prices for data that others are legally entitled to receive. Users of connected product and related services are entitled to data access free of charge. In all other cases of mandatory data sharing, compensation must be reasonable and non-discriminatory and, in the ordinary commercial case, may include a reasonable margin, subject to the Data Act’s FRAND and non-discrimination requirements. Only where the recipient is an SME with no larger affiliated group, or a not-for-profit research organisation, is compensation capped at the costs directly related to making the data available — with no margin. Importantly, this compensation is not a price for the data as an asset, and the Data Act grants data holders no new right to exploit the data. For management, this means that, for data covered by these obligations, pricing shifts toward transparent FRAND-models rather than open-market pricing, while purely voluntary data deals and data that a company genuinely infers or derives through its own proprietary algorithms remain outside these constraints and can still be commercialised freely.

5.3. European Health Data Space

The European Health Data Space (EHDS) establishes a framework for primary and secondary use of electronic health data across the EU. For secondary use, health data access bodies are expected to mediate access for research, innovation, policy-making, and regulatory purposes, subject to safeguards and approval procedures [20].

The EHDS illustrates a broader principle: when data has high public-interest value, regulation may prioritize controlled access, safeguards, and public-benefit reuse, while safeguarding rather than eliminating private intellectual-property and trade-secret rights. This can reduce some private monetization opportunities. For instance, by prohibiting advertising, marketing, and detrimental insurance uses, and by capping fees at cost, while increasing the societal and innovation value of data resources [20].

5.4. Contribution-based compensation in data ecosystems

Collaborative data ecosystems require fair rules for allocating costs, benefits, and rights to increase market efficiency [28] [27]. Data Shapley and related cooperative-game-theoretic tools can help estimate marginal contributions to model performance, but they must be complemented by legal agreements, governance rules, and operational metrics [1] [23] [27].

A practical compensation framework should distinguish between contribution to model performance, contribution to operational value, data preparation costs, compliance burden, quality assurance, and risk exposure [19] [32]. Mathematical fairness alone is insufficient because ecosystem participants also require enforceable contracts, trust, auditability, and dispute resolution [1] [19] [27].

Chapter 6



6. A Decision Framework for Enterprise Data Valuation

6.1. Step 1: Define the valuation object

The first step is to specify whether the valuation concerns raw data, a curated dataset, a continuously refreshed database, a data product, access rights, a model trained on data, or a broader data-enabled business capability [32] [38]. Without this definition, valuation results will be ambiguous and potentially misleading [32].

6.2. Step 2: Define the valuation purpose

The second step is to define the purpose: “including but not limited to acquisitions, mergers and sales of businesses, taxation, litigation, insolvency proceedings, and financial reporting” [32], potentially also collateralization and ecosystem compensation [26] [27]. A value estimate that is appropriate for internal portfolio prioritization may not be appropriate for audited financial statements or secured lending [32].

6.3. Step 3: Assess rights and constraints

The third step is to assess rights. Data has value only if the organization has legal and practical ability to use it for the relevant purpose [18] [19]. The rights review should cover consent, contracts, purpose limitation, trade secrets, IP restrictions, cross-border transfer, sector-specific rules, retention obligations, and data-subject rights [18] [50].

6.4. Step 4: Assess quality, freshness, and provenance

The fourth step is to assess data quality. Provenance, completeness, accuracy, timeliness, lineage, metadata, bias, representativeness, and refresh frequency materially affect data value [22] [31]. Poor quality can destroy value through wrong decisions, model degradation, compliance failures, or user mistrust [2] [28] [65].

6.5. Step 5: Identify the value mechanism

The fifth step is to identify how the data creates value. Common mechanisms include revenue generation, cost reduction, risk reduction, compliance enablement, customer retention, product personalization, operational automation, ecosystem access, AI training, and strategic optionality [38] [46]. For external purposes, the most promising target market must be identified [50].

6.6. Step 6: Select and triangulate methods

The sixth step is to select valuation methods that match the object and purpose. Replacement cost may establish a floor, income methods may estimate cash-flow contribution, relief-from-royalty may use licensing analogues, market methods may benchmark transactions, real options analysis may capture flexibility, and Data Shapley may allocate collaborative contribution [10] [23] [32] [50].

6.7. Step 7: Convert valuation into governance

The final step is to connect valuation to governance. High-value data resources should have accountable owners, documented rights, lifecycle controls, quality metrics, access policies, security classification, retention rules, and periodic impairment or relevance reviews [27] [31] [32]. Valuation without governance creates false precision; governance without valuation can weaken accountability [8] [38].

6.8. Enterprise Data Valuation Canvas

Question	Evidence	Output
What exactly is being valued?	Dataset, database, data product, access right, model, or capability.	Defined valuation object.
Who can use it and for what?	Contracts, privacy basis, licenses, access controls, purpose limitations.	Rights and constraints map.
How does it create value?	Revenue, cost, risk, compliance, AI, optionality, ecosystem participation.	Value mechanism statement.
How durable is the value?	Freshness, quality, decay, refresh cost, useful life	Useful-life and decay assumptions.
Which method fits?	Cost, income, market, RfR, real options, Shapley.	Valuation method and sensitivity range.
How should it be governed?	Ownership, lineage, controls, monitoring, security, audit.	Data asset governance plan.

Table 4. Guiding Questions for Enterprise Data Valuation



Figure 1. Enterprise Data Valuation Framework

Chapter 7



7. The 2035 Horizon

7.1. Three-layer architecture

The most plausible trajectory is a three-layer architecture rather than a single universal data-asset balance sheet. The first layer is accounting and disclosure, where recognition remains conservative but note disclosures and disaggregation improve [8] [30]. The second layer is internal management valuation, where firms use data asset registers, product P&Ls, investment cases, and governance metrics to allocate capital and accountability [27] [31] [38]. The third layer is transaction-based pricing, where AI licensing, data marketplaces, M&A, data-sharing agreements, and collateral structures generate market evidence [19] [32] [52].

This is a scenario, not a prediction. Standard-setters may move more slowly, regulators may constrain monetization, synthetic data may alter substitution dynamics, and litigation may reshape the economics of training data [30] [65].

7.2. What data is likely to command a premium

The most valuable data in 2035 is unlikely to be generic, stale, or weakly governed. It is more likely to be fresh, rights-cleared, domain-specific, difficult to replicate, embedded in workflows, and accompanied by strong provenance and metadata [32] [53] [65].

The premium categories are likely to include: real-time operational data, regulated-domain data with lawful access, expert-generated or expert-validated data, rare-event data, longitudinal ground-truth data, high-quality evaluation datasets, and data products embedded in critical decision workflows [53] [62] [65].

The value of generic data is more likely to decrease. Synthetic data, web-scale pretraining, open data, and regulatory access rights can reduce scarcity rents for datasets that are easy to approximate or substitute [19] [65].

7.3. Managerial imperatives

Executives should treat data valuation as a management discipline rather than a one-off finance exercise. The practical imperatives are to inventory material data resources, clarify rights, measure quality and freshness, link data to business outcomes, select appropriate valuation methods, disclose material assumptions where relevant, and assign board-level accountability for high-value data assets [8] [31] [32].

The organizations that benefit most will not be those that claim the largest data balance-sheet number. They will be those that can prove that their data is usable, governed, legally robust, analytically productive, and strategically embedded [6] [8] [35].

CONCLUSION



CONCLUSION:

From Data Hype to Data Discipline

Data valuation is moving from metaphor to method. The statement that “data is valuable” is no longer enough. Enterprises must specify which data, under which rights, for which use case, with which quality, over which useful life, and by which valuation method [32] [38].

The discipline is still immature. Accounting recognition remains narrow, transaction evidence remains sparse, and many internal valuations still confuse data volume with value [8] [30]. But the direction is clear: data assets are becoming more visible through investor demand, AI licensing contracts, purchase price allocations, data-sharing regulation, and internal governance systems [8] [19] [52].

The central managerial lesson is that data value is not inherent. It is produced by the interaction of data, rights, quality, models, workflows, governance, and complementary organizational capabilities [6] [35] [46]. The companies that master this interaction will value acquisitions more accurately, allocate capital more effectively, govern risk more rigorously, and convert data into defensible strategic advantage [8] [32].



Data value is created through productization, not by accumulation. Our experience with data analytics and applied AI, from early enterprise data marketplaces to today’s data factories, shows that raw data creates value only when made usable for a defined customer and use case. This requires clear rights, harmonized semantics, measurable quality, and enforceable governance; marketplaces also taught us that value lies in the complete product, not a simple per-record price. Bottom line: leadership needs to manage data like any other product, with defined users, ownership, quality expectations, and lifecycle accountability.”

Quote by Chris S. Langdon, PhD,
Director and Research Associate Professor Drucker Customer Lab

REFERENCES

An abstract geometric pattern in shades of blue and teal, featuring various shapes like triangles, rectangles, and circles, some of which are interconnected by lines, creating a complex, network-like structure. The pattern is more prominent in the lower half of the page.

References

- [1] AGARWAL, A., M. DAHLEH, and T. SARKAR, 2019. A Marketplace for Data. In: A. KARLIN, N. IMMORLICA, and R. JOHARI, eds. *Proceedings of the 2019 ACM Conference on Economics and Computation*. New York, NY, USA: ACM, pp. 701-726.
- [2] ALTENDEITERING, M., T.M. GUGGENBERGER, and F. MÖLLER, 2024. A design theory for data quality tools in data ecosystems: Findings from three industry cases [online]. *Data & Knowledge Engineering*, 153, 102333. Available from: [10.1016/j.datak.2024.102333](https://doi.org/10.1016/j.datak.2024.102333)
- [3] AMERICAN AIRLINES, 2021. *2021 Annual report on form 10-K*.
- [4] BLOOMBERG, n. d. Bloomberg Professional Services [online]. *Enterprise data and terminal product information* [viewed 8 July 2026]. Available from: <https://professional.bloomberg.com/products>
- [5] BRITTAİN, B., 2026. *US judge approves anthropics \$1.5 billion settlement copyright lawsuit* [online]. 20 July 2026, 12:00 [viewed 28 July 2026]. Available from: <https://www.reuters.com/world/us-judge-approves-anthropics-15-billion-settlement-copyright-lawsuit-2026-07-20/>
- [6] BRYNJOLFSSON, E. and L.M. HITT, 2000. Beyond Computation: Information Technology, Organizational Transformation and Business Performance [online]. *Journal of Economic Perspectives*, 14(4), 23-48. Available from: [10.1257/jep.14.4.23](https://doi.org/10.1257/jep.14.4.23)
- [7] CALDERÓN, J., B. SANTIAGO, and D.G. RASSIER, 2022. *Valuing Stocks and Flows of Data Assets for the U.S. Business Sector*. Presentation to the Bureau of Economic Analysis (BEA) Advisory Committee Meeting [viewed 23 July 2026]. Available from: <https://www.bea.gov/sites/default/files/2022-05/BEA-ACM-Data-Assets-Presentation-05132022.pdf>
- [8] CFA INSTITUTE, 2025. *Investor Perspectives: Intangible Assets. Before recognition, improved disclosures and disaggregation are needed*.
- [9] CIURIK, D., 2025. *Enterprise Value and the Value of Data*. CIGI Paper No. 327.
- [10] COYLE, D., S. DIEPEVEEN, J. WDOWIN, L. KAY, and J. TENNISON, 2020. *The Value of Data. Summary Report*.
- [11] DEHGhani, Z., 2022. *Data Mesh. Delivering Data-Driven Value at Scale*. 1st ed. Barcelona: O'Reilly Media.
- [12] DELOITTE, 2026. *Deloitte's Roadmap: Goodwill and Intangible Assets* [online] [viewed 8 July 2026]. Available from: <https://dart.deloitte.com/USDART/home/codification/assets/asc350-20/goodwill>
- [13] DIGITAL REALTY, 2023. *Data Gravity Index (DGx)™ 2.0*.
- [14] DUN & BRADSTREET, 2022. *2021 Annual report on form 10-K*.
- [15] DUN & BRADSTREET, 2025. *2024 Annual report on form 10-K*.
- [16] EBAY, 2008. *2007 Annual report on form 10-K*.
- [17] EUROPEAN COMMISSION, 2020. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A European strategy for data*.
- [18] EUROPEAN UNION, 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*.

- [19] EUROPEAN UNION, 2023. *Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*.
- [20] EUROPEAN UNION, 2025. *Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847*.
- [21] EXPERIAN, 2025. *Experian Annual Report 2025*.
- [22] GARTNER, 2024. *Data and analytics maturity score for CDAOs*.
- [23] GHORBANI, A. and J. ZOU, 2019. Data Shapley: Equitable Valuation of Data for Machine Learning. *Proceedings of the 36th International Conference on Machine Learning, PMLR*. Long Beach, California, pp. 2242-2251.
- [24] GOVINDARAJAN, V., A. SRIVASTAVA, and R. ZHAO, 2021. *Mind the GAAP* [online]. 4 May 2021, 12:00 [viewed 28 July 2026]. Available from: <https://hbr.org/2021/05/mind-the-gaap>
- [25] GREGORY, R.W., O. HENFRIDSSON, E. KAGANER, and H. KYRIAKOU, 2021. The Role of Artificial Intelligence and Data Network Effects for Creating User Value [online]. *Academy of Management Review*, 46(3), 534-551. Available from: 10.5465/amr.2019.0178
- [26] GUGGENBERGER, T.M. and J. BEER, 2026. Developing Collaborative Use Cases in Data Spaces: A Joint-Venture Perspective on Multi-Party Data Collaboration Governance. *Thirty-second Americas Conference on Information Systems*.
- [27] GUGGENBERGER, T.M., C. SCHLUETER LANGDON, and B. OTTO, 2025. Data spaces as meta-organisations [online]. *European Journal of Information Systems*, 34(5), 822-842. Available from: 10.1080/0960085x.2025.2451250
- [28] GUGGENBERGER, T.M., M. ALTENDEITERING, and C. SCHLUETER LANGDON, 2024. Quality Scoring of Data Products: Coping with Information Asymmetry in Data Markets. *Proceedings of the 57th Hawaii International Conference on System Sciences*.
- [29] IASB. IAS 38:2026, *Intangible Assets*: International Accounting Standards Board.
- [30] IFRS, 2026. *Intangible assets project work plan* [online]. 19 May 2026, 12:00 [viewed 10 July 2026]. Available from: <https://www.ifrs.org/projects/work-plan/intangible-assets/>
- [31] ISO. ISO/IEC 5259-1:2024:2024, *Artificial intelligence — Data quality for analytics and machine learning (ML)*: International Organization for Standardization.
- [32] IVSC, 2025. *International Valuation Standards*.
- [33] JIA, R., D. DAO, B. WANG, F.A. HUBIS, N.M. GUREL, B. LI, C. ZHANG, C. SPANOS, and D. SONG, 2019. Efficient task-specific data valuation for nearest neighbor algorithms [online]. *Proceedings of the VLDB Endowment*, 12(11), 1610-1623. Available from: 10.14778/3342263.3342637
- [34] JONES, C.I. and C. TONETTI, 2020. Nonrivalry and the Economics of Data [online]. *American Economic Review*, 110(9), 2819-2858. Available from: 10.1257/aer.20191330
- [35] KOHLI, R. and V. GROVER, 2008. Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times [online]. *Journal of the Association for Information Systems*, 9(1), 23-39. Available from: 10.17705/1jais.00147
- [36] KPMG, 2024. *Fair value measurement. Handbook US GAAP and IFRS® Accounting Standards*.
- [37] KWON, Y. and J. ZOU, 2022. Beta Shapley: a Unified and Noise-reduced Data Valuation Framework for Machine Learning. *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 8780-8802.

- [38] LANEY, D.B., 2018. Infonomics. *How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*: Routledge.
- [39] LEV, B. and F. GU, 2016. *The End of Accounting and the Path Forward for Investors and Managers*: Wiley.
- [40] MASTERCARD, 2021. *Form 10-K: Annual report 2020*.
- [41] MCCRORY, 2010. *Data Gravity – in the Clouds* [online]. 12 July 2010, 12:00 [viewed 9 July 2026]. Available from: <https://datagravitas.com/2010/12/07/data-gravity-in-the-clouds/>
- [42] MCKINSEY, 2025. *The Internet of Things: Mapping the value beyond the hype*.
- [43] MICROSOFT, 2012. *Microsoft Announces Non-Cash Accounting Charge* [online]. 20 February 2012, 12:00 [viewed 9 July 2026]. Available from: <https://news.microsoft.com/source/2012/07/02/microsoft-announces-non-cash-accounting-charge/>
- [44] MICROSOFT, 2017. *Annual Report 2017* [online]. 16.10.217 [viewed 9 July 2026]. Available from: <https://www.microsoft.com/investor/reports/ar17/>
- [45] MINISTRY OF FINANCE OF THE PEOPLE'S REPUBLIC OF CHINA, 2023. *Interim Provisions on Accounting Treatment of Enterprise Data Resources*.
- [46] MITHAS, S., N. RAMASUBBU, and V. SAMBAMURTHY, 2011. How Information Management Capability Influences Firm Performance¹ [online]. *MIS Quarterly*, 35(1), 237-256. Available from: 10.2307/23043496
- [47] OCEAN TOMO, 2026. *Study of Intangible Asset Market Value* [online].
- [48] OECD, 2019. *Enhancing Access to and Sharing of Data*: OECD Publishing.
- [49] OWENS, J.V. and L. WILHELM, 2017. *Alternative Data Transforming SME Finance* [viewed 8 July 2026]. Available from: <http://documents.worldbank.org/curated/en/701331497329509915>
- [50] PWC, 2022. *Fair value measurements*.
- [51] RAJKOMAR, A., E. OREN, K. CHEN, A.M. DAI, N. HAJAJ, M. HARDT, P.J. LIU, X. LIU, J. MARCUS, M. SUN, P. SUNDBERG, H. YEE, K. ZHANG, Y. ZHANG, G. FLORES, G.E. DUGGAN, J. IRVINE, Q. LE, K. LITSCH, A. MOSSIN, J. TANSUWAN, de WANG, J. WEXLER, J. WILSON, D. LUDWIG, S.L. VOLCHENBOUM, K. CHOU, M. PEARSON, S. MADABUSHI, N.H. SHAH, A.J. BUTTE, M.D. HOWELL, C. CUI, G.S. CORRADO, and J. DEAN, 2018. Scalable and accurate deep learning with electronic health records [online]. *NPJ digital medicine*, 1, 18. Available from: 10.1038/s41746-018-0029-1
- [52] REUTERS, 2024. *Sam Altman's OpenAI signs content agreement with News Corp* [online]. 22 May 2024, 12:00 [viewed 9 July 2026]. Available from: <https://www.reuters.com/technology/sam-altmans-openai-signs-content-agreement-with-news-corp-2024-05-22/>
- [53] REUTERS, 2025. *AI data startup Turing triples revenue to \$300 million* [online]. 28 January 2025, 12:00 [viewed 9 July 2026]. Available from: <https://www.reuters.com/technology/artificial-intelligence/ai-data-startup-turing-triples-revenue-300-million-2025-01-28/>
- [54] SCHOCH, D.S., 2020. *Mergers and Acquisitions in the Data Economy* [online]. SSRN Electronic Journal. Available from: 10.2139/ssrn.3686247
- [55] SHUMAILOV, I., Z. SHUMAYLOV, Y. ZHAO, N. PAPERNOT, R. ANDERSON, and Y. GAL, 2024. *AI models collapse when trained on recursively generated data* [online]. *Nature*, 631(8022), 755-759. Available from: 10.1038/s41586-024-07566-y
- [56] SRIVASTAVA, A., 2025. *Financial Reporting for the Knowledge Economy* [online]. SSRN Electronic Journal. Available from: 10.2139/ssrn.5279229
- [57] TIME, 2024. *Steve Huffman* [online]. 5 September 2024, 12:00 [viewed 9 July 2026]. Available from: <https://time.com/collections/time100-ai-2024/7012845/steve-huffman/>

- [58] TRANSUNION, 2022. *2021 Annual Report*.
- [59] TRIGEORGIS, L., 1996. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. Cambridge: The MIT Press.
- [60] UK ENDORSEMENT BOARD, 2026. *Intangible assets project* [online]. 21 May 2026, 12:00 [viewed 9 July 2026]. Available from: <https://www.endorsement-board.uk/projects/intangibles/>
- [61] VARIAN, H., 2019. Artificial Intelligence, Economics, and Industrial Organization. In: A. AGRAWAL, J. GANS, and A. GOLDFARB, eds. *The Economics of Artificial Intelligence*: University of Chicago Press, pp. 399-422.
- [62] VERISK, 2026. *2025 Annual Report*.
- [63] VERIZON, 2019. *Form 10-K: Annual report 2018*.
- [64] WALL STREET JOURNAL, 2024. *OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over \$250 Million* [online]. 22 May 2024, 12:00 [viewed 9 July 2026]. Available from: <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>
- [65] WEF, 2025. *Synthetic Data. The New Data Frontier. Briefing Paper*.
- [66] WIRED, 2013. *It's Official: Google Acquires Crowdsourced Navigation App 'Waze'* [online]. 11 June 2013, 12:00 [viewed 9 July 2026]. Available from: <https://www.wired.com/2013/06/google-waze-acquisition/>
- [67] WORLD BANK, 2025. *The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges*. Washington, DC: World Bank Group.
- [68] ZHANG, J., 2023. The paradox of data portability and lock-in effects. *Harvard Journal of Law & Technology*, 36(2), 657-706.

Authors

Dr. Tobias Guggenberger

Fraunhofer ISST

tobias.moritz.guggenberger@isst.fraunhofer.de

Dr.-Ing. Marcel Altendeitering

Fraunhofer ISST

marcel.altendeitering@isst.fraunhofer.de

Contact

Dr. Tobias Guggenberger

tobias.moritz.guggenberger@isst.fraunhofer.de

+49 231 97677-439

Fraunhofer-Institute for Software and Systems Engineering ISST

Speicherstraße 6

D-44147 Dortmund

Germany

www.isst.fraunhofer.de

Image Credits

©Boonyachot - istockphoto - 2267122414, Cover